CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Facing it: assessing the immediate emotional impacts of calorie labelling using automatic facial coding

Kate Laffan[1,2]* , Cass Sunstein[3] and Paul Dolan[1]

[1]Department of Psychological and Behavioural Science, London School of Economics and Political Science, London, UK, [2]School of Economics, University College Dublin, Dublin, Ireland and [3]Harvard Law School, Harvard University, Cambridge, MA, USA
*E-mail: k.m.laffan@lse.ac.uk

Abstract

Although there has been a proliferation of research and policy work into how nudges shape people's behaviour, most studies stop far short of consumer welfare analysis. In the current work, we critically reflect on recent efforts to provide insights into the consumer welfare impact of nudges using willingness to pay and subjective well-being reports and explore an unobtrusive approach that can speak to the immediate emotional impacts of a nudge: automatic facial expression coding. In an exploratory lab study, we use facial expression coding to assess the short-run emotional impact of being presented with calorie information about a popcorn snack in the context of a stylised 'Cinema experience'. The results of the study indicate that calorie information has heterogeneous impacts on people's likelihood of choosing the snack and on the emotions they experience during the moment of choice which varies based on their level of health-consciousness. The information does not, however, affect the emotions people go on to experience while viewing movie clips, suggesting that the emotional effects of the information are short-lived. We conclude by emphasising the potential of automatic facial expression coding to provide new insights into the immediate emotional impacts of nudges and calling for further research into this promising technique.

Introduction

Behaviourally informed approaches to public policy have gained traction over the past decade (OECD, 2017; Troussard & Van Bavel, 2018), especially choice-preserving interventions that steer people's behaviour in specific directions (Thaler & Sunstein, 2009; Loewenstein & Chater, 2017). Despite the proliferation of such 'nudges', most evaluations of these interventions stop far short of a consumer welfare analysis, most commonly presenting the magnitude of behaviour change achieved or

less frequently a form of cost-effectiveness analysis based on the resources used or saved by the nudge (Benartzi *et al.*, 2017).

Where attempts have been made to provide further insights into consumer welfare effects, they have used proxy measures of welfare, such as willingness to pay (WTP) (Allcott & Kessler, 2019; Rafiq, 2021) or elicited the short-term emotional impacts of an intervention using measures of subjective well-being (SWB) (Thunström, 2019). Asking about WTP and SWB yields important insights into the potentially heterogeneous effects of an intervention, but the self-reported nature of the data means that they are vulnerable to a host of reporting biases.

In this article, we explore an unobtrusive approach that sheds new light on the immediate emotional effects of nudges – an important component of their consumer welfare impact that has previously been assessed using self-report measures. Our approach involves using automatic facial expression coding to capture the immediate emotional impacts of behavioural interventions. Facial expression coding technology can read expressions via built-in webcams when people engage with tasks on a computer screen in the lab or the wild. Automatic facial expression coding is increasingly used to capture individuals' emotions for a variety of purposes, including marketing research where it has been used to measure individuals' responses to online advertising (Teixeira *et al.*, 2012); clinical settings where it has been used to monitor patients' depression symptoms (Girard *et al.*, 2013); and public policy research to assess individuals' responses to bureaucratic red tape (Hattke *et al.*, 2020). Automatic facial coding has not yet been used to examine people's emotional responses to behaviourally informed interventions, including nudges.

In the current work, we study the immediate emotional impact of a behavioural intervention that has been incorporated into policy in the US since 2014: namely, the mandatory provision of calorie information on movie theatre snack menus (US Food and Drug Administration, 2014). In an exploratory study, we invited participants to a 'Cinema experience' involving the screening of short movie clips. Before the screening, we offered the participants high-calorie sweet and salty popcorn, and randomly exposed them to see or not to see calorie information about the popcorn. Our key outcomes of interest are (1) people's choice about whether to have the popcorn, reflecting a standard 'behavioural' outcome and (2) the percentage of time spent in a negative emotional state during the moment of choice and subsequently while watching movie clips, both captured through automatic facial expression coding. We also use a survey measure to assess the potential role of health-consciousness as a moderator of the behavioural and emotional effects of being presented with the calorie information.

We find that, on average, presenting people with calorie information does not affect either their choice to have popcorn or the negative emotions they experience while making the choice. When we examine how the treatment interacts with people's health-consciousness, however, striking differences emerge. We find that the behaviour of health-conscious people is unaffected by being presented with calorie information. In other words, the nudge does not work as intended for these people. This finding is in line with some previous research into calorie labelling (Ellison *et al.*, 2013) and may be explained by the idea that the health-conscious are already more knowledgeable about nutrition and aware of the unhealthy nature of the snack

on offer. Interestingly, though, the messaging does take a short-lived emotional toll on this group: health-conscious people experience higher levels of negative affect during the moment of choice when presented with the calorie information but not subsequently while watching the movie clips. These are significant findings, confirming that whatever the effect of a nudge on behaviour, it can also produce immediate negative emotional consequences (Thunström, 2019).

In contrast, we find that, for the less health-conscious participants, being exposed to the calorie information significantly decreases their likelihood of choosing the popcorn, which means that the nudge changes the behaviour of those that calorie labelling laws are often trying to influence. At the same time, the intervention does not come at an immediate emotional cost to these people. It follows that there are no adverse emotional effects, at least immediate ones, to counter the beneficial health effects. Taken together, the results of this exploratory study demonstrate the ability of automatic facial coding data to provide valuable insights into heterogeneous emotional responses to behavioural interventions.

In what follows, we first review existing approaches to provide insights into the consumer welfare effects of nudges. We then present the methods and results of our exploratory 'Cinema experience' study which uses automatic facial coding to examine the immediate emotional responses to calorie information. Finally, we discuss the results and their limitations, and make some recommendations for future work examining the emotional impacts of nudges that build on our findings.

Existing approaches to provide insights into consumer welfare effects of nudges

Assessing the consumer welfare impacts of nudges requires a clear understanding of the much-disputed idea of 'welfare'. There are diverse philosophical perspectives on this question, with responses, broadly speaking, falling into three categories: the objective list account, the preference satisfaction account, and the mental state account (Adler, 2019). Under the objective list account, people's welfare is dependent on the extent to which their lives contain a list of things that are considered to be objectively good for human beings, such as good health and education. Approaches that have focused on the magnitude of behaviour change achieved by an intervention could be seen to fit within an objective list account insofar as the target behaviour falls on such a list (again, such as good health and education).

Recent efforts to provide insights into the consumer welfare effects of nudges fit most easily with accounts that rely on preference satisfaction or mental states. WTP falls under a preference satisfaction account of well-being – the most commonly accepted account of welfare in economics. This account holds that people are better off when they get more of what they prefer, with WTP measures providing proxy measures of the strength of those preferences. SWB measures, on the other hand, reflect a mental state account of well-being. This account posits that people are better off when they feel better. Reflecting this account, SWB measures are commonly used within psychology and increasingly economics to proxy how people feel (Dolan & Metcalfe, 2012). Given that there is significant disagreement about the relative merits of the different accounts of welfare, examinations of the consumer welfare effects of nudges that adopt different perspectives can provide valuable insights.

Willingness to pay

A standard economic approach to consumer welfare effects, used in Allcott and Kessler (2019), involves WTP. WTP is, of course, the measure of valuation in actual markets and, in hypothetical forms, it has also been used to value non-market goods, particularly in environmental economics (Atkinson *et al.*, 2012). In principle, it should capture everything that matters, that is, the entire welfare effect (at least if we indulge the admittedly unrealistic assumption that consumers are perfectly informed and free from behavioural biases; we will return to that assumption). If consumers are willing to pay a specified amount for shoes or sneakers, or for (say) information about fuel economy, that amount (given an individual's marginal utility of income) should capture both the positive and negative welfare effects of shoes or sneakers, or information about fuel economy.

Against this background, Allcott and Kessler (2019) use incentive-compatible WTP estimates to explore the consumer welfare impacts of receiving a home energy report that compares people's consumption to that of their neighbours – a classic nudge based on social norms. Their results find that traditional estimates, based on average economic savings, greatly overstate the benefits of these reports (estimating them to be approximately six times higher). The difference is driven by the fact that many people are willing to pay far less than the average economic savings that come from the nudge. Indeed, many would pay not to receive it at all. An important implication is that the potential economic savings provided by an informational nudge may overstate the welfare benefits of that nudge. This is because people might not enjoy receiving it, and/or might not enjoy engaging in the work that is necessary to obtain those savings.

In many contexts, WTP may be the best available measure of likely consumer welfare effects, but for informational nudges, it may not be sufficiently reliable (Sunstein, 2019). The most fundamental reason is that WTP requires us to solve a prediction problem, and people may not be in a good position to solve that problem. When people lack information or suffer from a behavioural bias, it is difficult for them to generate numbers that capture the actual welfare effects of receiving that information. To know the value of a home energy report, for example, one needs to know something about its likely effects, and many people lack the requisite foresight. Thunström (2019) finds that people are, on average, willing to pay $0.39 for calorie information. But what is the information on which they base that estimate? What do they know, for example, about the effects of high-calorie selections?

A similar problem impacts upon the use of WTP to specify the value of many nudges that we may be interested in evaluating, for example, information-based nudges such as social norms-based messages around healthy eating (Stok *et al.*, 2014) or a framing effect in climate change communication (Spence & Pidgeon, 2010). Given that most preference satisfaction accounts require that preferences are adequately informed (see Harsanyi, 1985), it is hard for people to estimate a WTP value for information that they lack, or for experiences they are yet to have.

Subjective well-being

Subjective well-being is an umbrella term that encompasses both individuals' reports about how they think about their lives and their feelings as they go about them

(Diener *et al.*, 2006). While evaluative measures of SWB, such as life satisfaction, can be understood as hybrid measures that reflect both people's preferences and their mental states, experiential measures, like self-reported happiness or anxiety in the moment, more closely approximate people's emotional experiences, and thus a mental state account of welfare. Thunström (2019) introduces experiential SWB measures into the evaluation of immediate emotional impacts of a behavioural intervention – an important short-run component of consumer welfare effect but, importantly, not the entire welfare impact. In this work, Thunström demonstrates that the presentation of calorie labels has heterogeneous effects on people in terms of both their hypothetical choices and their SWB, identifying self-control as an important moderator.

Many people consider SWB measures to capture reliable and useful information about individuals' experienced well-being (Stone & Mackie, 2013). This view is supported by evidence showing that SWB measures have reasonable test–retest scores (Krueger & Schkade, 2008) and that they are positively correlated with things that could be considered proxy measures for well-being, such as how often individuals' smile (Ekman *et al.*, 1990), as well as friends, family members, and clinical experts' ratings of their well-being (Frey & Stutzer, 2002; Schneider & Schimmack, 2009). SWB measures are also negatively associated with cortisol levels (Steptoe *et al.*, 2005), hypertension (Blanchflower & Oswald, 2008), and suicide rates (Koivumaa-Honkanen *et al.*, 2001).

SWB measures provide valuable information, but they should be approached with caution. A measure of self-reported well-being is not a direct measure of people's emotional state. With respect to happiness, for example, people might say that they are at 5 on a scale of 1–7, which is surely informative, but it is not entirely clear what the actual number means. The challenge becomes even harder when the goal is to assess the emotional impacts of an intervention (such as a fuel economy label) whose effect, even if real, is likely to be relatively small and short-lived.

SWB measures are also subject to reactivity effects when the act of asking people about their emotions influences how they report feeling. These effects can come about due to social desirability bias if people exaggerate (or downplay) the emotional impacts of an intervention if they feel they ought to be (ought not to be) affected and perceive being so as socially desirable (Krumpal, 2013). They can also occur if the act of asking the SWB questions artificially focuses attention on the intervention causing a form of focusing illusion (Kahneman *et al.*, 2006). Finally, when both WTP and SWB reports are elicited in response to hypothetical choices they are also vulnerable to hypothetical bias, as hypothetical WTP responses are not incentive compatible (Murphy & Stevens, 2004) and people may commit affective forecasting errors when reporting the emotions they expect to experience during a hypothetical or future scenario (Wilson & Gilbert, 2005).

Facial coding

In the current work, we explore an unobtrusive technique that can provide insights into the immediate emotional impacts of nudges: automatic facial expression coding. Automatic facial expression coding represents a technologically advanced alternative to the long-established Facial Action Coding System (FACS), which was first

developed in 1976 (Ekman & Friesen, 1976). FACS is a fully standardised classification system of facial expressions, used by trained human coders to analyse videos of people's faces and describe the occurrence of facial expressions.

In contrast, automatic facial coding is based on machine learning algorithms that have been trained on large databases of facial expressions to analyse emotions in the face (Imotions, 2017). A growing body of literature has explored whether automatic facial coding can reliably detect emotions and how it compares to previously established methods. This research literature suggests that the performance varies significantly across software (Dupré *et al.*, 2020) Affectiva – the software used in the current work has been shown to be able to recognise emotions from both photographs and videos (Stöckli *et al.*, 2018). For example, Stöckli *et al.* (2018) report a performance index of 73% on photographs. This is reported to be within the typical performance range for human coders range of 60% and 80% (Nelson & Russell, 2013). Stöckli *et al.* (2018) further documented that Affectiva classifies videos of emotional responses to emotionally evocative pictures – much like the treatment explored in the current work – with the correct valence 57% of the time. Importantly, and in particular, as we focus on negative affect in the current work, Affectiva was found to accurately classify 17% positive responses and 97% of negative ones. Affectiva has also been shown to generate values that correlate highly with those generated through electromyography – the current psychophysiological gold standard (Kulke *et al.*, 2020).

The above research suggests that Affectiva represents a useful tool to inquire into people's emotions based on their facial expressions. However, other research has raised important issues around the ability of automatic facial coding to pick up cross-cultural differences (Srinivasan & Martinez, 2018), potential biases within the estimates in relation to ethnicity (Rhue, 2018) and the need for further research into naturally occurring emotional facial reactions (Dupré *et al.*, 2020). The analysis in the current work is subject to all of these limitations and represents initial exploratory work using this fast-developing technology.

The major advantages of automatic facial coding over FACS-based analysis are that it is far less time consuming and that it does not require a high level of expertise to carry out the analysis. These advantages are of relevance when considering the potential applicability of this technique to the examination of the immediate emotional impacts of nudges in a policymaking setting.

Automatic facial coding also has both disadvantages and advantages over the existing approaches that have been used to provide insights into the consumer welfare effects of nudges. On the negative side, in contrast with WTP measures, which in theory assess the entire welfare effects of nudges, automatic facial coding only provides insights into people's immediate emotional responses. Moreover, in comparison with SWB measures which can be captured via surveys, automatic facial coding requires access to more advanced technology which can be costly.

On the positive side, automatic facial coding does not require the kind of information and foresight necessary to ensure that a WTP measure accurately captures actual consumer welfare effects. Moreover, in contrast to the SWB approach, facial coding is by its nature less vulnerable to reactivity effects and biases such as social desirability bias and focusing effects. A further advantage of facial coding is that it allows for the continuous monitoring of individuals' emotions while they are exposed to the nudge

and during follow-up experiences, facilitating the examination of the longevity of the emotional impacts of the nudge in question for as long as their facial expressions are captured. These advantages arise from the fact that facial coding can capture people's emotional responses to stimuli over time, rather than asking them to report preferences over a hypothetical experience or prior to an experience as is typically the case with WTP measures or interrupting people's experiences to ask them their feelings about them as SWB reports necessitate.

An exploratory study

We carried out an exploratory study that uses automatic facial expression coding to capture the emotional impacts of a behavioural intervention of widespread interest: the presentation of calorie information about an unhealthy popcorn snack in the context of a cinema experience. We use this approach to assess the immediate emotional impacts of the nudge during the period in which people are subject to it and during their experiences directly post-exposure. Negative emotions are understood to occur in response to being presented with tempting goods, involving short-term rewards coupled with larger long-term costs, as a part of a self-regulation strategy (Lowenstein & O'Donoghue, 2006). If participants in the study are tempted by the unhealthy popcorn snack, they may respond negatively to information that highlights its calorie content, including experiencing negative emotions such as anticipated regret or guilt. Importantly, we would not expect nutritional information in general to evoke negative emotions. Indeed, existing research based on self-reported emotional responses to a healthy food product found that people who were presented with nutritional information about its health benefits reported higher levels of positive emotion than those who did not see the information (Lagerkvist *et al.*, 2016). Given the popcorn on offer in the current experiment was unhealthy, being high in calories, fat sugar, and salt, we use automatic facial coding to explore whether information on its nutritional content evokes negative emotional responses.

Method

In Summer 2018, participants were recruited to take part in a study on video footage and decision making from the participant pool of the Behavioural Research Lab at the London School of Economics and Political Science. A total of 107 people participated in the study.

Upon arrival at the lab, we briefed all participants on the use of webcams, and they provided consent to participate in the study and for their facial expressions to be captured on camera and later analysed using automatic facial expression coding software. We then provided participants with instructions on how to adjust their chair height and seating position to ensure that their faces were in clear view of the webcams. Before beginning, we checked the video stream of all participants to confirm that their webcam and seating position allowed us to capture their facial expressions clearly.

We introduced all students to the 'Cinema experience', which we explained would involve viewing video clips of movies that they may or may not have seen before. The two movie clips were from the 2008 movie *Wall-E* and the 1993 movie *Searching for*
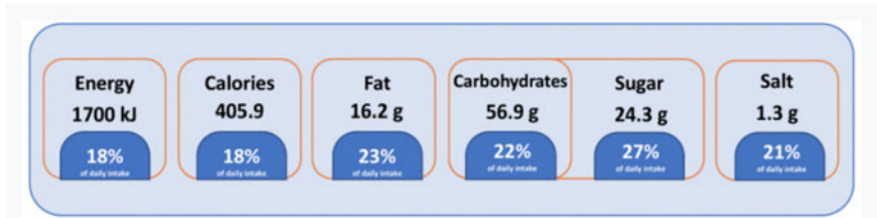
Figure 1. Nutritional Information Presented in the Treatment Condition.

Bobby Fischer. The emotions elicited by these clips have been investigated in previous research, with the two clips being rated highly for happiness and calmness, respectively (Gabert-Quillen *et al.*, 2015). We also made participants aware that they had the option of choosing to have a free packet of sweet and salty popcorn as part of the experience. The study design represents a simple experiment with participants randomly allocated to be in either the treatment or control group. We carried out our analysis on the videos for the three key time periods of interest. Six observations were excluded from the analysis due to poor data quality, leaving a complete sample of 101 people ($n = 48$ in the treatment and $n = 53$ in the control group).

In the treatment group, participants were presented with the nutritional information about the packet of popcorn on offer on-screen (see Figure 1). The packet of popcorn contained 405.9 calories or 18% of an average female adult's daily allowance. We took these figures from the original nutritional information provided on the packet of popcorn, which we concealed on all of the physical packets of popcorn on offer. In the control group, no information was available about the nutritional content of the popcorn until after the popcorn choice and movie clips.

The choice of whether to have the popcorn or not was assessed by the individuals' binary response to being offered the popcorn on screen. Participants that chose the snack were given a bag of the popcorn displayed on screen. The webcams captured video footage of the facial expressions of the participants throughout the experience. Clips relating to three key periods were analysed using the automatic facial coding: (1) the percentage of time spent in a negative emotional state during the choice of popcorn; (2) the percentage of time spent in a negative emotional state during the first movie clip from *Wall-E*; and (3) the percentage of time spent in a negative emotional state during the second movie clip from *Searching for Bobby Fischer*. Importantly, as participants were allowed to take unfinished popcorn with them when they left the laboratory, we are not able to track their emotions throughout the consumption or disposal of the snack. As a result, we do not claim to capture the entire emotional effects of intervention but rather only the immediate emotional responses to the nudge and the experiences directly after exposure while their facial expressions were still being captured.

The footage was analysed in the biosensor platform Imotions using the facial expression coding software Affectiva. This software has been used in publications on a range of topics including, for example, socially transmitted placebo effects (Chen *et al.*, 2019) and the impact of safety system warnings on people's emotions (Fahim *et al.*, 2019). The percentage of time spent in a negative emotional state

was calculated by Affectiva software according to the amount of time displaying negative emotion out of the total time recorded for stimulus. The software bases its prediction of negative emotion on its detection of the facial action units including inner brow raise, brow furrow, and lip press, for example. It then applies a confidence threshold to provide an estimate of the duration of the negative emotional response to the stimuli. The total time varied across stimuli: people viewed the screen with the nutritional information for an average of 34 s for the nutritional information, while both film clips were over 2 minutes.

We also collected a range of demographic information and an answer to a question that has been used in the existing literature to measure individuals' health-consciousness: their agreement with the statement 'Living life in the best possible health is very important to me' (Dutta-Bergman, 2004; Dutta, 2007). We measured health-consciousness on a Likert scale and then we split responses at 3 as a score of 4 or above indicated that people agreed either agree or strongly agree with the statement. See Table A1 of the Appendix for descriptive statistics. The choice of popcorn was modelled using a binary logistic regression and all other outcomes were examined using ordinary least squares regression in Stata 15. The treatment and control groups were balanced across age, gender, and whether the participants reported having a university degree (see Table Appendix Table A2).

Results

In the analysis, we explore the impact of the treatment on three outcomes of interest: the choice, the negative emotion during the moment of choice, and the negative emotion beyond the moment of choice while watching film clips.

The impact of the treatment on choice

Overall, the results indicate that being shown information about the popcorn's nutritional content has no significant impact on people's likelihood of choosing the popcorn (Model I). Looking at interactions between the treatment and the participants' health-consciousness, however, we find weak evidence that being shown the nutritional information reduces the likelihood of choosing popcorn among the non-health-conscious: those who do not report that living life in the best possible health is important to them (multiplying the odds by 0.214, significant at the 10% level). By contrast, being shown the nutritional information does not reduce the likelihood of choosing popcorn among the health-conscious (those who agree with the statement) (Model II).

The impact of the treatment on experienced well-being during choice

Looking at the overall sample, the results indicate that the treatment had no significant impact on people's negative emotions (Model III). This overall null effect disappears, however, when we examine interaction effects between the treatment and whether people report being health-conscious. For those who are health-conscious, the information increases the percentage of time spent in a negative emotional

state by 3.5% (significant at the 5% level). By contrast, the information does not affect the non-health-conscious (Model IV).

The impact of the treatment on experienced well-being beyond the moment of choice

We then investigate whether the impact of the treatment on people's emotions lasts beyond the moment of choice, spilling over into the people's levels of negative emotion during the subsequent film clips. On average, the percentage of time spent in a negative emotional state during each of the two film clips is unaffected by the presentation of calorie information (Models V and VII) and there is no significant interaction between the treatment effect and health-consciousness (Models VI and VIII) (Table 1).

Discussion

The current work

Some recent evaluations of nudge-type interventions have sought to go beyond capturing behavioural impacts to provide insights into how such interventions affect an individual's welfare (Allcott & Kessler, 2019; Rafiq, 2021; Thunström, 2019). To date, these efforts have involved eliciting either WTP or SWB reports. In the current work, we focus on an important component of the consumer welfare effects of nudges: their immediate emotional impacts. We explore the usefulness of automatic facial expression coding in shedding light on these impacts. This approach offers some advantages over existing methods in contexts where: (1) we have good reasons not to trust people's WTP reports as indicators of the welfare impacts of nudges, perhaps because of inadequate information; and (2) we have good reason to think that SWB measures potentially provides an inaccurate or biased picture, perhaps because of reactivity effects. Importantly, facial coding also facilitates the examination of the longevity of emotional responses to nudges.

In an exploratory study, we use automatic facial coding to gain insights into the immediate emotional impacts of a specific nudge: we present people with nutritional information about sweet and salty popcorn on offer as part of a 'Cinema experience'. The study represents a stylised representation of a real-world regulation – the mandatory provision of calorie information on cinema theatre menus – that exists in the US (US Food and Drug Administration, 2014). We explore whether the nutritional information we present affects people's propensity to choose to have popcorn and whether that information affects the percentage of time they spend in a negative emotional state during the choice and during their subsequent experiences watching movie clips.

The results of the study indicate that on average the intervention has no impact on people's consumption. In subgroup analysis, however, we find that seeing the nutritional information significantly affects the choices of those who do not report being health-conscious, leading them, perhaps counterintuitively, to make healthier choices. Similarly, in a study on calorie labels in restaurants, Ellison *et al.* (2013) found that labels worked best on the least health-conscious and that numeric labels (such as the

Table 1. The impact of the treatment on choice and negative emotion.

Variables	Model I Chose popcorn (Odds ratio)	Model II Chose popcorn (Odds ratio)	Model III Negative emotion during choice	Model IV Negative emotion during choice	Model V Negative emotion during Wall-E	Model VI Negative emotion during Wall-E	Model VII Negative emotion during Bobby	Model VIII Negative emotion during Bobby
Treatment	0.528 (0.216)	0.214* (0.198)	1 (0.724)	−1.76 (1.537)	1.899 (3.236)	6.042 (7.009)	0.678 (0.636)	−0.487 (1.355)
Health-conscious		0.432 (0.297)		−1.058 (1.178)		2.933 (5.373)		−1.951* (1.039)
Treatment* Health-conscious		3.152 (3.249)		3.511** (1.737)		−5.3 (7.919)		1.516 (1.531)
Constant	1.038 (0.285)	2 (1.225)	2.273*** (0.499)	3.092*** (1.036)	5.254*** (2.231)	2.985 (4.725)	3.059*** (0.438)	4.569*** (0.914)
Observations	101	101	101	101	101	101	101	101
R^2			0.019	0.062	0.0113	0.008	0.0113	0.047

Notes: Models I and II are estimated using binary logistic regression. All other models are estimated using ordinary least squares regression. Standard errors appear in parentheses.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

one used in the current work) had little impact on health-conscious consumers. Importantly, the evidence presented here is somewhat weak. A large literature exists that finds mixed results on the impact of calorie information on food choices with heterogeneous effects demonstrated across different types of people and information presentation formats (for reviews, see Harnack & French, 2008; Kiszko *et al.*, 2014).

In the current study, the immediate emotional impacts of the intervention are of central interest and so we analyse the data generated from the automatic facial coding procedure to investigate whether the treatment affects the participants' levels of negative emotion during the choice. In contrast with Hattke *et al.* (2020), who find that being exposed to bureaucratic red tape negatively impacts people's emotions, we do not find that calorie information negative impacts people's emotions in our sample overall. When we carry out interaction analysis, however, we find that seeing the nutritional information has a negative emotional impact on health-conscious people, while having no impact on those who do not report that living life in the best possible health is important to them. These findings have some relationship to those in Thunström (2019), in that she also finds heterogeneous effects of calorie information on emotion, with some participants being emotionally taxed by the calorie information, while others are subsidised.

Taken together, the asymmetrical behavioural and emotional effects have policy relevance. At first glance, the intervention appears to have no impact on either behaviour or emotions. But the interaction with health-consciousness suggests that it produces emotional costs for the health-conscious without providing any health benefits, while producing health benefits for those who are not health-conscious at no emotional cost. By themselves, those contrasting effects bear on appropriate policy. If increased well-being is our goal, of course, we might want to begin by measuring the net welfare costs and the net welfare benefits, and seeing which is higher. As our data can only speak to one component of the welfare costs – the immediate emotional impact – they do not permit quantitative measurements of that sort, but it points to their importance.

We also investigate the longevity of the *in situ* emotional impacts of the calorie information on people's 'Cinema experience' using the automatic facial coding data. This approach is novel. Hattke *et al.* (2020) who use automatic facial coding to explore responses to bureaucratic tape examined the immediate responses to 'red-tape' stimuli without following up to see whether the participants' emotions were affected post-exposure. Here, we find that the negative emotional effect that the calorie information has on the health-conscious is not sustained beyond the moment of choice. In other words, the *in situ* emotional costs of the intervention were quite short-lived. This finding is perhaps unsurprising given the small-scale of the behavioural intervention under examination, but it serves to emphasise the importance of considering how long an intervention continues to have emotional impacts and highlights the ability of automatic facial coding to speak to this question.

The study has limitations involving study design, and these suggest future directions for research. First, the sample size is small, limiting the robustness of the specific results and our ability to ask many questions of the data. Further investigations with larger samples could allow for more detailed analysis, which could include further interactions of theoretical interest. In particular, as the immediate emotional impact

of the nudge can act as a catalyst for a behavioural response, it can feed into both its short- and long-term welfare effects. Future work could examine interactions between different groups, emotional responses and behaviour to shed further light on the extent to which and how the intervention impacts on consumer welfare.

Second, the study did not include SWB reports of the participants' experienced well-being while making the choice. As a result, we cannot examine the extent to which they tell the same or a different story about the moment of choice when both the impacts of the information and vulnerability of focusing effects are likely to be greatest. It would be interesting in future work to capture both the objective and subjective emotion data. Such work could also involve WTP measures. Allcott *et al.* (2020) combine preferences measures with subjective well-being data to inquire into the welfare effects of social media. Future work on this and other topics could add automatic facial coding to produce even richer insights. Relatedly, future work could carry out more detailed analyses by examining a given intervention's impacts on specific emotions. This would be particularly useful should the technology further advance in ways that minimise the risk of miscategorisation of specific emotions (Höfling *et al.*, 2020).

Third, due to data limitations, the current work is not able to identify whether people suffering from eating disorders – who may be particularly sensitive to nutritional information about unhealthy foods – were included in our sample. Given that automatic facial coding has been used with clinical populations in prior work (Girard *et al.*, 2013), this approach may be able to provide valuable insights into the impact of nudges on vulnerable populations. Future work should consider which populations may be particularly sensitive to the specific nudge in question (nutrition labelling in the current work) and explore the use of automatic facial coding to inquire into their emotional responses to the intervention.

Despite these limitations, however, this exploratory study demonstrates that automatic facial coding can be used to examine the immediate emotional impacts of nudges and to help shed light on the longevity of any impacts arising from them.

Automatic facial coding going forward

The current work represents a first attempt at using facial coding to examine people's immediate emotional responses to nudge-type interventions. Assisted by software like Affectiva, automatic facial coding is easy and quick to carry out, opening up a new approach for researchers and policymakers to go beyond the question of whether a nudge affects people's behaviour, to gain insights into its immediate effects on how they feel. This technique may be of particular use in contexts where an intervention under consideration is effective at achieving the desired behaviour change but is contentious precisely because it is expected to negatively affect how people feel, for example, graphic warning labels on cigarette packages (Jung, 2016). In such cases, automatic facial coding data could bring evidence to bear on the extent and longevity of any emotional impacts so that they can be taken into account in the decision to implement the intervention or not. Facial coding may also be helpful in contexts where a given intervention has heterogeneous effects on behaviour, and the question of whether it emotionally taxes those whose behaviour it fails to affect is particularly pertinent.

Although facial coding represents a promising new approach, as previously discussed, it is a novel technology that requires further development. Important technical limitations, including concerning cross-cultural comparisons (Srinivasan & Martinez, 2018), will need to be addressed in order to maximise the technology's potential to inform public policy questions. There are also some other key issues that we believe are important to consider, relating to critiques of experienced utility as a welfare criterion, ethics, and the overall consumer welfare effects. We will take each of these in turn.

By assessing how people feel in the moment facial coding provides insights into people's welfare according to a mental state account of well-being. However, it is important to recognise that, in some situations, people may desire being exposed to a nudge, for example, be shown calorie information, even if it makes them unhappy. While such a preference may be based on an expectation that being exposed to the nudge will bring them happiness in the long-run, it may also indicate that mental states are not all that matter to people and that it is important to inquire into their preferences and not just their emotional experiences to gain insights into their welfare (Benjamin *et al.*, 2012).

It is also important to recognise that people's memory of the emotional profile of an experience is different to how it is experienced, with particular emphasis afforded to the peak and the end of experiences when they are encoded in memory (Oliver, 2017). On this basis, in situations where the emotional impacts of a nudge are expected to show up through the memories people have of them, solely assessing how any intervention impacts people in the moment might provide a biased picture and evaluative measures of SWB may be needed to complement the picture provided by facial coding.

Given the sensitivity of the data used in automatic facial coding and the privacy concerns around facial recognition, there are important ethical considerations involved in using this approach. We used facial coding in the current study in a lab setting and had the explicit permission of the participants to record and analyse their facial expressions during the 'Cinema Experience'. In terms of ecological validity, this is, of course, not equivalent to being examining people's responses to calorie information when they are in a queue for popcorn in a cinema. However, although automatic facial coding software can be used to facial expressions from clips of scenes from outside the lab, we have serious misgivings about capturing and analysing people's facial data in the wild. Limiting its use to controlled settings in which participants are aware that they are being recorded is on ethically safer ground and allows for the assessment of important moderators using survey instruments.

A final issue is that automatic facial coding produces insights into the immediate emotional responses to a nudge and not the welfare impact in its entirety (as WTP measures purport to do). The same limitation was acknowledged by Thunström in her work with SWB. While this technology represents a complementary approach to SWB in that is less vulnerable to reactivity effects and can more easily provide insights into people's flow of experiences, it still only provides a picture while the camera is still rolling. This precludes us from fully capturing all the impacts of a nudge on people's lived experiences. We could use this technique to examine people's immediate emotional responses to being defaulted into a pension scheme, for

example, but not to capture the decreased anxiety that might result from having a higher pay-out in their old age (Dobrescu *et al.*, 2018). Given this limitation, it is important to recognise that we are still a long way off capturing the entire welfare effects of nudges.

Conclusion

If nudges are to continue being widely used, there is a pressing need to examine their consumer welfare effects. Asking questions of people, including their WTP to be exposed to a nudge and about the effect they expect a nudge to have on their SWB, offers important insights but also faces some limitations. Though also not without challenges of its own, automatic facial expression coding represents a promising new approach that can help us better understand how behavioural interventions affect the immediate emotional responses of those subject to them, thereby contributing to our understanding of this important component of the consumer welfare effects of nudges.

References

Adler, M. D. (2019), *Measuring social welfare: An introduction*. New York, USA: Oxford University Press.
Allcott, H. and J. B. Kessler (2019), 'The welfare effects of nudges: A case study of energy use social comparisons', *American Economic Journal: Applied Economics*, **11**(1): 236–276.
Allcott, H., L. Braghieri, S. Eichmeyer and M. Gentzkow (2020), 'The welfare effects of social media', *American Economic Review*, **110**(3): 629–676.
Atkinson, G., I. Bateman and S. Mourato (2012), 'Recent advances in the valuation of ecosystem services and biodiversity', *Oxford Review of Economic Policy*, **28**(1): 22–47.
Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar and S. Galing (2017), 'Should governments invest more in nudging?' *Psychological Science*, **28**(8): 1041–1055.
Benjamin, D. J., O. Heffetz, M. S. Kimball and A. Rees-Jones (2012), 'What do you think would make you happier? What do you think you would choose?' *American Economic Review*, **102**(5): 2083–2110.
Blanchflower, D. G. and A. J. Oswald (2008), 'Hypertension and happiness across nations', *Journal of Health Economics*, **27**(2): 218–233.
Chen, P. H. A., J. H. Cheong, E. Jolly, H. Elhence, T. D. Wager and L. J. Chang (2019), 'Socially transmitted placebo effects', *Nature Human Behaviour*, **3**(12): 1295–1305.
Diener, E., M. Tamir and C. N. Scollon (2006), Happiness, Life Satisfaction, and Fulfillment: The Social Psychology of Subjective Well-Being.
Dobrescu, L. I., X. Fan, H. Bateman, B. Rhodri Newell, A. Ortmann and S. Thorp (2018), 'Retirement savings: A tale of decisions and defaults', *The Economic Journal*, **128**(610): 1047–1094.
Dolan, P. and R. Metcalfe (2012), 'Measuring subjective wellbeing: Recommendations on measures for use by national governments', *Journal of Social Policy*, **41**(2): 409–427.
Dupré, D., E. G. Krumhuber, D. Küster and G. J. McKeown (2020), 'A performance comparison of eight commercially available automatic classifiers for facial affect recognition', *PLoS ONE*, **15**(4): e0231968.
Dutta-Bergman, M. J. (2004), 'Primary sources of health information: Comparisons in the domain of health attitudes, health cognitions, and health behaviors', *Health Communication*, **16**(3): 273–288.
Dutta, M. J. (2007), 'Health information processing from television: The role of health orientation', *Health Communication*, **21**(1): 1–9.
Ekman, P. and W. V. Friesen (1976), 'Measuring facial movement', *Environmental Psychology and Nonverbal Behavior*, **1**(1): 56–75.

Ekman, P., R. J. Davidson and W. V. Friesen (1990), 'The Duchenne smile: Emotional expression and brain physiology: II', *Journal of Personality and Social Psychology*, **58**(2): 342.

Ellison, B., J. L. Lusk and D. Davis (2013), 'Looking at the label and beyond: The effects of calorie labels, health consciousness, and demographics on caloric intake in restaurants', *International Journal of Behavioral Nutrition and Physical Activity*, **10**(1): 21.

Fahim, M. A. A., M. M. H., Khan, T., Jensen, Y., Albayram, E., Coman and R. Buck (2019), 'Effect of feedback on users' immediate emotions: Analysis of facial expressions during a simulated target detection task', in *2019 International Conference on Multimodal Interaction*, 49–58.

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

Gabert-Quillen, C. A., E. E. Bartolini, B. T. Abravanel and C. A. Sanislow (2015), 'Ratings for emotion film clips', *Behavior Research Methods*, **47**(3): 773–787.

Girard, J. M., J. F., Cohn, M. H., Mahoor, S., Mavadati and D. P. Rosenwald (2013), 'Social risk and depression: Evidence from manual and automatic facial expression analysis', in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, IEEE, 1–8.

Harnack, L. J. and S. A. French (2008), 'Effect of point-of-purchase calorie labeling on restaurant and cafeteria food choices: A review of the literature', *International Journal of Behavioral Nutrition and Physical Activity*, **5**(1): 51.

Harsanyi, J. C. (1985), 'Rule utilitarianism, equality, and justice', *Social Philosophy and Policy*, **2**(2): 115–127.

Hattke, F., D. Hensel and J. Kalucza (2020), 'Emotional responses to bureaucratic red tape', *Public Administration Review*, **80**(1): 53–63.

Höfling, T. T. A., A. Gerdes, U. Föhl and G. W. Alpers (2020), 'Read my face: Automatic facial coding versus psychophysiological indicators of emotional valence and arousal', *Frontiers in Psychology*, **11**: 1388.

Imotions (2017), Facial Expression Analysis. Retrieved from: https://imotions.com/guides/facial-expression-analysis/

Jung, M. (2016), 'Implications of graphic cigarette warning labels on smoking behavior: An international perspective', *Journal of Cancer Prevention*, **21**(1): 21.

Kahneman, D., A. B. Krueger, D. Schkade, N. Schwarz and A. A. Stone (2006), 'Would you be happier if you were richer? A focusing illusion', *Science*, **312**(5782): 1908–1910.

Kiszko, K. M., O. D. Martinez, C. Abrams and B. Elbel (2014), 'The influence of calorie labeling on food orders and consumption: A review of the literature', *Journal of Community Health*, **39**(6): 1248–1269.

Koivumaa-Honkanen, H., R. Honkanen, H. Viinamaeki, K. Heikkilae, J. Kaprio and M. Koskenvuo (2001), 'Life satisfaction and suicide: A 20-year follow-up study', *American Journal of Psychiatry*, **158**(3): 433–439.

Krueger, A. B. and D. A. Schkade (2008), 'The reliability of subjective well-being measures', *Journal of Public Economics*, **92**(8–9): 1833–1845.

Krumpal, I. (2013), 'Determinants of social desirability bias in sensitive surveys: A literature review', *Quality and Quantity*, **47**(4): 2025–2047.

Kulke, L., D. Feyerabend and A. Schacht (2020), 'A comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for identifying facial expressions of emotion', *Frontiers in Psychology*, **11**: 329.

Lagerkvist, C. J., J. Okello, P. Muoki, S. Heck and G. Prain (2016), 'Nutrition promotion messages: The effect of information on consumer sensory expectations, experiences and emotions of vitamin A-biofortified sweet potato', *Food Quality and Preference*, **52**: 143–152.

Loewenstein, G. and N. Chater (2017), 'Putting nudges in perspective', *Behavioural Public Policy*, **1**(1): 26.

Loewenstein, G. and T. O'Donoghue (2006), '"We can do this the easy way or the hard way": Negative emotions, self-regulation, and the law', *The University of Chicago Law Review*, **73**(1): 183–206.

Murphy, J. J. and T. H. Stevens (2004), 'Contingent valuation, hypothetical bias, and experimental economics', *Agricultural and Resource Economics Review*, **33**: 182–192.

Nelson, N. L. and J. A. Russell (2013), 'Universality revisited', *Emotion Review*, **5**(1): 8–15.

OECD (2017), *Behavioural insights and public policy: Lessons from around the world*. Paris: OECD Publishing. https://doi.org/10.1787/9789264270480-en.

Oliver, A. (2017), 'Distinguishing between experienced utility and remembered utility', *Public Health Ethics*, **10**(2): 122–128.

Rafiq, S. (2021), 'How much is a calorie worth? A study of willingness to pay for calorie labels in restaurant menus', *Behavioural Public Policy*, **5**(3): 354–377.

Rhue, L. (2018), Racial influence on automated perceptions of emotions. Available at *SSRN 3281765*.

Schneider, L. and U. Schimmack (2009), 'Self-informant agreement in well-being ratings: A meta-analysis', *Social Indicators Research*, **94**(3): 363.

Spence, A. and N. Pidgeon (2010), 'Framing and communicating climate change: The effects of distance and outcome frame manipulations', *Global Environmental Change*, **20**(4): 656–667.

Srinivasan, R. and A. M. Martinez (2018), 'Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the wild', *IEEE Transactions on Affective Computing*.

Steptoe, A., J. Wardle and M. Marmot (2005), 'Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes', *Proceedings of the National Academy of Sciences*, **102**(18): 6508–6512.

Stöckli, S., M. Schulte-Mecklenbeck, S. Borer and A. C. Samson (2018), 'Facial expression analysis with AFFDEX and FACET: A validation study', *Behavior Research Methods*, **50**(4): 1446–1460.

Stok, F. M., K. T. Verkooijen, D. T. de Ridder, J. B. de Wit and E. de Vet (2014), 'How norms work: Self-identification, attitude, and self-efficacy mediate the relation between descriptive social norms and vegetable intake', *Applied Psychology: Health and Well-Being*, **6**(2): 230–250.

Stone, A. A. and C. E. Mackie (2013), *Subjective well-being: Measuring happiness, suffering, and other dimensions of experience*. Washington, DC: National Academies Press.

Sunstein, C. R. (2019), 'Ruining popcorn? The welfare effects of information', *Journal of Risk and Uncertainty*, **58**(2–3): 121–142.

Teixeira, T., M. Wedel and R. Pieters (2012), 'Emotion-induced engagement in internet video advertisements', *Journal of Marketing Research*, **49**(2): 144–159.

Thaler, R. H. and C. R. Sunstein (2009), *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, Connecticut: Penguin.

Thunström, L. (2019), 'Welfare effects of nudges: The emotional tax of calorie menu labeling', *Judgment and Decision Making*, **14**(1): 11.

Troussard, X. and R. van Bavel (2018), 'How can behavioural insights be used to improve EU policy?' *Intereconomics*, **53**(1): 8–12.

US Food and Drug Administration (2014), *Food Labeling; Nutrition Labeling of Standard Menu Items in Restaurants and Similar Retail Food Establishments*.

Wilson, T. D. and D. T. Gilbert (2005), 'Affective forecasting: Knowing what to want', *Current Directions in Psychological Science*, **14**(3): 131–134.

Appendix

Table A1. Descriptive statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	101	27.04	9.01	19	59
Female	101	0.63		0	1
University degree	101	0.77		0	1
Best possible health	101	0.78		0	1
Treated	101	0.48		0	1
Choose popcorn	101	0.44		0	1
Negative emotion choice	101	2.749	3.615	0	18.864
Negative emotion Wall-E	101	6.157	16.188	0	83.333
Negative emotion Bobby	101	3.318	3.193	0.008	15.77

Table A2. Balance checks.

Variable	Treatment	Control	p-value
Age	27.5	26.6	0.6
Female	60.4	66.7	0.51
University degree	77.4	77.1	0.97