Behavioural Public Policy (2022), 6: 2, 172–190
© *Cambridge University Press* doi:10.1017/bpp.2018.37 First published online 18 October 2018

Informing behavioural policies with data from everyday life

LEONHARD K. LADES*
EnvEcon, Blackrock, Co. Dublin, Ireland
and
University College Dublin, Dublin 4, Ireland
LUCIE MARTIN
Behavioural Insights Team, London, UK
LIAM DELANEY
University College Dublin, Dublin 4, Ireland
and
Stirling Management School, University of Stirling, Stirling, UK

Abstract: Naturalistic monitoring tools provide detailed information about people's behaviours and experiences in everyday life. Most naturalistic monitoring research has focused on measuring subjective well-being. This paper discusses how naturalistic monitoring can inform behavioural public policy-making by providing detailed information about everyday decisions and the choice architecture in which these decisions are made. We describe how the Day Reconstruction Method (DRM) – a naturalistic monitoring tool popular in the subjective well-being literature – can be used to: (i) improve ecological validity of behavioural economics; (ii) provide mechanistic evidence of the everyday workings of behavioural interventions; and (iii) help us to better understand people's true preferences. We believe that DRM data on everyday life have great potential to support the design and evaluation of behavioural policies.

Submitted 20 February 2018; revised 17 July 2018; accepted 14 September 2018

Introduction

Behavioural economic findings have not only influenced economic theory (Rabin, 1998; Chetty, 2015; Thaler, 2016), but also public policy-making (Thaler & Sunstein, 2008; Shafir, 2013; Sunstein, 2016b; Oliver, 2017). However, at least three methodological shortcomings have been identified

* Correspondence to: EnvEcon, 11 Priory Office Park, Blackrock, Co. Dublin, Ireland. Email: leonhard.lades@ucd.ie

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

that have the potential to reduce the effectiveness and ethical legitimacy of behavioural public policies. First, not much is known about the generalizability of behavioural findings from the laboratory to the real world where public policies take effect (Levitt & List, 2007; Gneezy & Imas, 2017; Galizzi & Navarro-Martinez, 2018). Second, much of the existing behavioural public policy literature has focused on identifying 'what works' and less on investigating 'why' behavioural interventions work. Mechanistic evidence about the 'why', however, is important to ensure that policies are effective, robust, persistent and welfare-improving in their target environments (Harrison, 2014; Grüne-Yanoff, 2015). Finally, the field has not agreed upon a way to identify true preferences when decision-making can be biased and thus lacks a welfare standard to evaluate behavioural public policies (Beshears *et al.*, 2008; Hausman, 2012; Infante *et al.*, 2016; Sugden, 2017).

This paper discusses how naturalistic monitoring of people's everyday decision-making biases can help to overcome these shortcomings. Naturalistic monitoring (or ecological momentary assessment) describes the observation of people's behaviours and experiences 'in the wild' (i.e., in people's natural environments where most of their economic decision-making takes place) (Shiffman *et al.*, 2008). It includes self-reported measures, but also more objective observations, for example of individuals' psychobiology (such as heart rate and skin conductance) and GPS data (Daly *et al.*, 2014). One of the most popular naturalistic monitoring tools is the Day Reconstruction Method (DRM). The DRM was developed by Kahneman *et al.* (2004a) as a cost-effective way to measure both how people allocate their time and how they feel in their everyday lives. So far, the use of the DRM has been limited to measuring the determinants and consequences of subjective well-being in everyday life.

This paper discusses how the DRM can improve behavioural public policy-making by providing data about when, how and why decision-making biases occur in everyday life. The DRM, and naturalistic monitoring more generally, can be a helpful complement to existing methods of obtaining data for behavioural policy design and evaluation. It can complement laboratory experiments by identifying the relevance of decision-making biases in the real world, as well as identifying situational factors that influence the extent to which decision-making is biased in everyday life. The DRM can complement field experiments and randomized controlled trials (RCTs) by providing information about the mechanisms that determine why a given policy intervention is effective in everyday life or not. This mechanistic evidence is essential to ensuring that the intervention will also work in other contexts. The DRM can also provide a way to identify what people want in their everyday lives (or their 'subjective

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

preferences'), which might be used as a standard to assess the welfare implications of behavioural policies.

While the DRM has shortcomings, which we discuss below, we see its biggest promise in its capacity to provide information about the high-frequency decisions that specific sub-populations make in their everyday lives. While common surveys and experiments are well-suited to analysing low-frequency decisions made with substantial deliberation, the DRM provides a window into the otherwise difficult to observe behavioural patterns of everyday life and their determinants. Some of these determinants are not under the influence of individuals or policy-makers (e.g., time of the day, weather, etc.), but others are (e.g., congestion, density of fast food restaurants, medication, work email policies, etc.). Due to its scalability, the DRM has the capacity to distinguish between different sub-populations and thus allows for the identification of situational factors that induce detrimental decision-making amongst the heterogeneous population (e.g., in terms of economic status, family status, income, education, financial literacy, geographical location, personality traits and economic preferences, etc.). Such a fine-grained analysis in terms of both the situational context determinants of behaviour and the different sub-populations is very suitable to supporting the discovery of actionable behavioural policy interventions. The DRM can thus complement the growing literature that uses behavioural findings to change the choice architecture of everyday life in order to encourage welfare-enhancing decision-making (Thaler & Sunstein, 2008).

Naturalistic monitoring and the DRM

Naturalistic monitoring describes the observation of people's behaviours and experiences 'in the wild' (i.e. in people's natural environments) (Shiffman *et al.*, 2008). The DRM is one of the most popular naturalistic monitoring tools. It was developed by Kahneman *et al.* (2004a) and combines features of experience sampling, time-budget elicitation and classical survey questions. The DRM aims to capture an entire day's experience while curtailing opportunities for recall biases, limiting participant burden and providing a cost-effective alternative to more burdensome experience sampling methodologies (Sonnenberg *et al.*, 2012; Diener & Tay, 2014). The DRM is widely used to measure subjective well-being in everyday life and has been applied both in large representative population surveys such as the American Time Use Survey and the German Socio-Economic Panel (Krueger *et al.*, 2009; Anusic *et al.*, 2017) and in smaller-scale studies focused on specific topics (Srivastava *et al.*, 2008; Daly *et al.*, 2010; Knabe *et al.*, 2010; Bakker *et al.*, 2013; Daly *et al.*, 2014; Ishio & Abe, 2017; Lee *et al.*, 2017).

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

In a typical DRM study, participants are first asked to complete a personal diary in which they divide their previous day into 'episodes', as if each episode were a scene in a movie. Participants are asked to reflect upon what they did and how they felt during each episode. In the second phase, participants complete a survey in which they are asked questions about each episode chronologically. These questions can address any themes the researchers are interested in. Kahneman *et al.* (2004a) asked about the episodes' start and end times, the type of activity participants were engaged in (e.g., commuting to work, having a meal, exercising), where they were and the emotional states they were in (e.g., happiness, boredom, hunger). In such a setup, the DRM can elicit whether particular subjective experiences are correlated with situational aspects in everyday life (for more detailed descriptions of the DRM, see Kahneman *et al.*, 2004a, 2004b; National Research Council (US), 2012; Diener & Tay, 2014).

The most prominent alternative to the DRM is the Experience Sampling Method (ESM). The ESM is a real-time data capture tool developed by Larson and Csikszentmihalyi (1983) in which participants are prompted at random intervals during the day through an electronic device (today it would be a mobile phone; Hofmann & Patel, 2015) to record what they are doing and feeling at that moment. The ESM is sometimes considered the gold standard of naturalistic monitoring as it circumvents memory biases and allows for the elicitation of additional objective data from the smartphone. However, ESM studies are relatively expensive and burdensome and can suffer from low response rates, which makes it difficult to obtain large samples and to attach the studies to ongoing surveys (National Research Council (US), 2012). DRM studies are comparably easy to conduct and can be managed by most researchers without the help of technicians (Anusic *et al.*, 2017). Moreover, due to the ESM's invasive nature, study participants might become conscious of their actions throughout the day, which could change their behaviour. Finally, the ESM and the DRM provide conceptually similar data (Kahneman *et al.*, 2004a; Dockray *et al.*, 2010; Sonnenberg *et al.*, 2012). While this paper makes a general case for informing behavioural policy-making with naturalistic monitoring techniques, the DRM is arguably the most suitable method due to its cost-effectiveness and scalability, especially when analysing large samples.

Using naturalistic monitoring to inform behavioural public policy

The DRM is frequently used to measure subjective well-being in daily life, and a number of researchers have suggested using subjective well-being data to refocus or inform policy decisions (Kahneman & Sugden, 2005). More

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

generally, subjective well-being research has emerged as a growing literature (Frey & Stutzer, 2002; Blanchflower & Oswald, 2004), and the case has been made to use affective, cognitive and eudaimonic forms of subjective well-being as alternative, non-monetary objectives for societal improvement (Layard, 2006; Dolan & White, 2007; Stiglitz *et al.*, 2010; Odermatt & Stutzer, 2017). This section discusses three reasons why we consider it useful for behavioural public policy-making to broaden the analysis of everyday life beyond subjective well-being and also to monitor decision-making biases using naturalistic monitoring tools such as the DRM.

Improving ecological validity

Behavioural economics is the study of how real-world 'Humans' (rather than the 'Econs' from most economics textbooks) make economic decisions (Thaler & Sunstein, 2008; Dhami, 2016). Despite behavioural economics' emphasis on the real world, however, the field is built on findings from laboratory experiments and surveys that are often abstract, that situate participants in artificial contexts and that often rely on student samples. Recognizing a gap between measurement in artificial contexts and the real world, Levitt and List (2007) argue that "perhaps the most fundamental question in experimental economics is whether findings from the lab are likely to provide reliable inferences outside of the laboratory" (p. 170). The related "fundamental" question about the extent to which people make the same mistakes in the real-world as those found in the laboratory has not yet been answered systematically. Recent studies do not paint an overly optimistic picture of the generalizability of behavioural laboratory findings, as they suggest in many cases that there are no significant associations between economic preferences measured in the laboratory and theoretically related behaviour outside the laboratory (Delaney & Lades, 2017; Gneezy & Imas, 2017; Galizzi & Navarro-Martinez, 2018). The relevance of behavioural economic laboratory experiments to help us better understand decision-making in everyday life may well be limited in the sense that experimental studies and their results do not tell us much about the real world. A range of behavioural biases identified in behavioural economics may characterize laboratory but not real-world decision-making.

Psychologists have long been concerned about how to generalize behaviour from experimental settings to the real world (e.g., Loewenstein, 1999; Kaplan & Stone, 2013). A common strategy in psychology to overcome problems of low ecological validity is to design experimental stimuli that are good representations of naturally occurring environments (Brunswik, 1956). But most experimental economic studies, on which many behavioural public policies

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

rely, cannot (and do not aim to) reflect contextual factors that affect decision-making in everyday life. On the contrary, most economic experiments are designed to represent context-free representations of the payoff structures in real-world situations. While economic experiments have many advantages (e.g., the ability to control other variables, ease of replication, high internal validity and reliability, focus on clean causal effects), they abstract from realistic frames, put participants into unfamiliar roles in unfamiliar contexts that do not reflect real-world situations and encourage reflective decision-making rather than the automatic decision-making that is more common in the real world.

In order to increase the relevance of behavioural economics for a better understanding of how people make decisions in their real lives, studies can be conducted '*in vivo*' (i.e., in the real world where everyday decision-making takes place). By definition, naturalistic measurement therefore does not share the ecological validity problems that characterize some laboratory experiments. Most importantly for behavioural scientists, naturalistic monitoring allows us to make inferences about how, when and where decision-making biases occur in real life. Providing information about the extent to which decision-making biases are relevant in everyday life is key for the cost–benefit analysis that should inform any policy or regulatory intervention (Sunstein, 2016a). As such, naturalistic monitoring can complement (not substitute) experimental studies by providing data on the real-world relevance of decision-making biases that were previously identified in highly controlled laboratory experiments.

For example, a recent stream of naturalistic monitoring research has provided novel findings on self-control (Hofmann *et al.*, 2012; Delaney & Lades, 2017; Milyavskaya & Inzlicht, 2017; Wilkowski *et al.*, 2018), which is historically among the most prominent topics in the behavioural sciences (Elster, 1979; Hoch & Loewenstein, 1991; Thaler, 2018). In their seminal study on self-control in everyday life, Hofmann *et al.* (2012) used the ESM to provide a detailed picture of everyday desires and self-control failures. Among other results, they found that desires are frequent, variable in intensity and mostly unproblematic in everyday life. However, they also found a non-trivial amount of self-control failures. Their participants enacted 17% of the desires despite resistance attempts. The study by Delaney and Lades (2017) used the DRM to replicate most of the findings presented by Hofmann *et al.* (2012) and showed that participants enact more than 30% of the desires despite resistance attempts. This study also showed that the most typical behavioural economic measure of self-control, namely present bias as measured using a financial inter-temporal choice task, is not significantly correlated with any aspect of self-control in everyday life. These findings show both that self-control failures are indeed prevalent in everyday life and that typical

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

experimental measures of self-control in behavioural economics seem to measure another phenomenon entirely.

Obtaining mechanistic evidence

At the core of behavioural economics' relevance for policy-making is the suggestion of changing the choice architecture in addition to (or even rather than) educating individuals to change behaviour and intervening with harder regulation or mandates (Thaler & Sunstein, 2008). For example, presenting healthy food first in cafeterias and simplifying forms to nudge individuals to behave differently can complement and substitute for awareness campaigns that aim to encourage healthy eating and filling forms correctly. Data on the effectiveness of such nudges often come from RCTs that are conducted with high ecological validity in the relevant real-world contexts, often in collaboration with businesses and/or policy-makers (Harrison & List, 2004; Halpern, 2015; Duflo, 2017; Gneezy & Imas, 2017). RCTs are often considered the gold standard if the aim is to identify the effect of an intervention in a given context, and an impressive body of literature has now emerged examining the causal impact on behaviour of changing aspects of the choice environment that, from a neoclassical perspective, should not impact on people's decisions.

However, several scholars have pointed out that the treatment effects from RCTs themselves cannot provide a basis for developing theoretical accounts of decision-making in real-world economic environments. RCTs have also been criticized for providing information that is not necessarily generalizable to other contexts, for being limited to evaluations of observables and average effects and for not providing data on the latent welfare consequences of interventions (Harrison, 2014; Grüne-Yanoff, 2015; Deaton & Cartwright, 2018). In the context of this paper, most importantly, RCTs do not explore the mechanisms that explain *why* interventions work, but instead focus on 'what works'. These mechanisms are typically assumed relying on theory inspired by experimental findings from laboratory environments.

The rich data that naturalistic monitoring provides allow for the testing of detailed mechanisms of behavioural change. Similar to Harrison (2014), who suggested complementing field studies with experimental studies on risk attitudes, subjective risk perception and time preferences, we suggest that it is of value to complement RCTs with naturalistic monitoring data. The main advantage of naturalistic monitoring for the evaluation of behaviourally informed policies is that we can directly measure the choice architecture, as well as its influence on real-world decision-making and everyday behaviour. While the DRM has not yet been used to evaluate treatment effects on behavioural change, it has been used to evaluate the effects of an early intervention

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

policy on the subjective well-being of mothers (Doyle *et al.*, 2017), which shows that it is possible to integrate a DRM element into RCTs. By measuring the effects of different variations of the choice architecture on decision-making patterns and the outcomes of these decisions, we will better understand the mechanisms that can explain *why* an intervention is effective or not. This will help us to better understand *how* choice architecture informs individual decision-making in daily life. As such. RCTs augmented by naturalistic monitoring can help address one of the key methodological weaknesses of behaviourally informed policies by supporting the design of policies that are 'effective, robust, persistent or welfare-improving' in their target environments (Grüne-Yanoff, 2015).

For example, the recent naturalistic monitoring studies on everyday desires and self-control (e.g., Hofmann *et al.*, 2012; Delaney & Lades, 2017) provide mechanistic evidence about when, where, whether and why self-control failures in everyday life occur. These studies show that it is possible to identify different decision-making processes that can lead to similar behaviours. For example, eating a delicious but unhealthy snack might be the result of a self-control failure (e.g., when the person is on a diet) or might be in line with higher-order goals (e.g., when the snack is considered a reward or a treat). By asking whether the person attempted to resist eating the snack, researchers are able to differentiate between these decision-making processes leading to the same outcome.[1] Naturalistic monitoring studies can also identify individual differences and their links to everyday behaviours. Hofmann *et al.* (2012) showed that personality is a relatively strong predictor of desire strength and conflict strength in everyday life, while situational factors, such as alcohol consumption, are stronger predictors of attempts to resist and eventually enacting desired behaviours. This suggests that behavioural interventions modifying the choice architecture are most effective when attempting to influence later stages in the decision-making process from desire to enactment. Finally, the findings from Delaney and Lades

1 Using resistance attempts as the defining criterion to indicate whether a behaviour is a self-control failure or not assumes the narrow definition of self-control as a process of effortful inhibition of impulses. It is up for debate whether more preventive or proactive strategies that help to avoid impulses (de Ridder *et al.*, 2012; Duckworth *et al.*, 2016) should be captured by the notion of self-control, self-regulation or something else (Fujita, 2011). For example, employing the narrow definition of reactive self-control means that 'resigned addicts' who have progressed so far in their addiction that they have given up resistance to the desire to consume the substance do not experience self-control failures. Also, 'unapologetic hooligans' who enact impulsive aggression without experiencing any desire–goal conflict and hence experience no resistance attempts are not experiencing self-control failures in the narrow definition. Thanks are given to an anonymous reviewer for providing these examples.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

(2017) suggest that everyday self-control failures are more likely to be due to visceral influences rather than the decreasing impatience that much of the inter-temporal choice literature suggests as the most likely mechanism for self-control failures.

Identifying true preferences

Arguably the most important implication of behavioural economics for welfare economics is that choices do not necessarily reveal true preferences if these choices are based on cognitive biases. Since Pareto's (1971) argument for the removal of psychological considerations from economic theory and Samuelson's (1938) development of revealed preference theory, economic analysis has been focused on choice and on other observable conditions. If people choose good A when good B is available, good A is 'revealed preferred' to B. Policies that provide people with A are assumed to be better for individual welfare than policies that provide B. However, if people do not always decide rationally and with perfect willpower (e.g., when the decision to choose A is the result of a self-control failure), choices do not necessarily reveal true preferences, and policies might fail to maximize individual welfare as they recover preferences from biased or weak-willed choices. And even under the assumption that people make rational and controlled decisions, choice alone, in the absence of knowledge about individual beliefs, cannot reveal preference (Hausman, 2000). Thus, new ways of identifying individual preferences, or other appropriate welfare measures, are needed in order to evaluate policy interventions (Beshears *et al.*, 2008; Chetty, 2015).

Some soft paternalists argue that the achievement of generally held higher-order goals (e.g., in terms of health, wealth and well-being) is a good welfare standard, as these higher-order goals represent what individuals want when not influenced by bounded rationality and limited willpower (Thaler & Sunstein, 2008). Others are sceptical as to whether policy-makers can identify whether people have these higher-order preferences (Rizzo & Whitman, 2009). Strategies have been devised to identify preferences when decisions can be biased. For example, Beshears *et al.* (2008) provide several strategies to align revealed and normative preferences, Hausman's (2012) 'preference purification' approach suggests reconstructing preferences by controlling for bias and Bernheim (2016) argues for using only the subset of decisions that is clearly unbiased when identifying preferences. All of these approaches have in common that they require some information on the behavioural mechanisms that underlie potentially biased decision-making.

As mentioned in the previous subsection, naturalistic monitoring can provide such information on behavioural mechanisms. For welfare analysis, it is

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

particularly useful to differentiate between biased and unbiased decision-making, as well as between dynamically consistent and dynamically inconsistent decision-making (e.g., to identify whether a behaviour is or is not the result of a self-control failure). Following Bernheim (2016), one could use naturalistic monitoring techniques in order to identify those behaviours that are not driven by cognitive biases and self-control failures and use only those unbiased and dynamically consistent choices to recover true preferences. Similarly, naturalistic monitoring can help test whether people have stable and well-defined preferences – conditional on context. By measuring decisions and associated contexts repeatedly, clear patterns might emerge that allow the predicting of future choices in similar contexts and that provide a basis for recovering stable underlying preferences. Finally, by using naturalistic monitoring techniques, one can elicit the subjective beliefs that individuals have at the moment of making a decision. This approach might overcome the problem posed by Hausman (2000), who argues that choices alone cannot reveal beliefs nor preferences, because any choice is consistent with any preference, given the right set of beliefs.[2]

Naturalistic monitoring also allows for a direct measurement of what people want in their everyday lives and whether they get it. For example, Hofmann *et al.* (2012) and Delaney and Lades (2017) provide detailed accounts of everyday desires and their satisfaction. Preference satisfaction approaches have historically been used by many economists to guide welfare evaluations, and data from everyday life can inform us as to whether people satisfy their short-term preferences or not. This approach of measuring subjective preferences directly is analogous to previous research that measures experiential utility in everyday life (Kahneman *et al.*, 2004a). But rather than measuring how people feel, these new approaches measure what people want (i.e., their 'wantability'; Fisher, 1918) in everyday life. Such direct measures of preference and of their satisfaction could complement indirect preference measures that rely on choice data in order to reveal preferences. If the satisfaction of short-term desires is taken as the guide to welfare, a policy that gives people what they want at any given moment would be preferred over a policy that restricts choice. Similarly, different desires could be ranked according to their normative weight in the sense of a hierarchy of needs, as suggested already by Maslow (1943) and discussed in Witt (2017). A benevolent dictator might come up with a list of short-term desires that policies should encourage and with another list of desires to discourage. We do not claim that these would be the normatively best evaluation criteria, but merely state that naturalistic monitoring allows for quantifying preferences and their satisfaction in everyday life. Whatever welfare criterion

2 We thank an anonymous reviewer for this suggestion.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

related to everyday life one favours, naturalistic monitoring can provide data to assess the success of a policy based on this welfare criterion.[3]

Finally, the most common way to identify preferences using naturalistic monitoring techniques is to decompose and compare different components of subjective well-being. Based on the distinction between life satisfaction as the evaluate component of subjective well-being and momentary happiness as its experiential component, Knabe *et al.* (2010) used the DRM to show that unemployed people are dissatisfied with life, but have a good day in terms of momentary happiness. They explain the relatively high momentary happiness of the unemployed by the lack of time spent at work, which is one of the least enjoyable activities. In a follow-up study, Knabe *et al.* (2017) show that workfare participants' life satisfaction is between that of employed and unemployed people and that workfare participants' emotional well-being is the highest of these three groups.

Assessing the DRM as a tool for behavioural public policy: methodological considerations

There are several situations in which the DRM can be used in public policy contexts. However, the benefits and costs of using the DRM, and naturalistic monitoring more generally, depend on a number of methodological considerations. This section discusses the opportunities and limitations of using the DRM with the purpose of informing behavioural public policies.

Methodological options and opportunities

The aim of the seminal DRM study by Kahneman *et al.* (2004a) was to measure experiences, settings, activities and time allocation in peoples' daily lives. The study's focus was on presenting affect ratings and their correlations with contemporaneous situational factors such as activities, interaction partners and time of the day. The study design, however, is versatile and allows for various modifications, depending on the specific research question.[4] For example, in Kahneman *et al.* (2004a), participants completed the survey on a paper-and-pencil basis. But recent research has moved towards digital versions of the DRM, sometimes completely digital and sometimes with a paper diary to keep personal details anonymous (e.g., National Research Council (US), 2012; Bakker *et al.*, 2013; Delaney & Lades, 2017). Moreover, in most DRM studies,

3 We are currently working on a related manuscript that discusses these normative issues in more detail, focusing on the case of self-control failures as judged by individuals themselves.

4 We are currently preparing a methodology brief with more details on the method.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

participants are told that the typical length of an episode is between 15 minutes and 2 hours, but it is also possible to predefine certain time intervals, for example 2-hour intervals, and ask participants follow-up questions regarding this interval. The time intervals can be shortened and the absolute number of episodes reduced if the research is about rather short-lived aspects of everyday life, such as specific decisions and their psychological and situational correlates. The predefined time intervals also reduce individual heterogeneity in terms of the number of reported episodes (Diener & Tay, 2014). It is also possible to deviate from the episode as the unit of analysis according to which the second phase of the DRM is structured. The follow-up questions would then be concerned with one of the other questions asked in the first phase, such as activity, decision or social interaction. For example, in the American Time Use Survey (ATUS) well-being module, after documenting time use for the full previous day, respondents are asked to rate their affect during certain activities rather than episodes (National Research Council (US), 2012).

Abbreviated versions of the DRM have also been used to keep participant burden low. For example, in the German Socio-Economic Panel, participants are asked to complete the full first-phase diary, but in the second phase, follow-up questions are asked only about some randomly sampled episodes (Anusic *et al.*, 2017). Such abbreviated versions of the DRM provide results that are similar to more comprehensive versions (Miret *et al.*, 2012). It is also possible to ask participants to recall only a subset of episodes in the first phase (e.g., only the morning, afternoon or evening, or beginning with a randomly specified point in time yesterday) and to ask participants to answer follow-up questions for all recalled episodes. In the ATUS, the full day is reconstructed in terms of activities, and three activities are randomly selected for follow-up questions. The English Longitudinal Study of Ageing focuses on seven specific activities (watching TV, working or volunteering, walking or exercising, engaging in health-related activities other than walking/exercising, travelling or commuting, spending time with family or friends and spending time at home alone) and asks participants follow-up questions only for these. Some research has also used the same concept but has focused on specific events, creating so-called 'event reconstruction studies' (Grube *et al.*, 2008). For these studies, it is essential to invite only participants who had the event under investigation in the recent past.

Limitations of the DRM as a tool for decision-making research

While we have suggested that the DRM is a valuable tool for obtaining detailed information on decisions made in everyday life in order to inform behavioural public policy-making, it is important to be clear about the limitations of the method in these contexts. In particular, the DRM is not suited to investigating

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

possible factors behind low-frequency decisions. For example, when interested in whether a situational factor influences the decision to buy a car, which mortgage to get or whom to marry, DRM data will not be useful, as they will likely miss the right moment. But even when analysing high-frequency decisions, some limitations need to be acknowledged before using DRM data for policy purposes. Some of the limitations are the same as in other studies that rely on self-reports (e.g., the validity, reliability and sensitivity of measurement instruments, dishonest reporting, social desirability bias, norms, self-image considerations and reactivity to assessment procedures), but others are more DRM-specific. The National Research Council (US) (2012) discusses question-order effects, scale effects and survey-mode effects.

The most obvious DRM-specific limitation is that it requires a reliance on participants' memory. Memories are subject to a number of biases, such as a reliance on routine to infer yesterday's likely activities, the peak-end bias in memory and overstating of previous emotions and preferences (Diener & Tay, 2014). Moreover, the mood during the completion of the questionnaire can influence the revivification of yesterday and influence how participants recall their past day (Schwarz & Clore, 1983). Another important aspect to consider when looking at the reliability of participants' memories is the structure of these memories. People rely on mental scripts and episodes in their memory to retrace their steps – and they are comfortable with a certain level of detail (e.g., "I had breakfast," not "I had a meal" or "I had a continental breakfast") (Tourangeau *et al.*, 2000). Deviating from this preferred level of detail could mean a deterioration in the quality of the data being collected. Finally, individual differences in people's memories should also be considered – age and health, in particular, are likely to play a role. That said, the DRM was explicitly designed to minimize memory bias, and evidence from studies comparing the outcomes of the DRM with experience sampling measures in real time suggests that the DRM largely achieves its goals of assessing people's episodic feelings and experiences without being distorted by memory and other biases (Dockray *et al.*, 2010; Sonnenberg *et al.*, 2012; Kim *et al.*, 2013).

Another potential issue relates to the episode as the unit of measurement. A single episode might contain various feelings and decisions, and it is not known whether single responses can represent a full episode that might last up to several hours (Diener & Tay, 2014). Moreover, differential responding patterns of subgroups can lead to misleading conclusions about actual experiences. For example, if older people were more open to acknowledge a bias than younger people, or if men were more likely to underreport socially inappropriate behaviours than women, differences in DRM data would not reflect real differences in people's everyday lives. This would raise doubts about the comparability across subgroups of the population.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

Finally, naturalistic monitoring does not allow for much control over the study environment compared to laboratory studies or RCTs. This means that a DRM study might be hampered by more threats to internal validity: confounding factors affecting outcomes are more difficult to avoid, and this may threaten causal inferences and external validity (e.g., Jiménez-Buedo, 2011). Where spurious relationships cannot be ruled out, rival hypotheses to the original causal inference hypothesis of the researcher may be developed. However, this is precisely where the features of the DRM can shine, through follow-up questions about episodes that allow for in-depth probing in order to map out potential factors, identify causal mechanisms and identify alternative relationships.

Conclusion: evaluating the effects of behavioural policies in everyday life

Behavioural economics shows that people are boundedly rational, have limited willpower and do not always act in ways economics textbooks would suggest (Dhami, 2016). These deviations from rationality and dynamic consistency are often systematic and predictable, as shown repeatedly in laboratory experiments (Camerer *et al.*, 2004). Such findings have started to change economic theory (Rabin, 1998) and have substantially reformed policy-making worldwide, particularly in the UK, through the foundation of behavioural insights teams and nudge units (Jolls *et al.*, 1998; Thaler & Sunstein, 2008; Halpern, 2015). The evidence on which many of these behavioural interventions rely often comes from laboratory environments (which often put study participants into rather artificial decision situations) and RCTs (which provide information about what works, but not about the underlying mechanisms).

This paper discussed how a popular naturalistic measurement tool, the Day Reconstruction Method (DRM), can be used to inform behavioural public policies by providing mechanistic evidence on how people make decisions in the real world. We suggest that the DRM is a valuable addition to the behavioural scientist's toolbox and can complement ordinary surveys, observational data, laboratory experiments and RCTs. The key benefit of the DRM, which sets it apart from alternative approaches, is that it allows for measuring decision-making in naturalistic, everyday contexts. It is thus a method that helps to quantify the extent to which behavioural biases change our behaviour in the real world and to identify where, when and why decision-making biases occur. The DRM can show, for example, whether there are correlations between biases and simultaneous situational factors such as location, activity, social interaction partner and internal state. Measuring everyday contexts and their effects on decision-making can also help to design better behavioural policies that change the choice architecture in order to nudge people to make better decisions. Such behavioural public policy interventions should be

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

informed by domain-specific naturalistic monitoring studies in which detailed information about a particular type of phenomenon is elicited and where domain-specific context variables can be identified.

For future research, there are several potential applications of the DRM in behavioural science and behavioural public policy. The method can be used to measure the prevalence of almost any behavioural concept in every domain of life. It can measure, for example, in which real-life situations people are particularly risk or loss averse, or similarly examine the influence of everyday anchors, defaults and social norms or identities on everyday economic behaviour. The key challenge for these future studies is to design survey questions that are as similar as possible to the concepts usually identified in decision-making experiments. For several concepts (e.g., risk aversion), verbal survey questions that measure individual differences have already been designed (Weber *et al.*, 2002). Future research can adapt these questions to relate them to intra-individual changes that can differ across situations in everyday life. Such studies will then be able to quantify how prevalent behavioural biases are, and also explore in what contexts these biases are particularly likely to arise.

A key challenge for future research is to integrate DRM studies in causal designs. Since the DRM is a survey that can be completed in one sitting, it can be easily added to existing RCTs. DRM studies also lend themselves well to evaluating large policies where DRM data from before the policy implementation can be compared to DRM data gathered after the implementation. Another branch of future research should deal with methodological issues. For example, different versions of the DRM (full versus abbreviated, online versus analogue, one day versus multiple days, different reinstantiation procedures, etc.) should be compared in order to identify the effects that design choices have on participants' response patterns. It will also be important to further test the reliability, validity and accuracy of DRM data by comparing it with experience sampling data. Cognitive testing in interviews and focus groups should be conducted to make sure that the question wording used does not confuse the participants. Moreover, filling out the DRM itself can change behaviour, and the potential of the DRM as a behavioural intervention should be explored. If we better understand these methodological issues, DRM studies measuring behavioural concepts could be integrated into existing large-scale, nationally representative time use surveys.[5] This strategy would help us to gain a better understanding as to how individuals from different

5 In an ongoing project, we measure desires and self-control failures among a nationally representative sample of 955 individuals using a short (~10-minute) online version of the DRM, showing that large-scale DRM studies are feasible.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

sub-populations differ in terms of the decisions they make in their everyday lives.

Acknowledgements

Leonhard Lades has been supported by a grant from the Irish Environmental Protection Agency (project name: Enabling Transition; project number: 2017-CCRP-FS.32).

References

Anusic, I., R. E. Lucas and M. B. Donnellan (2017), 'The validity of the Day Reconstruction Method in the German Socio-economic Panel Study', *Social Indicators Research*, **130**(1): 213–232.

Bakker, A. B., E. Demerouti, W. Oerlemans and S. Sonnentag (2013), 'Workaholism and daily recovery: A day reconstruction study of leisure activities', *Journal of Organizational Behavior*, **34**(1): 87–107.

Bernheim, B. D. (2016), 'The good, the bad, and the ugly: a unified approach to behavioral welfare economics', *Journal of Benefit-Cost Analysis*, **7**(1): 12–68.

Beshears, J., J. J. Choi, D. Laibson and B. C. Madrian (2008), 'How are preferences revealed?' *Journal of Public Economics*, **92**(8): 1787–1794.

Blanchflower, D. G. and A. J. Oswald (2004), 'Well-being over time in Britain and the USA', *Journal of Public Economics*, **88**(7): 1359–1386.

Brunswik, E. (1956), *Perception and the Representative Design of Psychological Experiments*, Berkeley, CA: University of California Press.

Camerer, C., G. Loewenstein and M. Rabin (2004), *Advances in Behavioral Economics*, Princeton, NJ: Princeton University Press.

Chetty, R. (2015), 'Behavioral economics and public policy: a pragmatic perspective', *American Economic Review*, **105**(5): 1–33.

Daly, M., R. F. Baumeister, L. Delaney and M. MacLachlan (2014), 'Self-control and its relation to emotions and psychobiology: evidence from a Day Reconstruction Method study', *Journal of Behavioral Medicine*, **37**(1): 81–93.

Daly, M., L. Delaney, P. P. Doran, C. Harmon and M. MacLachlan (2010), 'Naturalistic monitoring of the affect-heart rate relationship: a day reconstruction study', *Health Psychology*, **29**(2): 186–195.

Deaton, A. and N. Cartwright (2018), 'Understanding and misunderstanding randomized controlled trials', *Social Science & Medicine*, **210**, 2–21.

Delaney, L. and L. K. Lades (2017), 'Present bias and everyday self-control failures: a Day Reconstruction Study', *Journal of Behavioral Decision Making*, **30**(5): 1157–1167.

de Ridder, D. T. D., G. Lensvelt-Mulders, C. Finkenauer, F. M. Stok and R. F. Baumeister (2012), 'Taking stock of self-control: a meta-analysis of how trait self-control relates to a wide range of behaviors', *Personality and Social Psychology Review*, **16**(1): 76–99.

Dhami, S. (2016), *The Foundations of Behavioral Economic Analysis*, Oxford, UK: Oxford University Press.

Diener, E. and L. Tay (2014), 'Review of the day reconstruction method (DRM)', *Social Indicators Research*, **116**(1): 255–267.

Dockray, S., N. Grant, A. A. Stone, D. Kahneman, J. Wardle and A. Steptoe (2010), 'A comparison of affect ratings obtained with ecological momentary assessment and the Day Reconstruction Method', *Social Indicators Research*, **99**(2): 269–283.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

Dolan, P. and M. P. White (2007), 'How can measures of subjective well-being be used to inform public policy?' *Perspectives on Psychological Science*, **2**(1): 71–85.

Doyle, O., L. Delaney, C. O'Farrelly, N. Fitzpatrick and M. Daly (2017), 'Can early intervention improve maternal well-being? Evidence from a randomized controlled trial', *PLoS ONE*, **12**(1): e0169829.

Duckworth, A. L., T. S. Gendler and J. J. Gross, (2016), 'Situational strategies for self-control', *Perspectives on Psychological Science*, **11**(1): 35–55.

Duflo, E. (2017), 'Richard T. Ely Lecture: The Economist as Plumber', *American Economic Review*, **107**(5): 1–26.

Elster, J. (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge, UK: Cambridge University Press.

Fisher, I. (1918), 'Is "utility" the most suitable term for the concept it is used to denote?' *The American Economic Review*, **8**(2): 335–337.

Frey, B. S. and A. Stutzer (2002), 'What can economists learn from happiness research?' *Journal of Economic Literature*, **40**(2): 402–435.

Fujita, K. (2011), 'On conceptualizing self-control as more than the effortful inhibition of impulses', *Personality and Social Psychology Review*, **15**(4): 352–366.

Galizzi, M. M. and D. Navarro-Martinez (2018), 'On the external validity of social preference games: a systematic lab-field study'. *Management Science*, Epub ahead of print.

Gneezy, U. and A. Imas (2017), 'Chapter 10 – lab in the field: measuring preferences in the wild', In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, (Vol. **1**, pp. 439–464). North-Holland.

Grube, A., J. Schroer, C. Hentzschel and G. Hertel (2008), 'The event reconstruction method: An efficient measure of experience-based job satisfaction', *Journal of Occupational and Organizational Psychology*, **81**(4): 669–689.

Grüne-Yanoff, T. (2015), 'Why behavioural policy needs mechanistic evidence', *Economics and Philosophy*, 1–21.

Halpern, D. (2015), 'The rise of psychology in policy: The UK's de facto Council of Psychological Science Advisers', *Perspectives on Psychological Science*, **10**(6): 768–771.

Harrison, G. W. (2014), 'Cautionary notes on the use of field experiments to address policy issues', *Oxford Review of Economic Policy*, **30**(4): 753–763.

Harrison, G. W. and J. A. List (2004), 'Field experiments', *Journal of Economic Literature*, **42**(4): 1009–1055.

Hausman, D. M. (2000), 'Revealed preference, belief, and game theory', *Economics & Philosophy*, **16**(1): 99–115.

Hausman, D. M. (2012), *Preference, Value, Choice, and Welfare*, Cambridge, UK: Cambridge University Press.

Hoch, S. J. and G. Loewenstein (1991), 'Time-inconsistent preferences and consumer self-control', *Journal of Consumer Research*, **17**(4): 492–507.

Hofmann, W., R. F. Baumeister, G. Förster and K. D. Vohs (2012), 'Everyday temptations: An experience sampling study of desire, conflict, and self-control', *Journal of Personality and Social Psychology*, **102**(6): 1318–1335.

Hofmann, W. and P. V. Patel (2015), 'SurveySignal: a convenient solution for experience sampling research using participants' own smartphones', *Social Science Computer Review*, **33**(2): 235–253.

Infante, G., G. Lecouteux and R. Sugden (2016), 'Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics', *Journal of Economic Methodology*, **23**(1): 1–25.

Ishio, J. and N. Abe (2017), 'Measuring affective well-being by the combination of the Day Reconstruction Method and a wearable device: case study of an aging and depopulating community in Japan'. *Augmented Human Research*, **2**(1): 2.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

Jiménez-Buedo, M. (2011), 'Conceptual tools for assessing experiments: some well-entrenched confusions regarding the internal/external validity distinction', *Journal of Economic Methodology*, **18**(3): 271–282.

Jolls, C., C. R. Sunstein and R. H. Thaler (1998). 'A behavioral approach to law and economics', *Stanford Law Review*, 1471–1550.

Kahneman, D., A. B. Krueger, D. A. Schkade, N. Schwarz and A. A. Stone (2004a). 'A survey method for characterizing daily life experience: the Day Reconstruction Method', *Science*, **306**(5702): 1776–1780.

Kahneman, D., A. B. Krueger, D. A. Schkade, N. Schwarz and A. A. Stone (2004b). *The Day Reconstruction Method (DRM). Instrument Documentation*. Supporting Online Material for Kahneman *et al.* (2004a). URL https://dornsife.usc.edu/assets/sites/780/docs/drm_documentation_july_2004.pdf.

Kahneman, D. and R. Sugden (2005), 'Experienced utility as a standard of policy evaluation'. *Environmental and Resource Economics*, **32**(1): 161–181.

Kaplan, R. M. and A. A. Stone (2013), 'Bringing the laboratory and clinic to the community: mobile technologies for health promotion and disease prevention', *Annual Review of Psychology*, **64**, 471–498.

Kim, J., H. Kikuchi and Y. Yamamoto (2013), 'Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults', *British Journal of Health Psychology*, **18**(1): 155–167.

Knabe, A., S. Rätzel, R. Schöb and J. Weimann (2010), 'Dissatisfied with life but having a good day: time-use and well-being of the unemployed', *The Economic Journal*, **120**(547): 867–889.

Knabe, A., R. Schöb and J. Weimann (2017), 'The subjective well-being of workfare participants: insights from a day reconstruction survey', *Applied Economics*, **49**(13): 1311–1325.

Krueger, A. B., D. Kahneman, D. Schkade, N. Schwarz and A. A. Stone (2009), 'National time accounting: The currency of life'. In *Measuring the subjective well-being of nations: National accounts of time use and well-being*, (pp. 9–86). Chicago, IL: University of Chicago Press.

Larson, R. and M. Csikszentmihalyi (1983), 'The experience sampling method', *New Directions for Methodology of Social & Behavioral Science*, **15**, 41–56.

Layard, R. (2006), 'Happiness and public policy: a challenge to the profession', *The Economic Journal*, **116**(510): C24–C33.

Lee, P. H., A. C. Y. Tse and K. Y. Lee (2017), 'A new statistical model for the Day Reconstruction Method', *International Journal of Methods in Psychiatric Research*, **26**(4): e1547.

Levitt, S. D. and J. A. List (2007), 'What do laboratory experiments measuring social preferences reveal about the real world?' *The Journal of Economic Perspectives*, **21**(2): 153–174.

Loewenstein, G. (1999), 'Experimental economics from the vantage-point of behavioural economics', *The Economic Journal*, **109**(453): 25–34.

Maslow, A. H. (1943), 'A theory of human motivation', *Psychological Review*, **50**(4): 370–396.

Milyavskaya, M. and M. Inzlicht (2017), 'What's so great about self-control? Examining the importance of effortful self-control and temptation in predicting real-life depletion and goal attainment', *Social Psychological and Personality Science*, **8**(6): 603–611.

Miret, M., F. F. Caballero, A. Mathur, N. Naidoo, P. Kowal, J. L. Ayuso-Mateos and S. Chatterji (2012), 'Validation of a measure of subjective well-being: an abbreviated version of the Day Reconstruction Method', *PLoS ONE*, **7**(8): e43887.

National Research Council (US) (2012), *The Subjective Well-Being Module of the American Time Use Survey: Assessment for Its Continuation*, Panel on Measuring Subjective Well-Being in a Policy-Relevant Framework. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press (US).

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press

Odermatt, R. and A. Stutzer (2017), 'Subjective well-being and public policy'. In E. Diener, S. Oishi, and L. Tay (Eds.), *Handbook of Well-Being*, Salt Lake City, UT: DEF Publishers.
Oliver, A. (2017), *The Origins of Behavioural Public Policy*, Cambridge, UK: Cambridge University Press.
Pareto, V. (1971), *Manual of Political Economy*, Basingstoke, UK: Macmillan.
Rabin, M. (1998), 'Psychology and economics', *Journal of Economic Literature*, **36**(1): 11–46.
Rizzo, M. J. and D. G. Whitman (2009), 'The knowledge problem of new paternalism', *Brigham Young University Law Review*, **2009**(4): 905–968.
Samuelson, P. A. (1938), 'A note on the pure theory of consumer's behaviour', *Economica*, **5**(17): 61–71.
Schwarz, N. and G. L. Clore (1983), 'Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states', *Journal of Personality and Social Psychology*, **45**(3): 513–523.
Shafir, E. (2013), *The Behavioral Foundations of Public Policy*, Princeton, NJ: Princeton University Press.
Shiffman, S., A. A. Stone and M. R. Hufford (2008), 'Ecological momentary assessment', *Annual Review of Clinical Psychology*, **4**, 1–32.
Sonnenberg, B., M. Riediger, C. Wrzus and G. G. Wagner (2012), 'Measuring time use in surveys – concordance of survey and experience sampling measures', *Social Science Research*, **41**(5): 1037–1052.
Srivastava, S., K. M. Angelo and S. R. Vallereux (2008), 'Extraversion and positive affect: a day reconstruction study of person–environment transactions', *Journal of Research in Personality*, **42**(6): 1613–1618.
Stiglitz, J. E., A. Sen and J.-P Fitoussi. (2010), *Report by the Commission on the Measurement of Economic Performance and Social Progress*, Paris, France: Commission on the Measurement of Economic Performance and Social Progress.
Sugden, R. (2017), 'Do people really want to be nudged towards healthy lifestyles?' *International Review of Economics*, **64**(2): 113–123.
Sunstein, C. R. (2016a), 'Cost–benefit analysis, who's your daddy?' *Journal of Benefit–Cost Analysis*, **7**(1): 107–120.
Sunstein, C. R. (2016b), *The Ethics of Influence: Government in the Age of Behavioral Science*, New York, NY: Cambridge University Press.
Thaler, R. H. (2016), 'Behavioral economics: past, present, and future', *American Economic Review*, **106**(7): 1577–1600.
Thaler, R. H. (2018), 'From cashews to nudges: the evolution of behavioral economics', *American Economic Review*, **108**(6): 1265–1287.
Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, CT: Yale University Press.
Tourangeau, R., L. J. Rips and K. Rasinski (2000), *The Psychology of Survey Response*, Cambridge, UK: Cambridge University Press.
Weber, E. U., A.-R. Blais and N. E. Betz (2002), 'A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors', *Journal of Behavioral Decision Making*, **15**(4): 263–290.
Wilkowski, B. M., E. L. Ferguson, L. Z. Williamson and S. K. Lappi (2018), '(How) does initial self-control undermine later self-control in daily life?' *Personality and Social Psychology Bulletin*, **44**(9): 1315–1329.
Witt, U. (2017), 'The evolution of consumption and its welfare effects', *Journal of Evolutionary Economics*, **27**(2): 273–293.

https://doi.org/10.1017/bpp.2018.37 Published online by Cambridge University Press