



Behavioral policies and inequities: the case of incentivized smoking cessation policies

O. Çağlar Dede

To cite this article: O. Çağlar Dede (2019) Behavioral policies and inequities: the case of incentivized smoking cessation policies, Journal of Economic Methodology, 26:3, 272-289, DOI: [10.1080/1350178X.2019.1625223](https://doi.org/10.1080/1350178X.2019.1625223)

To link to this article: <https://doi.org/10.1080/1350178X.2019.1625223>



Published online: 07 Jun 2019.



Submit your article to this journal [↗](#)



Article views: 23



View Crossmark data [↗](#)



Behavioral policies and inequities: the case of incentivized smoking cessation policies

O. Çağlar Dede

Erasmus Institute for Philosophy and Economics (EIPE), Erasmus School of Philosophy, Erasmus University Rotterdam, The Netherlands

ABSTRACT

Behavioral policies, such as nudges and boosts, are gaining prominence. Such policies are advertised as evidential public policies. Yet, they have significant evidential problems. I analyze an important example of behavioral policy, so-called Incentivized Smoking Cessation Policies. I focus on their evaluation with respect to health inequities. I demonstrate that the evaluations of Incentivized Smoking Cessation Policies can be characterized by a plurality of researchers making use of different kinds of evidence gathering methods. I argue that the evaluation of Incentivized Smoking Cessation Policies' impact on health inequities can best be accomplished by integrating different evidence gathering methods. More generally, pluralism of evidence gathering methods adds another consideration to the debate about evidential requirements of behavioral policy assessment.

ARTICLE HISTORY

Received 18 February 2018

Accepted 27 August 2018

KEYWORDS

Behavioral policies; evidence; mixed methods; policy evaluation; nudge; inequality

1. Introduction

Behavioral policies aim to change people's behavior through interventions to choice contexts and psychological mechanisms of decision-making. Some well-known examples of behavioral policies are nudges (Thaler & Sunstein, 2008), boosts (Hertwig & Grüne-Yanoff, 2017), and behavioral-economics-informed incentivized programs for behavior change (Loewenstein & Chater, 2017). A growing body of research investigates the justifiability of behavioral policies. Behavioral economists, law scholars, public policy specialists, and those invested in evidence-based policy aim to identify the ethical, scientific, and institutional grounds for putting behavioral public policies into practice.

There is also a distinct *philosophical* literature on behavioral policies. Philosophers raise ethical worries about behavioral policies by, for instance, discussing the desirability of the normative principles these policies aim to instantiate and the attainability of these goals by different types and tokens of behavioral policies (e.g. Bovens, 2010; Hausman & Welch, 2010). Philosophers also raise epistemic worries about behavioral policies, asking 'how do we know that a behavioral policy is successful in achieving a desideratum in a certain environment?' (Barton & Grüne-Yanoff, 2015; Grüne-Yanoff, 2016; Grüne-Yanoff, Marchionni, & Feufel, 2018; Heilmann, 2014). The arguments focusing on this methodological question have been informed by the theories of human action underlying behavioral policies, such as the dual systems approach (Kahneman, 2011), the methodological accounts of experimental social science (e.g. Guala, 2005; Steel, 2008), and the philosophical literature investigating the epistemic and methodological requirements for the evaluation of evidence-based biomedical and social policies (Cartwright, 2012; Clarke, Gillies, Illari, Russo, & Williamson, 2014; Cartwright & Hardie, 2012).

In this paper, I contribute to this growing body of specialized methodological literature that assesses how empirical researchers predict and evaluate the success of behavioral policies (see Grüne-Yanoff, 2016 for a paradigmatic account). A common worry raised by the commentators in this literature is that the methodology used for evaluating behavioral policies is not conducive to a comprehensive analysis of these policies, as the evidence typically gathered establishes that these policies work, but without indicating how they do. Here, the focus is on the adequacy of Randomized Controlled Trials (RCTs) in assessing behavioral public policies. The use of RCTs, so the critics argue, does not serve researchers in investigating long-term, unintentional or distributional consequences of behavioral policies.

I address this literature through an analysis of how Incentivized Smoking Cessation Programs (*henceforth*, ISCP) are evaluated. ISCP are prominent examples of evidence-based behavioral health policies (Bhargava & Loewenstein, 2015, p. 400; Cabinet Office UK, 2011; Loewenstein & Chater, 2017; Volpp, Asch, Galvin, & Loewenstein, 2011). ISCP are thoroughly investigated or implemented in the UK and US in the last ten years. My analysis of ISCP is illuminating for the methodological literature for several reasons. First of all, ISCP are instances of a prominent type of behavioral public policies, so called ‘incentivized behavioral policies’, that so far have not been investigated by the specialized methodological literature. Secondly, and more importantly, the evidence-based evaluation of ISCP, as it is practiced in the UK, offers us rich resources for investigating specific methodological issues. Specifically, the evaluative perspectives and questions engaged in the evaluations of ISCP tend to be more comprehensive, unlike the evaluations of other commonly known behavioral policies, such as nudges. This includes the alternative methodological approaches to the use of RCTs.

I focus on the evaluation of ISCP’s impact on *health inequities*, hence the assessment of ISCP’s long-term effectiveness across specific sub-groups in the population. I argue that RCTs, when combined and synchronized with different evidence gathering methods, have distinct advantages in delivering inequity-relevant evidence over primarily RCT-based evaluations. More generally, I contend that this example gives us a reason to believe that a more pluralist evaluative methodology for behavioral public policies rectifies some of the commonly stated methodological limitations of primarily RCT-based extant behavioral policy evaluations.

The paper proceeds as follows. In section 2, I introduce and explain what Incentivized Smoking Cessation Policies are and argue that assessing how ISCP are evaluated affords promising insights that are relevant for the methodological literature on the evaluations of behavioral public policies. In section 3, I explicate what it means to evaluate a policy’s impact on health-inequities, and how ISCP are evaluated in this respect. In section 4, I assess the extant evaluative practices for ISCP’s impact on health-inequities and argue for a pluralist methodology. In section 5, I reflect on the implications of this analysis for the philosophical literature on the evaluations of behavioral public policies. In section 6, I conclude by emphasizing the importance of pluralism of evidence gathering methods and the community of researchers’ capacity to synchronize diverse methods to achieve more comprehensive evaluations of behavioral public policies.

2. Incentivized smoking cessation programs (ISCP)

2.1. What are ISCP?

It is well documented that smoking contributes to the development of serious non-communicable diseases such as type 2 diabetes, respiratory and cardiovascular diseases, and lung cancer (WHO, 2011).¹ Most smokers acknowledge such malignant health consequences of smoking. Yet, they often fail in their attempts to quit. There are various policy instruments to influence people’s smoking behavior. These include taxation of tobacco consumption, health-information campaigns, promotion of anti-smoking culture, and more coercive forms of regulation such as limiting the supply of tobacco and mandating smoking-free zones in cities. However, smoking remains a significant public health problem (e.g. NHS, 2017). Governments, therefore, actively seek new approaches

to smoking cessation policies (e.g. Department of Health UK, 2010, 2011; Commission on Social Determinants of Health [CDSH, WHO], 2008).

Behavioral economics offers a novel approach to smoking cessation (Cabinet Office UK, 2004, 2011; Dolan, Hallsworth, Halpern, King, & Vlaev, 2010; Loewenstein, Brennan, & Volpp, 2007; Loewenstein, Asch, Friedman, Melichar, & Volpp, 2012; Volpp et al., 2011). A well-studied example of the behavioral-economics-inspired smoking policies is that of Incentivized Smoking Cessation Policies (*henceforth*, ISCP) (e.g. Halpern et al., 2015; Sunstein, 2015; Volpp et al., 2009). ISCP promote quitting through *monetary or non-monetary rewards*. The supporters of ISCP (such as the Behavioral Insights Team in the UK, and the scholars such as Kevin Volpp, George Loewenstein, and Cass Sunstein) consider them applicable to small-scale environments such as firms or neighborhoods. They also consider ISCP as decentralized policies: implementers of ISCP could be a regional health service agency, or a non-governmental organization, or a private company. ISCP have been found attractive mostly in the US and in the UK. In the context of the US, large private companies have an interest in smoking cessation among their employees, as it reduces the insurance costs (Strickland, 2014). In the context of the UK, governments find it more desirable to approach smoking cessation at the regional, rather than national scale (Department of Health UK, 2011).

Some unsophisticated variants of ISCP were already implemented in the UK. For instance, the *Quit4U* program in Scotland (Ormston, van der Pol, Ludbrook, McConville, & Amos, 2015) and the *Give It Up For Baby* program in the UK (Ballard & Radley, 2009 and Radley et al., 2013) used financial rewards (such as food and shopping vouchers) to reduce smoking in socio-economically disadvantaged smokers. Based on results from behavioral economics, the proponents of ISCP suggest that the way in which incentives are presented can make a difference in changing smoking behavior (Loewenstein et al., 2012; Adams, Giles, McColl, & Sniehotta, 2013). It is, therefore, important not to conflate those standard policy-interventions that only alter financial incentives with the behavioral incentivized policies that alter incentives in sophisticated ways based on evidence regarding the cognitive or psychological models of decision-making.

Consider, for instance, the finding that *explicit* financial rewards, such as direct cash payments or holiday vouchers, are more likely to lead to behavior change than 'relatively more invisible incentives' of the same magnitude, such as costs tied to insurance premiums (Strickland, 2014; Volpp et al., 2009). The more salient the incentives are, the more likely the behavior change will be, or so the proponents of ISCP argue. Similar to the salience or visibility of incentives, there are other and much more complicated aspects of incentive-provision for behavior change that the proponents of ISCP deem helpful for designing and implementing ISCP (see Congdon, Kling, & Mullainathan, 2011; Tversky & Kahneman, 1974; Elster & Skog, 1999; Bhargava & Loewenstein, 2015 for the relevant behavioral economic literature). Importantly, behavioral economic studies suggest that people who engage in non-volitional health behaviors (such as smoking, binge-eating, and excessive gambling) might suffer from the effects of 'cognitive biases' that most humans have,² and that policy makers can make use of these biases for incentivized behavior change by harnessing them through offering incentives in varying immediacy, duration, frequency, or timing. Accordingly, the proponents of ISCP contend, for instance, that increasing the immediacy and frequency of incentives may render the incentivized quitting more likely to be effective, as it harnesses certain known cognitive biases of smokers (such as 'choice-bracketing' and 'present bias'; see Loewenstein et al., 2012 for a review).

At first glance, it seems that ISCP are promising evidence-based instruments for smoking policy that should perhaps be used much more widely. Should, for instance, governments encourage private companies and regional health services to adopt ISCP? In the rest of the paper, I will examine the methodological dimension of this question, leaving important ethical and political issues regarding ISCP aside (see Bovens, 2016; Schmidt, 2016; and Kelly, 2016 for a discussion of these matters). Indeed, I am solely interested in the question of whether ISCP should be adopted on *evidential grounds*; more specifically, which evidential sources warrant a positive evidential assessment of ISCP. It is also important to note that by 'an evidential assessment of ISCP' or 'an evaluation of ISCP' impact', I mainly refer to an *ex-ante* evaluation of evidence-base for the adoption of ISCP.³ To

this end, I will critically examine the methodologies researchers currently employ to evaluate ISCP. I will then discuss whether the employed evaluative methods are entirely adequate to assess these policies.

2.2. The importance of ISCP evaluations

While philosophers have analyzed the evaluations of well-studied types of behavioral public policies such as boosts and nudges (Grüne-Yanoff, 2016, Grüne-Yanoff et al. forthcoming), ISCP can neither be defined as boosts nor nudges. Nudges and boosts alter people's cognitive biases and heuristics to change behavior *while leaving the incentives intact*. ISCP and other incentivized behavioral policies do not fit this definition because they alter people's cognitive biases and heuristics in order to *increase the effectiveness of an incentive-altering intervention or an incentive-structure existing in the target environment*.

The categorical difference between nudge/boost and incentivized behavioral policies does not render the latter less of an important type of behavioral public policy, for two reasons. Firstly, incentivized behavioral policies are different from the traditional use of incentives in policy making, because the former make use of behavioral economic insights about how people respond to different ways of presenting incentives and information. Secondly, and more importantly, the *proponents* of the nudge agenda regard incentivized behavioral policies as a prominent type of behavioral public policies. Loewenstein and Chater (2017), for example, argue that behavioral public policy making should not be equated with nudging because the majority of available tokens of behavioral-economics-inspired policies are *not* instances of nudge (and nor of boosts for that matter), but rather of incentivized behavioral policies (see also Chetty, 2015; Bhargava & Loewenstein, 2015 for similar arguments).⁴

How incentivized behavioral policies are evaluated has not yet been investigated in the methodological literature focusing on the evaluation of behavioral public policies. ISCP certainly are prominent and well-referenced examples of this latter category of behavioral policies (see, for instance, Sunstein, 2015; Bhargava & Loewenstein, 2015). An investigation of how ISCP are evaluated is, therefore, an important addition to the methodological literature in its own right. Beyond these, there are two more aspects of ISCP and their evaluation that render them interesting.

Firstly, ISCP are evaluated by a wider range of evaluators than the range of evaluators corresponding to nudges or boosts. Nudges (and boosts, for that matter) have so far been evidentially assessed primarily by behavioral economists or behavioral scientists. As I will demonstrate in the next section, incentivized health policies such as ISCP, on the other hand, are evaluated not only by behavioral economists (e.g. Loewenstein et al., 2007; Giné, Karlan, & Zinman, 2010), but also by public policy scholars specialized in smoking cessation (e.g. Halpern et al., 2015), social epidemiologists (e.g. Adams, 2009), scholars from bio-medicine and preventive medicine (e.g. Bickel, Moody, & Higgins, 2016; Matjasko, Cawley, Baker-Goering, & Yokum, 2016), as well as prominent systematic reviewers of evidence-based public health interventions such as the Cochrane Collaboration, Campbell Collaboration, King's Fund, and NICE (e.g. Cahill, Hartmann-Boyce, & Perera, 2015; Jochelson, 2007).⁵ Importantly, the plurality of evidential evaluators of ISCP also implies a plurality of policy desiderata with respect to which behavioral public policies can and should be evaluated. For instance, the evidential evaluations of nudges have so far primarily focused on the question of whether nudges are effective, or the extent to which they are. The methodological literature thereafter questioned the extent to which the evaluations of nudging realizes the evaluative goal of short or long term effectiveness (e.g. Grüne-Yanoff, 2016). But the evaluators of ISCP, as we will see, are focusing on further evaluative goals of ISCP, such as the effectiveness of these policies in specific subgroups in a population, or in reducing smoking-related health inequities, or the minimization of unintended policy consequences.

Secondly, the evaluations of ISCP, as they are practiced in the UK, are based on a *plurality of methods*, which makes the case of evaluating ISCP different from the evaluation of nudges.⁶ The

evaluation of ISCP is primarily based on evidence gathered through randomized controlled trials (RCTs) and systematic reviews of RCTs.⁷ Systematic reviews of ISCP typically excludes or gives lower grades to non-randomized trials. Yet, importantly, some reviews consist of studies that make use of non-experimental and observational evidence such as the ones gathered through in-depth interviews.⁸ Just as the extant methodological studies questioned in how far the evaluation of nudging is practiced on sound methodological grounds, we may also ask in how far the evaluation of ISCP is based on the adequate use of the evidence gathering methods available. As the background research regarding the evaluation of public health policies tends to be comprehensive and well-documented in the UK, the evaluations of ISCP help us to focus on specific methodological problems regarding the use of experiments in the evaluation of behavioral policies discussed in the literature.

I have so far argued that we have good reasons to investigate the evaluation of ISCP from a methodological perspective. In the rest of the article, I will focus on the evaluation of ISCP's *impact on health inequities*. The evaluation of ISCP's impact on health inequities helps to further analyze the interesting and novel characteristics of the evaluation of incentivized behavioral policies that I listed above. More broadly, the health inequity focus will allow me to reflect on the question of whether behavioral public policies, as a new type of public policy, alter existing inequities and how this aspect of behavioral public policies should be evaluated.

In the next section, I will first explain what it means to evaluate a public policy's impact on health-inequities. I will then review what we know about the evaluation of ISCP's impact on health-inequities, introducing diverse perspectives of different kinds of researchers who provide us with relevant information for the assessment of how ISCP are evaluated.

3. The evaluation of ISCP with respect to health inequity

3.1. What does it mean to evaluate a policy's impact on health inequity?

What does the term 'health-inequity' mean? To start with, inequities in health outcomes should not be conflated with *inequalities* in health outcomes. Health *inequalities* are measurable differences in health outcomes across different populations and individuals. Health *inequities*, on the other hand, are health inequalities that result from people's unequal access to health services and capabilities to sustain healthier lives. For instance, health inequities in health outcomes in a country may result from its citizens' differential access or ownership of health services, nutritional sources, or health-related social capital. Being in worse health condition due to these disadvantaging and contextual factors is different than being worse-off due to unchangeable biological factors or volitional preferences. Hence, health inequities are commonly regarded as unnecessary, avoidable, unfair, and unjustifiable inequalities in health outcomes, which should be addressed by public policy interventions (O'Neill et al., 2014, Tugwell et al., 2010, Whitehead, 1992; WHO, 2012).

An effective health policy intervention may not be successful in reducing the inequalities in health outcomes between the most and the least disadvantaged individuals. For instance, it might be that an intervention is effective overall but ineffective or less effective for disadvantaged people. Similarly, a specific way to offer a public health intervention may discourage certain group of individuals from taking up the treatment, although that was not intended by the policy designers. The resulting inequities may result directly from the intervention itself or may appear as one of the already existing inequities prior to the intervention but exacerbated by the intervention (Lorenc, Petticrew, Welch, & Tugwell, 2013). Health inequities may, therefore, remain intact or exacerbated due to the mistakes in the implementation and design of public health policies as well as the knowledge gaps in the evaluation of these policies.

Accordingly, when I talk about 'evidence-based evaluation of policies with respect to inequities', I refer to the empirical investigation of the potential ways in which a policy may generate inequities once it is implemented. These kinds of investigations, then, aim at detecting if there are any

inequity-relevant mistakes in the implementation of the policies in question, or if there are knowledge gaps in this respect (see O'Neill et al., 2014 for a more detailed description).

Our question is: what sort of methodological practices lend us the evidential basis for making this sort of judgment about ISCP and similar behavioral public policies? I now turn to answering it.

3.2. A plurality of perspectives for the evaluation of the ISCP's impact on health-inequities

We now have fixed an understanding of what it means to evaluate a policy's impact on health inequity. I will now present an overview of how ISCP are evaluated in this regard. As we will see, it would be misleading to characterize the assessment of ISCP as a scientific activity that is governed by a homogeneous set of methodological principles, advising the use of a single method such as RCTs. As I articulate in this section, it is more appropriate to describe ISCP's assessment as a scientific activity that involves *different types of researchers* who have *different evaluative goals* and who make use of *different evidence gathering methods* in evaluating ISCP. Each type of researchers delivers some evidential output that is relevant for the assessment of ISCP with respect to inequities. It is precisely this plurality of evaluative and methodological perspectives that makes the investigation of ISCP interesting for informing the methodological debates regarding the evaluation of behavioral public policies. In the following, I will first describe the evaluation of ISCP with respect to inequities, demonstrating how the plurality of perspectives plays an important role. I will thereafter focus on the methodological lesson we should draw from this practice.

Table 1 offers a structured summary of my characterization of the specialized literature evaluating ISCP. I list different *Types of Researchers* who contribute to the evaluation of ISCP. I also give a few *Representative Examples* for the respective type of researchers. I indicate *Primary Evaluative Goals* that the different types of researchers aim for. I specify *Evidence Gathering Methods* the different types of researchers commonly use to gather the evidence in question. Finally, I mention some *References* from the literature that report or exemplify the kind of research in question.

Table 1 characterizes a number of categories for each of the *Type of Researchers* listed on the left. Let's start with *Decision-Makers*. *Decision-Makers* are policy makers or public policy agents who seek information to determine the justifiability of implementing ISCP in a certain context. Although *Decision-Makers* are, strictly speaking, not types of scientific researchers, they search for scientific consultants, commission reports of available evidence, and seek evidence-based arguments for making particular types of policy making. Two good examples of *Decision-Makers* in the context of ISCP are the NHS or the UK Department of Health, who not only consider political, ethical, economic or other, pragmatic, concerns for implementing ISCP, but also seek information about available evidence concerning the impact of ISCP (UK Department of Health, 2011). As policy makers, *Decision-Makers* demand certain types of evidential output, which then indirectly determines the kinds of evaluative goals the other *Types of Researchers* seek to deliver. For instance, evidence regarding health inequities is undoubtedly very important for *Decision-Makers* all around the world (CSDH, 2008; Petticrew, 2004; UK Department of Health, 2011). More specifically, smoking control is one of those areas in public health where the evidence on policies' impact on specific disadvantaged groups and inequities is highly important.⁹ Because there are widely accepted smoking-related inequities in health across various socio-economic-demographic strata, health policy's success in reducing health inequities is crucial to investigate.

Who then are the *Types of Researchers* that are supposed to assemble the evidence relevant for the assessment of ISCP with respect to health inequities? Firstly, there are the *Proponents of Behavioral Public Policies* who advocate ISCP. Secondly, we have the *Evidence-based Policy Specialists* who assess ISCP and similar policy proposals based on evidence. Thirdly, we have *Systematic Reviewers* who review, rate, and report available evidence-based assessment of ISCP. Finally, there are different kinds of scientists associated with different research disciplines such as *Behavioral Economists* and *Social Epidemiologists* who provide all the other *Types of Researchers* with the relevant

Table 1 The characterization of the specialized literature evaluating incentivized smoking cessation policies: a plurality of evaluative and methodological perspectives.

Type of Researchers	Representative Researcher	Primary Evaluative Goal	Evidence-Gathering Methods	References
<i>Decision-Makers</i>	National Health Services (NHS) in the UK	Specify evidence required for policy making (e.g. ISCP's effect on health inequities)		CSDH (2008), Petticrew (2004), Department of Health (2011)
<i>The Proponents of Behavioral Policies</i>	Behavioral Insights Team in the UK	Defend behavioral policies such as ISCP based on evidence and propose new tokens of behavioral policies	Behavioral economic theory, experimental evidence, RCTs	Haynes et al. (2012), Dolan et al. (2010)
<i>Evidence-based Policy Specialists</i>	(a) RCT specialist Policy Evaluators (e.g. Kevin Volpp, Scott Halpern) (b) Public Health Evaluators (e.g. Jean Adams, Heather Morgan, Gill Thomson NICE in the UK, Higgins)	Measure the ISCP's (and similar Incentivized policies') impacts such as effectiveness, persistence, cost-effectiveness, unintended consequences, effects on specific populations	RCTs Observational Methods, Systematic Reviews, Mix-methods	Halpern et al. (2015), Volpp et al. (2009), Morgan et al. (2015), Higgins and Solomon (2016)
<i>Behavioral Economists</i>	George Loewenstein, Paul Slovic	Pursuing empirical and theoretical knowledge about the psycho-cognitive factors that determine the differential effects of incentives and information across different types of smokers.	Theory, lab and field experiments (methodologically individualist orientation).	
<i>Social Epidemiologists</i>	Micheal Kelly Hilary Graham Jennie Popay Stanley Blue	Empirical and theoretical knowledge about the structural factors that determine various aspects of smoking behavior, the health inequalities and differential effects of public health policies.	Theory, observational methods including econometrics descriptive statistics and qualitative methods (with a holistic methodological orientation)	Kelly (2010), Graham (2011), Popay (2008), Blue et al. (2016)
<i>Systematic Reviewers</i>	Cochrane Collaboration, Campbell Collaboration, Kings's Fund	(i) Review, rate, and report available evidence on various impacts of ISCP; (ii) inform the design of new evidence-based policy analysis (e.g. by generating new hypothesis, defining gaps in evidence)	All of above and methods of structured review	Cahill and Perera (2011, 2015), Jochelson (2007)

theoretical and empirical input gathered through the practicing so called ‘primary science’ regarding smoking behavior and health inequities. Let’s first focus on *Proponents of Behavioral Policies*.

The *Proponents of Behavioral Policies* are policy researchers who advocate a particular policy approach, that is the behavioral approach. They aim to inform or else convince *Decision-Makers* based on scientific evidence regarding the performance of behavioral policies. In the context of behavioral public policies and ISCP, the researchers in the Behavioral Insights Team (BIT) of the UK are a good example of this category of researchers. BIT aims to justify behavioral public policies based on evidence. Hence, its main evaluative goal is to gather or report evidence that supports various tokens of behavioral public policies such as ISCP. Now, it is important for our discussion to note that BIT in the UK, and similar so-called Nudge Units around the world, adopt a particular methodological strategy to evaluate behavioral policies. That strategy prioritizes RCTs as the evidence gathering method that should be used for evaluating behavioral policies. This strategy is often explicitly stated. For instance, in one of BIT’s methodological reports, the researchers claim that ‘Randomized Controlled Trials are at the heart of the Behavioral Insights Team’s methodology’ and that ‘RCTs are the best way of determining whether a policy is working’ (Haynes, Goldacre, & Torgerson, 2012, p. 4). We now have a more detailed characterization of the *Proponents of Behavioral Policies* in the context of UK, and their methodological strategy. Let’s focus on the next type of researchers engaging in the evaluation of ISCP: *The Evidence-based Policy Specialists*.

The *Evidence-based Policy Specialists*’ primary aim is to measure ISCP’s (and similar Incentivized Policies’) success with respect to various policy desiderata. Effectiveness is their most commonly presupposed policy desideratum; however, other evaluative goals such as investigating a policy’s cost-effectiveness, persistence, unintended consequences, and impacts on specific populations or inequities are also pursued. I consider all advanced researchers who pursue this kinds of evaluative goals as evidence-based policy specialists. However, for the purpose of characterizing diverse methodological lines in the ISCP’s evaluation with respect to inequities, I focus on two qualitatively different examples of *Evidence-based Policy Specialists*: (i) *RCT-specialists* and (ii) *Public Health Evaluators*.

Similar to BIT’s methodological approach, the empirical literature evaluating ISCP is marked by the prominence of randomized trials comparing the effectiveness of various ISCP (e.g. Halpern et al., 2015). *RCT-specialists* evaluating ISCP are primarily interested in determining whether a token of ISCP is effective in an environment for a particular target. As it is well-known, RCTs work perfectly in pursuing that particular evidential output. Hence, it is no surprise that *Evidence-based Policy Specialists* who investigate the effectiveness of ISCP base their research primarily on RCTs.

Public Health Evaluators, on the other hand, are akin to social policy scholars who assess health policies from the perspective of public health. Hence, their primary research interest and evidential output goes beyond the specification of ISCP’s effectiveness and includes the specification of ISCP’s effectiveness in certain groups in a population; for instance, ISCP’s impact on inequities, or ISCP’s cost-effectiveness. *Public Health Evaluators* are therefore not a homogenous category of researchers. Rather, they occupy roles in different disciplines such as epidemiology, social epidemiology, social policy, economics, sociology, preventive medicine, etc. This variety is then reflected in the different evidence gathering methods employed by Public Health Evaluators across disciplinary backgrounds. For instance, researchers in NICE tend to use qualitative and observational types of evidence gathering methods, which are pertinent to disciplines such as epidemiology, social epidemiology, or sociology of health (e.g. NICE, 2007). Preventive medicine scholars, on the other hand, tend to pursue their primary evaluative goal regarding ISCP through laboratory experiments (e.g. Higgins et al., 2012).

I should emphasize that these two categories of example researchers, *RCT-specialists* and *Public Health Evaluators*, are not mutually exclusive. In other words, an RCT-specialist may also be a public health evaluator (e.g. Jean Adams). I distinguish between the two in order to emphasize the following point: *Evidence-based Policy Specialists* use different evidence gathering methods for evaluating various aspects of ISCP. More generally, that is to say, *Evidence-Based Policy Specialists*

as a type of researcher does not correspond to a homogenous body of researchers with respect to the primary evaluative goal and evidence gathering methods used.

The next *Type of Researchers* are scientists who deliver primary theoretical input and empirical evidence relevant to the evaluation of ISCP. Since smoking cessation is a complex and multi-faceted scientific subject, there are many such different kinds of 'primary scientists' involved. As they are the most relevant ones for the assessment of ISCP's impact on health inequity, I focus here on *Behavioral Economists* and *Social Epidemiologists*. Behavioral economists assemble empirical and theoretical knowledge on the psycho-cognitive factors that determine the differential effects of incentives and information across different types of smokers (Giné et al., 2010; Loewenstein et al., 2012). *Social epidemiologists*, on the other hand, analyze the impacts of social-structural factors on individual, population health states, health-related social practices and health behavior (Honjo, 2004). Social-structural factors are commonly referred as 'wider determinants of health' by epidemiologists. Social epidemiologists generally make use of observational evidence gathering methods to investigate health inequities.

Finally, we have the *Systematic Reviewers* who review, rate, report and evaluate available evidence about various impacts of ISCP. The most prominent examples of *Systematic Reviewers'* research in the context of health interventions and ISCP are assembled by major evidence-based policy institutions in public health and biomedicine such as the Cochrane Collaboration, Campbell Collaboration, and Kings' Foundation in the UK. Systematic reviewers play a major, arguably the most crucial, role in the evaluations of health interventions. The most explicit and the main contribution of *Systematic Reviewers* is the reporting of evidence in a way that is useful for decision-making. Hence, *Systematic Reviewers'* judgments directly inform *Decision-Makers* (at least in the case of UK). A less salient but a very important contribution of *Systematic Reviewers* is the evaluation of available evidence so as to inform the design and implementation of new evidence-based policy assessment, ex-ante. That is to say, *Systematic Reviewers'* evidential output informs *Evidence-based Policy Specialists'* research. Specifically, *Systematic Reviewers* do so by generating new hypotheses, determining gaps in evidence or theories provided by *Primary Scientists*, and communicating the relevant evidential demands on behalf of *Decision-Makers*.

Scientists delivering primary theoretical and empirical evidence relevant to the evaluation of ISCP (e.g. *Behavioral Economists* and *Social Epidemiologists*) may also benefit from the *Systematic Reviews'* research output in the same way as *Evidence-based Policy Specialists* do. However, the interaction between *Systematic Reviewers* and *Evidence-based Policy Specialists* is much more direct than the one between *Systematic Reviewers* and *Primary Scientists* in practice. This is the case firstly because *Evidence-Based Policy Specialists* draw generally on evidence available from the systematic reviews (e.g. Evidence-based policy specialists' research articles would generally include a 'background' section where the evidential output of relevant Systematic Reviews is reported). Secondly, it is often the case that an Evidence-based policy specialist is also a specialist in systematic reviews (e.g. researchers such as Jean Adams, Gill Thomson, and Heather Morgan).

What do we know about the evidence gathering methods *Systematic Reviewers* use in pursuing their primary evaluative goal? As I described above, *Systematic Reviewers* deliver two types of evidential input for the evaluation of ISCP: one that is relevant for reporting, another one that is relevant for further evidence-based policy assessment. The systematic reviews rely on observational evidence gathering methods in reviewing available evidence. However, it is also appropriate to speak of heterogeneity of ISCP's systematic reviews in terms of the kinds of evidence selected for the reviews. Depending on the aims and the orientations of *Systematic Reviewers*, systematic reviews of ISCP sometimes draw only on RCT-based evaluations of ISCP (e.g. when assessing overall effectiveness, see for instance, Cahill & Perera, 2011). Yet, they also draw on other evidential output assembled by the use of alternative evidence gathering methods including the *Public Health Evaluators'* assessment based on non-RCT studies, theoretical and empirical evidence delivered by *Social Epidemiologists* and *Behavioral Economists* (e.g. when assessing aspects of interventions other than the effectiveness, see for instance, Thomson et al., 2014). Systematic reviews sometimes integrate

different evidence gathering methods. Such reviews, therefore, significantly contribute to methodologically more integrative evaluations of ISCP. *Systematic Reviewers'* research output, which may be reinforced by the multiple evidence gathering methods, also serves as observational evidence that informs *Evidence-based Policy Specialists' research* (see section 4 for an example).

Here, I have offered a general definition of what it means to evaluate a policy's impact on health-inequities (3.1) and have reviewed the different sources and types of available evidence relevant for judging whether ISCP reduce health inequities in diverse contexts and demonstrated that there are different methods of evidence gathering involved (3.2).

I will now assess how ISCP's impact on health-inequities is evaluated. Based on this analysis, I will specifically argue that the evaluation of ISCP through the combination of different evidence gathering methods has distinct advantages in delivering inequity-relevant evidence in comparison to primarily RCT-based evaluations [advocated by the proponents of behavioral public policies and some of the RCT-specialized policy evaluators represented in the table (e.g. Haynes et al., 2012)]. More generally, I contend that this example gives us a reason to believe that a more pluralist evaluative methodology for behavioral public policies rectifies some of the commonly stated methodological limitations of extant behavioral policy evaluations which are pertinent to reliance primarily on RCTs.

4. RCTs integrated with different evidence gathering methods for the evaluation of ISCP's impact on health inequity

I have offered an overview of different types of evaluators of ISCP and different evidence gathering methods they use for delivering inequity-relevant evidence. I will now examine the adequacy of these evidence gathering methods in investigating ISCP's success with respect to the reduction of health inequities. My purpose is not to propose a single best methodology for the evaluations of ISCP; however, I will argue that RCTs fare better in delivering relevant evidence *when integrated with alternative evidence gathering methods*. This argument, to the extent that it is an argument regarding the use of RCTs in policy evaluation, is concerned with how RCTs can be used in a better way rather than stating how limited RCTs are. In this section, I will first review how well-known limitations of RCTs arise in the context of evaluating ISCP's impact on health inequities. I will then illustrate how some evaluative studies integrate RCTs with different evidence gathering methods and argue for a more integrated evaluative methodology for the evaluation of ISCP.

Let me first review the limitations of RCTs as an evidence-gathering method used for the evaluation of ISCP's impact on health inequities. To do so, I consider the evidential output delivered by the RCT-specialists who are *Evidence-based Policy Evaluators* and the *Proponents of Behavioral Policies* who primarily rely on RCTs.

I will offer two examples of the kind of evidential output that is relevant for evaluating ISCP's impact on inequities, yet not delivered by *Evidence-based Policy Evaluators* and the *Proponents of Behavioral Policies* who only use RCTs.

The first kind of evidence is concerned with sub-groups, specifically the group of socioeconomically disadvantaged smokers. Based on available systematic reviews, what we know is that primarily RCT-based assessments which report that some tokens of ISCP are significantly effective in ceasing smoking (e.g. Halpern et al., 2015; Volpp et al., 2009) are not informative about effectiveness in *specific subgroups*, such as disadvantaged smokers. The following comment that appeared in a Cochrane Collaboration systematic review concerning these studies is quite telling in this respect:

Since both trials enrolled employees of large American companies, who were predominantly white and enjoyed relatively high levels of education and income, their success may not be readily generalizable to other populations of smokers, with different regional, socio-economic and ethnic mixes. (Cahill & Perera, 2011)

The lack of evidence regarding ISCP's impact on specific disadvantaged groups is a major limitation for making evidence-based judgments about ISCP's impact on health inequities. For instance, without comprehensive information about the subgroups of the population under investigation, one cannot

judge whether ISCP's effectiveness is modified across smokers with different demographic characteristics, or whether some overall effective ISCP are not successful in ceasing smoking for disadvantaged groups. A possible methodological reply to this challenge might be to carefully stratify the population of the experiment prior to the experiment, thus to define the different subgroups. However, doing so in the right way in fact invites researchers to use alternative evidence gathering methods such as observational studies and descriptive statistics together with RCTs, as I will illustrate further below.

The second example I would like to put forward concerns the lack of evidence regarding the *long-term effectiveness* of ISCP. The long-term effectiveness is crucial for understanding ISCP's impact on health inequities. There is well-known social epidemiological evidence, based on qualitative and quantitative observational studies, reporting on the specific challenges of *long-term* smoking cessation in the context of disadvantaged smokers (e.g. due to stressors associated with social and economic exclusion). Many social epidemiologists would consequently argue that post-ISCP smoking behavior would be different across groups, even if an ISCP is initially successful, anticipating that disadvantaged smokers are more likely to relapse (Blue, Shove, Carmona, & Kelly, 2016; Popay, 2008). Although such social epidemiological evidence is theoretically plausible, it gives us at best an *indirect* or *prima facie* reason to believe that disadvantaged quitters are more likely to relapse months after a successful, ISCP-generated, abstinence. But those who favor ISCP may always demand further and stronger evidence to believe in social epidemiologists' arguments against the effectiveness of ISCP. Although evidence gathered through RCTs are generally considered stronger than observational evidence, RCT-based studies of ISCP fail to provide us with evidence confirming or disconfirming arguments pro or against ISCP's long-term effectiveness in disadvantaged smokers. Specifically, primarily RCT-based studies of ISCP do not deliver information about the distribution of relapse behavior across different strata of smokers in the long term.

Based on the reports of systematic reviews, what we know is that the smoking abstinence generated by effective ISCP usually does not last longer than a couple of months after the incentives are withdrawn (Cahill & Perera, 2011; Jochelson, 2007; Marteau & Mantzari, 2015). Yet, based on the considerations I stated above, it would be crucial to have RCT-based evidence indicating whether the post-ISCP relapse behavior is stratified and modified by the characteristics of disadvantages, such as unemployment or social exclusion, as many social epidemiologists would anticipate. As I will illustrate further below, such information can be more easily delivered when RCTs are integrated with alternative evidence gathering methods that are more suitable for *predicting* the stratification or modification effects in the long-term (e.g. by studying specific mechanisms of behavior change through qualitative studies or further modeling by primary scientists).

Now, it is *not* surprising that the evaluations of ISCP that rely *only* on evidence gathered through RCTs do *not* deliver evidence on these two aspects. RCTs are considered more adequate tools for determining the overall effectiveness of interventions rather than the variation of effectiveness across sub-groups. Similarly, since RCTs are not supposed to give information regarding *how* or through which mechanisms an intervention works, they are similarly not informative about the long-term impacts. These issues have been widely discussed in the relevant philosophical literature on policy evaluation in general (e.g. Cartwright & Hardie, 2012), and the behavioral public policy evaluation in particular (e.g. Grüne-Yanoff, 2016). It is therefore well known that RCTs have limitations in delivering comprehensive evidence regarding the effects of interventions (such as evidence characterizing the heterogeneity of subgroup in the target population and differential distribution of effects across different subgroups, longer-term effects of the intervention, or how the intervention interacts with the context of the target environment). My aim is not to advance upon on these well-known critiques of RCTs, or to offer a new one. I fully acknowledge these criticisms and point out that the same issues arise in the context of ISCP evaluation as well. I also do not suggest that RCTs are *in principle* uncondusive to investigate health inequities. My aim is rather to make a constructive methodological claim regarding how RCTs, as they are currently employed in the case of ISCP, might be designed and harnessed better for the purpose of making judgments about how behavioral policies such as ISCP fare with inequities.

To this end, I suggest that the *Evidence-based Policy Specialists'* assessments of ISCP can and do actually deliver the necessary evidence when they *integrate different evidence gathering methods with RCTs*. To demonstrate this and to exemplify what kinds of evidence gathering methods are needed, let me offer a closer look at those methodologically more integrative evaluations of ISCP and why they perform better.

Consider Morgan et al.'s (2015) investigation of the incentives for smoking cessation during pregnancy conducted for the NHS in the UK. This social scientific research involves multiple methodological steps, but it is possible to represent Morgan and her colleagues' investigation in two parts for the sake of understanding how they integrate multiple evidence gathering methods. The first part of their study involves a systematic review of RCT-based evaluations of various ISCP's effectiveness, a report of qualitative and theoretical literature about the mechanism of incentive-based behavior change, and a collection of socio-epidemiological and behavioral scientific evidence regarding the barriers and facilitators of smoking cessation during pregnancy in the context of socioeconomic disadvantage. Morgan et al.'s aim in the first part is to integrate these different pieces of available evidence so as to inform the *design* and the *scope* of the second part of their study. The second part of the study involves conducting primary qualitative studies (based on structured interviews) to understand how the target audience of ISCP trials, which were carefully pre-selected in the first part of the study, responds to incentive provision. That is to say, Morgan and her colleagues make use of qualitative studies in order to assemble comprehensive evidence regarding the working and the consequences of RCT-based trials. In a simultaneous study based on the same data, Thomson et al. (2014) gather evidence specifically relevant for making judgments about various *unintended consequences* of ISCP under investigation for disadvantaged smoking pregnant women in the UK. Their results do also inform judgments about ISCP's impact on health inequities based on empirical evidence gathered in these two simultaneous studies.

The integrated methodology, exemplified in Morgan and her colleagues' study, extracts inequity-relevant information from the available RCT-based studies of ISCP which do not necessarily have inequity-relevant content. It does so by making use of different evidential sources such as the evidence delivered by *Primary Scientists* regarding the possible mechanisms, barriers and facilitators of smoking cessation, and evidence delivered by qualitative literature concerning which ISCP trials have failed, for which sub-groups, and how. In doing so, it reveals comprehensive evidence regarding subgroups, which is not readily available from RCTs. Moreover, the integrated methodology also better predicts the potential long-term consequences of selected ISCP by conducting follow up structured interview studies in order to extract information about potential modifications or unintended consequences of the interventions. It thereby succeeds in rectifying the abovementioned limitations of primarily RCT-based evaluations.

Moreover, the studies which integrate multiple methods do also a better job in inequity assessment than those studies which rely *only* on social epidemiological evidence or qualitative methods. While the latter provides *prima facie* indirect descriptive and theoretical evidence for ISCP's various impacts, the former deliver more direct and comprehensive causal evidence, which also makes use of available social epidemiological evidence.

This kind of more integrated methods are increasingly advocated by the *Evidence-based Policy Specialists* and *Systematic Reviewers* (O'Neill et al., 2014; Petkovic et al., 2017; Welch et al., 2010; Welch et al., 2015), and the number of similar studies is increasing, as demands for evidence-based evaluations of inequity increases (NICE, 2013). To the best of my knowledge, none of the behavioral public policies or incentivized behavioral health policies has been evaluated in this manner, but there is no principled reason against doing so.

I have suggested that the evaluations of ISCP that integrate different evidence gathering methods with RCTs fare better in delivering pieces of evidence relevant for the inequity assessment in comparison to those which rely only on RCTs, or only on social epidemiological methods. This argument gives us a nuanced view regarding what RCTs can and cannot do in evaluating behavioral public

policies. I will now also discuss in what way the argument advanced so far also contributes to the following more general epistemological question: what kind of *evidence* does the justifiability of behavioral public policies require?

5. Discussing evidence gathering methods in the philosophy of behavioral public policies

I have illustrated that there are different kinds of evidence gathering methods involved in the evaluations of ISCP. I then have claimed that the evaluation of ISCP through the *integration* of different evidence gathering methods has distinct advantages in delivering inequity-relevant evidence in comparison to studies relying on single methods. I will now specify how my argument relates to more general philosophical debates.

Philosophers contribute to the evidence-based policy making by specifying the kinds of ideal epistemic requirements (pertaining to the nature of causal knowledge necessary for policy-making purposes) the evidence-based policies should meet, and by determining the sources of evidential gaps existing in the practice of evidence-based public policy justification (e.g. Cartwright & Hardie, 2012). Philosophers specialized in this literature often make a distinction between two broad categories of evidence: evidence of difference-making and evidence of mechanisms. There are controversies regarding what mechanisms are (Illari & Williamson, 2012), what counts as mechanistic or difference-making evidence (Illari, 2011), and how mechanistic evidence relates to the difference-making evidence (Clarke et al., 2014).

A general contention of the philosophers regarding the received methodology of evidence-based policy assessment (such as the idea of ‘evidence hierarchies’ in bio-medical and public health interventions) has been that the assessment of these policies is based on a limited array of evidence, although inclusion or prioritization of different categories of evidence is sometimes necessary. Evidence-based public policy assessment is primarily based on randomized controlled trials, which is commonly seen as delivering evidence of difference making. Philosophers often demand evidence of mechanisms for enhancing the evaluation of evidence-based policies (Russo & Williamson, 2007; Clarke et al., 2014). The former denotes the kind of evidence establishing that a policy-intervention or a treatment makes a difference in a target environment. The latter denotes the kind of evidence establishing how (through which mechanisms) a policy intervention makes a difference in a target environment. Grüne-Yanoff (2016) discusses these issues in the context of behavioral public policies such as Nudges. He observes that behavioral economists typically do not gather evidence of mechanisms to assess behavioral policies and considers this an important drawback for the justification of these policies.¹⁰

The dichotomy between evidence of difference-making and evidence of mechanisms definitely helps us understand some of the limitations of policy evaluations that are based *primarily on RCTs* (see, for instance, Russo & Williamson, 2007 in the context of evidence-based medicine). I am broadly supportive of this conclusion, as I too try to spell out how RCT-based evaluations of behavioral policies can be improved. Yet, the argument I advanced in favor of the pluralism of evidence gathering methods is different from the arguments made in favor of or against mechanistic evidence in at least two ways. Firstly, the former and the latter have different evaluative goals. My analysis aims to answer how behavioral policies’ *impact on inequities* should be analyzed; whereas the call for mechanistic evidence is motivated to answer how behavioral policies’ *efficacy* should be assessed in different contexts. Secondly, my argument and the argument(s) for mechanistic evidence operate on different levels. While mine is a demand for a pluralism of *methods*, a demand for mechanistic evidence is about the *content of evidence* required for the respective evaluative goals. Of course, an implication of evidential pluralism as a thesis about the content of evidence for policy evaluation might be that we need to integrate different evidence gathering methods to deliver evidence of difference making and mechanisms. To the extent that evidential pluralists are happy to endorse

the pluralism of evidence gathering methods for policy evaluation, this article provides further support for evidential pluralism in the context for behavioral public policy evaluation.

I would be broadly sympathetic to draw closer and more precise connections between the demand for pluralism of evidence gathering methods and the demand for evidence of mechanisms. For instance, one might argue that the integration of evidence gathering for inequity assessment demands only specific kinds of mechanistic evidence and specific kinds of difference-making evidence. Or one might also quite naturally contend that the demands for pluralism of evidence gathering methods and need for mechanistic evidence could imply or complement each other, depending on how these positions are formulated exactly, or to which case they apply. Spelling out an exact relationship between my analysis and the need for mechanistic evidence, however, goes beyond the purpose of the present article. My aim has been to advance a specific perspective on the evaluations of behavioral public policies, one that I hope will be a useful addition to the discussions about the need for evidence of mechanism in this broader philosophical literature as well.

6. Conclusion

I have offered new insights for evidential evaluations of behavioral policies by focusing on evaluative challenges of a prominent example of behavioral health policies (Incentivized Smoking Cessation Policies). I have contended that evaluators of behavioral policies should go beyond a primarily RCT-based methodology for more comprehensive evaluations. I have also offered a pluralist evaluative methodology in which RCTs are integrated with alternative evidence gathering methods.

I focused on the *evaluation* of incentivized behavioral policies; however, I also believe that my comprehensive and practice-oriented analysis of ISCP also expands our understanding of behavioral policies more broadly. My focus on a fundamental public policy problem, the evaluation of health inequity aspects, invites readers to think about my analysis also in a different way: as exploration of a salient way in which behavioral economics, the ‘run-away success story of contemporary economics’ (Angner, 2015), modifies standard public policies, conventional ‘welfare regimes’ (Esping-Andersen, 1990/2013), and claims to rectify the standard economics approach in addressing one of the centuries-old central problems of welfare-provision, how to address inequities. My article, therefore, provides a rationale for why philosophers and evaluators of behavioral policies may want to adopt a more interdisciplinary toolbox and mindset in addressing the potential merits and pitfalls of behavioral policies as *evidence-based public policies*.

Notes

1. The numbers cited by the WHO report about the non-communicable diseases around the globe are striking:

Almost 6 million people die from tobacco use each year, both from direct tobacco use and second-hand smoke [...] Smoking is estimated to cause about 71% of lung cancer, 42% of chronic respiratory disease and nearly 10% of cardiovascular disease. (p.1)

2. For example, the so-called ‘peanut effect’ consists in underestimation of cumulated long-term effects of minor negative health behavior (Weber & Chapman, 2005). ‘Present bias’ is the label for overvaluation of prospects that occur sooner in time (O’donoghue & Rabin, 2000). The so-called ‘choice-bracketing’ is defined as valuing a particular decision in isolation from the larger incentive structure (Read, Loewenstein, & Rabin, 1999). The so-called ‘loss aversion’ is the overvaluation of small losses than gains of equal magnitude (Thaler, Tversky, & Kahneman, 1997). These are some of the various heuristics and biases that behavioral economists use to explain smoking behavior and its change.
3. As I will explain in section 3, it is possible that an ex-post evaluation of evidence-base for an already-implemented ISCP is considered as an evidential source in an ex-ante evaluation of ISCP (e.g. the use of systematic reviews in the evaluations of ISCP).
4. Loewenstein & Chater, 2017 also argue that incentivized behavioral policies have a wider range of applicability than nudging because the conditions for the proper applications of ideally defined nudges are harder to instantiate in practice than that of incentivized behavioral policies.
5. At least, this is the case in the UK where the comprehensive evaluation of ISCP is mostly practiced.

6. Grüne-Yanoff, 2016 argues that the evaluations of nudges are primarily based on lab and field experiments.
7. See, for example, Halpern et al., 2015 as a typical example of an evaluation of ISCP based on RCTs and Cahill et al., 2015 for a prominent example of a systematic review which evaluates the Halpern et al.'s study along with other empirical investigations of ISCP.
8. See Morgan et al., 2015; Thomson et al., 2014; Adams et al., 2013; Graham, Sowden, Flemming, Heirs, & Fox, 2012, for an examples of systematic reviews of ISCP that make use of non-experimental evidence.
9. Statistical data show that although overall smoking has decreased over time in many countries, smoking prevalence is unevenly distributed across different socio-economic classes in all countries (NHS, 2017). For instance, according to the most recent statistics in the UK, where overall adult smoking prevalence is lower than in many other European countries, socioeconomically disadvantaged citizens are more likely to be a smoker than affluent ones (classified as those with lower income, less education, routine and manual jobs; as those with higher income, higher education, managerial and professional jobs; respectively) (NHS, 2017, part 3). In the same way, the influence of being an unemployed person or a member of an ethnic minority in the UK increases the likelihood of being a smoker. Moreover, smoking during pregnancy is highest in the economically poorest regions of the country. Also, smoking during pregnancy is more common among young mothers who are likely to experience financial and social distress. Similar facts about the socio-economic distribution of tobacco use also hold for other West-European countries where overall smoking prevalence is very low in comparison to the rest of the world (Verdurmen, Monshouwer, & Van Laar, 2015; Baha, Boussadi, & Le Faou, 2016).
10. He provides examples from the literature so as to demonstrate that a lack of sufficient mechanistic evidence regarding how these policies make a difference in people's behavior avoids us to assess various policy desiderata attached to these policies. For instance, without mechanistic information, Grüne-Yanoff argues, one cannot determine whether a default nudge program will be effective across different target environments, and whether it will lead to robust, persistent, and welfare-improving behavioral changes. This is the evidence that is assembled by behavioral economists corresponding to their explanatory models of choice behaviors. For example, default nudges make a difference in people's behavior because of a mechanistic model A that specifies 'inertia' as the main causal entity or process that leads to the conclusion, or model B cites recommendation effect, or C loss-aversion etc.

Acknowledgement

I would like to thank in particular Conrad Heilmann and Jack Vromen, and also George Ainslie, Anna Alexandrova, Emrah Aydinonat, James Grayot, Till Grüne-Yanoff, Donal Khrosrowi, Magdalena Malecka, Michiru Nagatsu, Federica Russo, Attilia Ruzzene, Daphne Truijens, Sarah Wieten, and Nicolas Wütrich, as well as two anonymous referees of this journal and audiences in Cape Town, Enschede, Istanbul, Helsinki, Rotterdam, and London for their very helpful feedback.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

O. Çağlar Dede is a PhD candidate in philosophy at the Erasmus Institute for Philosophy and Economics (EIPE) in Rotterdam. His research focuses on philosophy of science and the relation of science and policymaking. In particular, he investigates the evaluation of evidence in behavioral public policies and discusses the role of values and citizen involvement in evidence-based policy.

References

- Adams, J. (2009). The role of time preference and perspective in socioeconomic inequalities in health-related behaviours. In S. J. Babones (Ed.), *Social inequality and public health* (pp. 9–25). Bristol: Policy Press.
- Adams, J., Giles, E. L., McColl, E., & Sniehotta, F. F. (2013). Carrots, sticks and health behaviours: A framework for documenting the complexity of financial incentive interventions to change health behaviours. *Health Psychology Review*, 8(3), 286–295.
- Angner, E. (2015). To navigate safely in the vast sea of empirical facts. *Synthese*, 192, 3557. doi:10.1007/s11229-014-0552-9
- Baha, M., Boussadi, A., & Le Faou, A.-L. (2016). French smoking cessation services provide effective support even to the more dependent. *Preventive Medicine*, doi:10.1016/j.ypmed.2016.06.024
- Ballard, P., & Radley, A. (2009). Give It Up For Baby-A smoking cessation intervention for pregnant women in Scotland. *Cases in Public Health Communication & Marketing*, 3, 147–160.

- Barton, A., & Grüne-Yanoff, T. (2015). From Libertarian Paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, 6(3), 341–359. doi:10.1007/s13164-015-0268-x
- Bhargava, S., & Loewenstein, G. (2015). Behavioral economics and public policy 102: Beyond nudging. *American Economic Review*, 105(5), 396–401. doi:10.1257/aer.p20151049
- Bickel, W. K., Moody, L., & Higgins, S. T. (2016). Some current dimensions of the behavioral economics of health-related behavior change. *Preventive Medicine*, 92, 16–23. doi:10.1016/j.ypmed.2016.06.002
- Blue, S., Shove, E., Carmona, C., & Kelly, M. P. (2016). Theories of practice and public health: Understanding (un)healthy practices. *Critical Public Health*, 26(1), 36–50. doi:10.1080/09581596.2014.980396
- Bovens, L. (2010). Ethics of nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference change: Approaches from philosophy, economics and psychology* (pp. 207–220). Dordrecht: Springer.
- Bovens, L. (2016). Don't mess with my smokes: Cigarettes and freedom. *The American Journal of Bioethics*, 16(7), 15–17. doi:10.1080/15265161.2016.1180461
- Cabinet Office UK. (2004). Personal responsibility and changing behaviour: The state of knowledge and its implications for public policy. Retrieved from <http://webarchive.nationalarchives.gov.uk/+http://www.cabinetoffice.gov.uk/media/cabinetoffice/strategy/assets/pr2.pdf>
- Cabinet Office UK, Behavioral Insights Team. (2011). *Applying behavioural insight to health*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60524/403936_BehaviouralInsight_acc.pdf
- Cahill, K., Hartmann-Boyce, J., & Perera, R. (2015). Incentives for smoking cessation. *Cochrane Database of Systematic Reviews* (5), Art. No.: CD004307.
- Cahill, K., & Perera, R. (2011). Competitions and incentives for smoking cessation. *Cochrane Database of Systematic Reviews* (4), Art. No.: CD004307. doi:004310.001002/14651858.CD14004307.pub14651854
- Cartwright, N. (2012). Will this policy work for you? Predicting effectiveness better: How philosophy helps. *Philosophy of Science*, 79, 973–989.
- Cartwright, N., & Hardie, J. (2012). *Evidence-Based policy: A practical guide to doing it better*. New York, NY: Oxford University Press.
- Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5), 1–33. doi:10.1257/aer.p20151108
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, 33, 339–360.
- Commission on Social Determinants of Health (CSDH) World Health Organization. (2008). *Closing the gap in a generation: health equity through action on the social determinants of health. Final report*. Retrieved from http://apps.who.int/iris/bitstream/handle/10665/43943/9789241563703_eng.pdf;jsessionid=BA3A7DB491AE0306866B416371749C6?sequence=1
- Congdon, W. J., Kling, J. R., & Mullainathan, S. (2011). *Policy and choice: Public finance through the lens of behavioral economics*. Washington, DC: Brookings Institution Press.
- Department of Health UK. (2010). *Healthy lives, healthy people: Our strategy for public health in England*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216096/dh_127424.pdf
- Department of Health UK. (2011). *Healthy lives healthy people: Tobacco control plan for England*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213757/dh_124960.pdf
- Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). *MINDSPACE: influencing behavior for public policy: Institute of Government*. <http://www.instituteforgovernment.org.uk/sites/default/files/publications/MINDSPACE.pdf>
- Elster, J., & Skog, O.-J. (1999). Getting Hooked: Rationality and Addiction. In *Getting hooked: Rationality and addiction*. Cambridge University Press.
- Esping-Andersen, G. (1990/2013). *The three worlds of welfare capitalism*. Cambridge: Polity Press.
- Giné, X., Karlan, D., & Zinman, J. (2010). Put your money where your butt is: A commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4), 213–235. doi:10.1257/app.2.4.213
- Graham, H. (2011). *Understanding health inequalities*. Maidenhead: Open University Press.
- Graham, H., Sowden, A., Flemming, K., Heirs, M., & Fox, D. (2012). *Public Health Research Consortium: Using qualitative research to inform interventions to reduce smoking in pregnancy in England: a systematic review of qualitative studies*. Retrieved from http://phrc.lshtm.ac.uk/project_2005–2011_a810.html
- Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, 32(03), 463–483. doi:10.1017/s0266267115000425
- Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). Toward a Framework for Selecting Behavioural policies: How to Choose between boosts and nudges. *Economics and Philosophy*, 34(2), 43–66.
- Guala, F. (2005). *The methodology of experimental economics*. Cambridge: Cambridge University Press.
- Halpern, S. D., French, B., Small, D. S., Saulsgiver, K., Harhay, M. O., Audrain-McGovern, J., ... Volpp, K. G. (2015). Randomized trial of four financial-incentive programs for smoking cessation. *N. Engl. J. Med*, 372, 2108–2117.
- Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, 18(1), 123–136. doi:10.1111/j.1467-9760.2009.00351.x

- Haynes, L., Goldacre, B., & Torgerson, D. (2012). *Test, learn, adapt: developing public policy with randomised controlled trials*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/TLA-1906126.pdf
- Heilmann, C. (2014). Success conditions for nudges: A methodological critique of libertarian paternalism. *European Journal for Philosophy of Science*, 4(1), 75–94.
- Hortwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973–986.
- Higgins, S. T., & Solomon, L. J. (2016). Some recent developments on financial incentives for smoking cessation among pregnant and newly postpartum women. *Current Addiction Reports*, 3(1), 9–18.
- Higgins, S. T., Washio, Y., Heil, S. H., Solomon, L. J., Gaalema, D. E., Higgins, T. M., & Bernstein, I. M. (2012). Financial incentives for smoking cessation among pregnant and newly postpartum women. *Preventive Medicine*, 55), doi:10.1016/j.ypmed.2011.12.016
- Honjo, K. (2004). Social epidemiology: Definition, history, and research examples. *Environmental Health and Preventive Medicine*, 9(5), 193–199. doi.org/10.1007/BF02898100
- Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo–williamson thesis. *International Studies in the Philosophy of Science*, 25, 139–157.
- Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal of Philosophy of Science*, 2, 119.
- Jochelson, K. (2007). *Paying the patient: improving health using financial incentives*. King's Fund. Retrieved from www.wpro.who.int/NR/rdonlyres/BCC2F5C3-B685-4FF1AE7739C52F4ED247/0/payingthepatient.pdf
- Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kelly, M. P. (2010). The axes of social differentiation and the evidence base on health equity. *Journal of the Royal Society of Medicine*, 103(7), 266–272. doi:10.1258/jrsm.2010.100005
- Kelly, M. P. (2016). The politics of behaviour change. In F. Spotswood (Ed.), *Beyond behaviour change: Key issues, interdisciplinary approaches and future directions* (pp. 11–26). Bristol: Policy Press.
- Loewenstein, G., Asch, D. A., Friedman, J. Y., Melichar, L. A., & Volpp, K. G. (2012). Can behavioural economics make us healthier? *BMJ*, 344(may23 1), e3482–e3482. doi:10.1136/bmj.e3482
- Loewenstein, G., Brennan, T., & Volpp, K. G. (2007). Asymmetric paternalism to improve health behaviors. *JAMA*, 298, 2415–2417.
- Loewenstein, G., & Chater, N. (2017). Putting nudges in perspective. *Behavioural Public Policy*, 1(01), 26–53.
- Lorenc, T., Petticrew, M., Welch, V., & Tugwell, P. (2013). What types of interventions generate inequalities? Evidence from systematic reviews: Table 1. *Journal of Epidemiology and Community Health*, 67(2), 190–193. doi:10.1136/jech-2012-201257
- Marteau, T. M., & Mantzari, E. (2015). Public health: The case for pay to quit. *Nature*, 523(7558), 40–41.
- Matjasko, J. L., Cawley, J. H., Baker-Goering, M. M., & Yokum, D. V. (2016). Applying behavioral economics to public health policy. *American Journal of Preventive Medicine*, 50(5), doi:10.1016/j.amepre.2016.02.007
- Morgan, H., Hoddinott, P., Thomson, G., Crossland, N., Farrar, S., Yi, D., ... Hislop, J. (2015). Benefits of incentives for breastfeeding and smoking cessation in pregnancy (BIBS): a mixed-methods study to inform trial design. *Health Technology Assessment*, 19(30), 1–522.
- National Health Service (NHS). (2017). *Statistics on Smoking, England 2017*. Retrieved from <https://www.gov.uk/government/statistics/statistics-on-smoking-england-2017>
- NICE (The UK National Institute for Health and Excellence). (2007). *Evidence for the effect on inequalities in health of interventions designed to change behaviour*. Retrieved from <https://www.nice.org.uk/guidance/ph6/documents/evidence-for-the-effect-on-inequalitiesdesigned-to-change-behaviour2>
- NICE (The UK National Institute for Health and Excellence). (2013). *Stop smoking services*. Retrieved from <https://www.nice.org.uk/guidance/ph10/chapter/1-Key-priorities>
- O'donoghue, T., & Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making*, 13 (2), 233–250. doi:10.1002/(sici)1099-0771(200004/06)13:2<233::aid-bdm325>3.0.co;2-u
- O'Neill, J., Tabish, H., Welch, V., Petticrew, M., Pottie, K., Clarke, M., ... Evans, T. (2014). Applying an equity lens to interventions: Using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of Clinical Epidemiology*, 67(1), 56–64. doi:10.1016/j.jclinepi.2013.08.005
- Ormston, R., van der Pol, M., Ludbrook, A., McConville, S., & Amos, A. (2015). quit4u: The effectiveness of combining behavioural support, pharmacotherapy and financial incentives to support smoking cessation. *Health Education Research*, 30(1), 121–133. doi:10.1093/her/cyu024
- Petkovic, J., Welch, V., Jull, J., Petticrew, M., Kristjansson, E., Rader, T., ... Tugwell, P. (2017). How health equity is reported and analyzed in randomized trials. *Cochrane Database of Systematic Reviews* 2017 (8). Art. No.: MR000046.
- Petticrew, M. (2004). Evidence for public health policy on inequalities: 1: The reality according to policymakers. *Journal of Epidemiology & Community Health*, 58(10), 811–816. doi:10.1136/jech.2003.015289
- Popay, J. (2008). Should disadvantaged people be paid to take care of their health? No. *British Medical Journal*, 337(Jul08 3), a594. doi:10.1136/bmj.a594

- Radley, A., Ballard, P., Eadie, D., MacAskill, S., Donnelly, L., & Tappin, D. (2013). Give it up for baby: Outcomes and factors influencing uptake of a pilot smoking cessation incentive scheme for pregnant women. *BMC Public Health*, 13, 343.
- Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19, 171–197.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2), 157–170. doi:10.1080/02698590701498084
- Schmidt, A. T. (2016). Withdrawing versus withholding freedoms: Nudging and the case of tobacco control. *American Journal of Bioethics*, 16(7), 3–14.
- Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford: Oxford University Press.
- Strickland, S. (2014). Does it work to pay people to live healthier lives? *British Medical Journal*, 348, g2458.
- Sunstein, C. R. (2015). Nudging smokers. *New England Journal of Medicine*, 372(22), 2150–2151.
- Thaler, R. H., & Sunstein, C. (2008). *Nudge*. New Haven, Conn.: Yale.
- Thaler, R. H., Tversky, A., & Kahneman, D. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *Quarterly Journal of Economics*, 112, 647–661.
- Thomson, G., Morgan, H., Crossland, N., Bauld, L., Dykes, F., & Hoddinott, P. (2014). Unintended consequences of incentive provision for Behaviour change and Maintenance around Childbirth. *PLoS ONE*, 9(10), doi:10.1371/journal.pone.0111322
- Tugwell, P., Petticrew, M., Kristjansson, E., Welch, V., Ueffing, E., Waters, E., ... Kelly, M. P. (2010). Assessing equity in systematic reviews: Realising the recommendations of the commission on social determinants of health. *British Medical Journal*, 341(sep13-1), c4739–c4739.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185 (4157), 1124–1131.
- Verdurmen, J., Monshouwer, K., & Van Laar, M. (2015). *Factsheet Continu Onderzoek Rookgewoonten 2014*. Utrecht: Trimbos-instituut.
- Volpp, K. G., Asch, D. A., Galvin, R., & Loewenstein, G. (2011). Redesigning Employee health incentives — Lessons from behavioral economics. *New England Journal of Medicine*, 365(5), 388–390. doi:10.1056/nejmp1105966
- Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., & Asch, D. A. (2009). A randomized controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*, 360, 699–709.
- Weber, B. J., & Chapman, G. B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles? *Organ Behav Hum*, 97, 31–46.
- Welch, V., Petticrew, M., Petkovic, J., Moher, D., Waters, E., White, H., ... the PRISMA-Equity Bellagio group. (2015). Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration. *International Journal for Equity in Health*, 14, 92. doi:10.1186/s12939-015-0219-2
- Welch, V., Tugwell, P., Petticrew, M., de Montigny, J., Ueffing, E., Kristjansson, B., ... Smylie, J. (2010). How effects on health equity are assessed in systematic reviews of interventions. *Cochrane Database of Systematic Reviews*, (12), Art. No.: MR000028. doi:10.1002/14651858.MR000028.pub2
- Whitehead, M. (1992). The concepts and principles of equity and health. *International Journal Health Service*, 22(3), 429e45.
- World Health Organization (WHO). (2011). *Global status report on noncommunicable diseases 2010*. Retrieved from http://www.who.int/nmh/publications/ncd_report2010/en/
- World Health Organization (WHO). (2012). *Health impact assessment, glossary of terms*. Retrieved from <http://www.who.int/hia/about/glos/en/index1.html>