School of Economics
University of East Anglia
Norwich NR4 7TJ, United Kingdom

‘As judged by themselves’: Do people really want to be nudged towards healthy lifestyles?

Robert Sugden
University of East Anglia

Paper to be presented at conference on Economics, Health and Happiness
Lugano, 14–16 January 2016

It's generally known that mortality and morbidity are strongly affected by lifestyle choices. In particular, people can expect longer, healthier lives if:

-- their diets are low in saturated fats, salt and sugar, and include a lot of fruit and vegetables;

-- they don't smoke;

-- they don't drink more than small amounts of alcohol;

-- they take regular exercise.

But in fact, many people don't follow this advice.

Behavioural economists often argue that governments should respond by using **healthy lifestyle nudges**.

I.e. construct choice environments to engage with 'non-rational' psychological mechanisms so as to make healthier choices more likely, while not restricting choice opportunities.

This paper is about one common justification for healthy-lifestyle nudges – **that individuals want to make the choices that that they are being nudged towards**.

My interest in this: I am trying to develop a general form of normative economics that is compatible with behavioural findings.

So I want to understand and appraise the principles that behavioural economists use when they do normative analysis. This isn't easy, because behavioural economists tend to be very casual in normative argument.

One fixed point is that when behavioural economists justify nudge policies, the claim that people want to be nudged is often centre-stage.

I want to ask:

-- What exactly does this claim mean?

-- Is it internally coherent?

-- Is it consistent with evidence?

-- Is it consistent with the policy recommendations it is supposed to justify?

This paper is not concerned with other arguments for nudging, such as:

-- Individuals' welfare would be greater with healthier lifestyles, whatever they may believe.

-- Nudges don't interfere with choice, so there is no need to justify them.

-- Public policy favours the adoption of healthy lifestyles. The job of the economist is to find out 'what works' in achieving that objective. No further justification is needed.

-- One person's poor health imposes external costs on others.

… **or with arguments against nudging**, such as:

-- Nudges are ineffective in the long run.

-- Nudges compromise individual autonomy.

(I have views about these other arguments, but let's take one issue at a time!)

I'll focus on the arguments by Sunstein and Thaler in *Nudge* (2008). These have been very influential; and similar arguments are used by other leading behavioural economists.

[See: Infante, Lecouteux and Sugden, 'Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics', forthcoming in *Journal of Economic Methodology*.]

Sunstein and Thaler go out of their way to claim that the criterion for nudging is to make individuals better off, as judged by themselves…

This is said explicitly in the opening chapter: their recommendations for nudging are designed to 'make choosers better off, *as judged by themselves*' (p. 5; italics in original).

The italicised clause is repeated (with minor variations) at other points in the book:

'So long as people are not choosing perfectly, some changes in the choice architecture could make their lives go better (**as judged by their own preferences, not those of some bureaucrat**)' (p. 10).

Using the distinction between 'Econs' (i.e. individuals as modelled in neoclassical economics) and 'Humans':

'Some firms sell cigarettes; others sell products that help you quit smoking. Some firms sell fast food; others sell diet advice. If all consumers are Econs, there is no reason to worry about which of these competing interests wins. But if some of the consumers are Humans who sometimes make bad choices (**as judged by themselves, of course**), then all of us may have an interest in which set of firms wins the battle' (p. 80).

Sunstein and Thaler see this clause as important. E.g. In his book *Misbehaving: The Making of Behavioral Economics* (2015, pp. 326–326), Thaler emphasises 'a point that critics of our book [*Nudge*] seem incapable of getting'.

The point is that S&T 'have no interest in telling people what to do. **We want to help them achieve their *own* goals**'.

Thaler points to the '*as judged by themselves*' clause:

'The italics are in the original but perhaps we should also have used bold and a large font, given the number of times we have been accused of thinking that we know what is best for everyone. … **We just want to reduce what people would themselves call errors**.'

It seems clear that the argumentative purpose of the clause is to head off criticisms that nudge policies are unacceptably paternalistic.

But what exactly does the clause mean?

Immediately after the remark about making choosers better off, as judged by themselves, S&T say they will show that:

'in many cases, individuals make pretty bad decisions – decisions that they would not have made if they had **paid full attention** and **possessed complete information, unlimited cognitive abilities, and complete self-control'**.

Such decisions are 'inferior decisions in terms of their [i.e. the individuals'] own welfare'.

The most obvious interpretation is that S&T's normative criterion is individuals' judgements about their own welfare, and that these are revealed in the decisions that individuals would make in the absence of error, i.e. with full attention, complete information, unlimited cognitive ability and complete self-control.

So the normative economist's job is to reconstruct individuals' latent (i.e. error-free) preferences and try to satisfy these.

This approach ('behavioural welfare economics') is widely used in behavioural economics.

With Infante and Lecouteux, I have tried to reconstruct the logic of this approach and subjected it to a methodological critique.

We argue that this approach works as if a human being has an inner rational agent (with integrated preferences), which can interact with the world only through an error-prone psychological shell – i.e. the Human as a faulty Econ. But…

-- This model lacks psychological foundations (the existence and rationality of latent preferences is assumed, not explained).

-- Although latent preferences are supposed to be subjective, we are given no independent criterion for identifying errors and hence for identifying a person's latent preferences (by 'purification', i.e. abstraction of error).

Thus, the claim that behavioural welfare economics uses individuals' own judgements about their welfare turns out to have little content.

Notice that this claim is a typical feature of paternalist arguments! E.g. the parent who requires the child to eat up his vegetables. The parent says that the child would recognise that vegetables were good for him, if he had sufficient knowledge, reasoning ability, attention and self-control.

Implication: if the 'as judged by themselves' clause is read in terms of the inner rational agent model, it isn't a convincing response to the criticism that nudge policies are unacceptably paternalistic.

So is there any other way of reading it? Possibly…

Alternative reading:

Latent preferences are not a hypothetical construct (what the individual would prefer if ideally rational); they are the preferences that the individual actually endorses in some independently definable circumstances, and which in some sense he continues to acknowledge even when he doesn't act on them.

In other words: this is a model of akrasia/ failure of self-control.

This reading would fit better with the argumentative purpose of the 'as judged by themselves' clause.

But remember S&T's characterisation of error-free decision making in terms of 'full attention, … complete information, unlimited cognitive abilities, and complete self-control'. If we are to interpret latent preferences as actually-endorsed preferences, I think we have to drop the conditions of complete information and unlimited cognitive ability.

Let's look at what S&T say about healthy-lifestyle nudges, and try to discover what they mean by 'as judged by themselves'.

As part of their discussion of the difference between Humans and Econs:

'Consider the issue of obesity. Rates of obesity in the United States are now approaching 20 per cent, and more than 60 per cent of Americans are considered either obese or overweight. There is overwhelming evidence that obesity increases risks of heart disease and diabetes, frequently leading to premature death. It would be quite fantastic to suggest that everyone is choosing the right diet, or a diet that is preferable to what might be produced with a few nudges. … We do not claim that everyone who is overweight is necessarily failing to act rationally, but we do reject the claim that all or almost all Americans are choosing their diet optimally… With respect to diet, smoking, and drinking, people's current choices cannot reasonably be claimed to be the best means of promoting their well-being. Indeed, many smokers, drinkers and overeaters are willing to pay third parties to help them make better decisions' (p. 7).

Notice that, apart from the final sentence (about the small minority of people who pay others to help them make decisions about diet etc.), the argument makes no reference to individuals' own judgements about their welfare. The argument seems to be:

The 60+ per cent of Americans who choose unhealthy lifestyles are not maximizing their welfare on any 'reasonable' criterion of welfare.

A fully rational person (= an Econ) would maximise some reasonable criterion of welfare.

Therefore: these people must be making errors.

If this argument is supposed to be consistent with the 'as judged by themselves' clause, that clause has no real content. The appeal is to 'reasonable' judgement, not people's actual judgements.

S&T on ‘self-control problems’ (pp. 40–42):

The opening example is of the plate of cashew nuts at Thaler’s dinner party. The guests are eating so many nuts that Thaler thinks their appetites for the meal will be spoiled. He moves the nuts out of sight, and immediately the guests thank him.

S&T say this is an example of temptation. The implication is that the guests are made better off by a nudge, and they want to be nudged. This fits the self-control reading of ‘as judged by themselves’.

Another example…

‘Tom is on a diet and agrees to go out to a business dinner, thinking he will be able to limit himself to one glass of wine and no dessert. But the host orders a second bottle of wine and the waiter brings the dessert cart….’

S&T say this is an example of the hot/cold empathy gap (Loewenstein), i.e. the failure of Cold Tom (Tom before the dinner) to empathise with Hot Tom’s (Tom at the dinner) responses to the cues of the restaurant environment.

Cold Tom wants to go to the dinner and have one glass of wine and no dessert. Looking ahead, he thinks of the behaviour of Hot Tom as an error, and one he would not make. Perhaps even when he is drinking the second glass and eating the dessert, he has some sense that Cold Tom’s preferences are ‘really’ his – in which case, this fits the self-control reading of ‘as judged by himself’.

More clues in the chapter ‘When do we need a nudge?’

This begins with the ‘New Year’s resolution test’:

‘[I]t seems safe to say that not many people are resolving on New Year’s Eve to floss less next year and to stop using the exercise bike so much. …[H]ow many people vow to smoke more cigarettes, drink more martinis, or have more chocolate donuts in the morning next year? Both investment goods [e.g. dental care, exercise] and sinful goods [e.g. cigarettes, alcohol] are prime candidates for nudging.’ (p. 73)

The implication seems to be that New Year’s resolutions provide evidence about people’s ‘as judged by themselves’ preferences.

Perhaps… But here the argument is getting thin. S&T are privileging preferences that are expressed on a special occasion with peculiar cues (including cues that favour making resolutions) over preferences that people act on consistently throughout the year.

Some other examples in this chapter don't fit the self-control interpretation at all well:

'Unfortunately, some of life's most important decisions do not come with many opportunities to practice. Most students choose a college only once. Outside of Hollywood, most of us choose a spouse, well, not more than two or three times… [R]are, difficult choices are good candidates for nudges' (pp. 74–75)

'Someone can eat a high-fat diet for years without having any warning signs of a heart attack. When feedback does not work, we may benefit from a nudge' (p. 75)

These are supposed to be cases in which people make errors because of lack of opportunities to learn the effects of alternative choices. But then it's hard to claim that people are conscious at the time that their choices are errors. (E.g. the person on the high-fat diet is not thinking that this diet is bad for his heart.)

Thinking more about the case of healthy and unhealthy diets…

What kinds of argument would be needed to justify nudges towards healthy diets?

It would be implausible to claim that this is a case in which the problem is lack of information. Most people (certainly in Britain) know the main dietary recommendations of health experts, along with simple formulae for acting on them (e.g. 'five portions of fruit or vegetables a day', 'not more than 14 units of alcohol per week).

So how do we explain why so many people fail to follow these recommendations?

A thought experiment …

Consider Jane, one of the customers at Sunstein and Thaler's cafeteria.

Jane is currently in good health but seriously overweight. Professional dieticians would agree that her long-term health prospects would be better if she ate less high-fat, high-sugar food (e.g. the cafeteria's cream cake) and more fruit and vegetables (e.g. the cafeteria's fresh fruit option). But, despite knowing what dieticians advise, she usually chooses the cake.

S&T's recommendation is that Jane should be nudged towards the fruit. Their claim is that Jane knows that her cake-eating is an error. ['We just want to reduce what people would themselves call errors.']

If the 'as judged by themselves' clause is to be taken seriously, this recommendation has to be compatible with whatever reasons Jane finds adequate to explain why, despite knowing what dieticians advise, she usually chooses cake.

Suppose we use a questionnaire approach….

Imagine a questionnaire which asks: **'Which of the following responses best represents your reasons for choosing cake rather than fruit, contrary to the recommendation of health experts?'**

The thought experiment is to think of statements that Jane might plausibly assent to, and to diagnose the modes of reasoning, heuristics or biases that they reveal. There should be one that picks up failure of self-control:

(a) I always go into the cafeteria having resolved not to choose cake, but when I see the cake at the front of the counter, I can't resist the temptation.

Now for some others:

(b) I get a lot of pleasure from eating sweet and fatty foods, and the thought of living to a great age doesn't appeal to me.

The first part is a plausible taste, but the second part may pick up a young/old empathy gap, combined with lack of experience of old age (young people under-estimate the self-reported happiness of the old).

(c) When I am a few years older I will adopt a healthier diet, so my current eating habits are not a problem.

Procrastination, or a young/old or now/later empathy gap (I imagine that my older self will be less susceptible to temptation, or will get less pleasure from food, or will be more far-sighted, than me).

(d) The expert advice sets unrealistic standards. Most of the people I know eat at least as much sugar and fat as I do.

Social proof (i.e. the heuristic of matching your behaviour to others'); self-serving bias in selection of comparators.

(e) All my grandparents were thin but died relatively young. It's quite likely that I will die young too, whatever I eat.

Over-weighting of personal experience relative to base rates; perhaps self-serving bias in selection of comparators.

(f) All my grandparents were fat but lived to ripe old ages. It's quite likely that I will have a long life too, whatever I eat.

As for (e).

(g) Expert health advice is always changing; in a few years time, experts may be recommending high-fat, high-sugar diets.

Cognitive dissonance (i.e. adjusting your beliefs to make unwelcome facts less painful, as in 'sour grapes').

[But not quite as unreasonable as it might seem: Flegal et al., *Journal of the American Medical Association* 2013 report a massive meta-analysis which finds that overweight (but not obese) people are 6 per cent *less* likely to die in a given period than individuals of normal weight.]

(h) Whatever I eat, I put on weight, so for me there is no point in trying to eat healthier food.

Cognitive dissonance; excessive readiness to see oneself as a special case. Possibly outright self-deception (i.e. in fact I eat much more food than is recommended, but I claim not to).

The point of this thought experiment is that there are lots of ways in which a person can explain why she knowingly acts against well-grounded but unwelcome recommendations from experts. Failure of self-control is only one of these.

My conjecture: 'failure of self-control' is relatively rare as a self-ascribed explanation (remember we are not asking about actual explanations):

-- As the thought experiment suggests, there are many other potential 'explanations' which are common in everyday conversation.

-- These 'explanations' involve modes of reasoning/ heuristics/ biases that are known to be properties of human psychology.

-- 'Failure of self-control' differs from all the others by admitting to error, which most people probably prefer not to do.

-- 'Failure of self-control' invokes a relatively sophisticated (and not necessarily true) dual-self theory of mind.

Implication: even if we (= behavioural economists, social planners) feel confident that people's lifestyle choices are based on some form of error, we shouldn't jump to the conclusion that the error is a failure of self-control.

So from the fact that a person's choices seem not to maximise any reasonable measure of welfare, we are not entitled to infer that the person is making an error as judged by herself. (To the contrary: one might expect it to be a general characteristic of errors of reasoning that the erroneous reasoning feels correct to the person who is reasoning!)

This raises the question: Why are behavioural economists predisposed to explain deviations from neoclassical rationality by using models of failures of self-control, rather than of all the other psychological mechanisms that are known to exist? I suggest two reasons:

(1) This strategy allows behavioural economists to use existing theoretical tools of economics and game theory, rather than having to develop more psychologically-based theories. Dual-self reasoning is more sophisticated than simple maximising. This may make it less psychologically plausible, but more interesting to economic theorists.

(2) Attributing 'irrational' behaviour to failures of self-control allows paternalistic interventions to be justified as satisfying preferences that individuals consciously endorse, even when they don't act on them.

But are these good reasons? I think not.

Thank you for listening.