REINFORCEMENT LEARNING, COGNITIVE SCARCITY AND BEHAVIOURAL BAND-AIDS

RACHEL KANE

SENIOR FRESH

"Behavioural interventions such as the 'nudge' have become increasingly popular in the context of health-related choices. While the use of this policy tool relies upon the existence of bounded rationality, remarkably little attention is given to the variables underpinning bounded rationality and the wider decision-making and learning processes. In this essay, Rachel Kane combines perspectives from behavioural and computational neuroscience to explore the role of cognitive load in the socioeconomic gradient of obesity. Kane concludes that the obesity nudge is a mere behavioural band-aid to this complex policy issue; it treats the symptoms of excessive caloric consumption and physical inactivity, rather than attempting to treat the problem at its root through the relaxation of cognitive constraints."

Introduction

'Felix, qui potuit rerum cognosere causas'

'Fortunate, who was able to know the causes of things' (Virgil, 29 BC)

The policy 'nudge' is a behavioural intervention that aims to improve individual welfare by encouraging the selection of a particular choice by strategically manipulating the existing choice architecture. First popularised by Thaler and Sunstein, nudges find their justification in bounded rationality; the limitations of both human knowledge and computational capacity arising from the widespread existence of cognitive biases and heuristics (Simon, 1990; Thaler and Sunstein, 2008). According to Thaler and Sunstein (2003), these constraints on rationality cause individuals to make 'inferior decisions in terms of their own welfare' and necessitate state intervention.

Unlike traditional incentive or law-based regulation, nudges are said to be functions of libertarian paternalism insofar as they do not punish nor forbid the selection of a particular choice. Instead, nudges harness behavioural insights to make the selection of 'better' choices easier. Sunstein (2013) takes the view that modern governments should 'make people's lives easier and get rid of unnecessary complexity'. While this is an honourable goal, nudges are seriously limited in their ability to achieve this due to their ends-focused nature. This essay argues that nudges aiming to tackle obesity are mere *behavioural band-aids* to such a complex policy issue. They fail to account for certain mediating variables in the individual decision-making and learning processes and are thus incapable of generating significant and sustained behavioural changes in this area.

Bounded Rationality as Justification for Nudges

Neoclassical microeconomic theory rests upon the *assumption of homo economicus*; the idea that individuals are utility-maximisers who possess unbounded rationality, willpower, and selfishness (Elahi, 2015). The Behavioural School rejects this positive description of behaviour and instead assumes that cognitive biases and heuristics lead to serious and systematic errors in the decision-making process (Tversky and Kahneman, 1974). These breakdowns in the individual decision-making process, often labelled as 'reasoning failures', result in a divergence between some *theorised* preference-consistent behaviour and the observed (and supposed preference-inconsistent) behaviour (Le Grand, 2022). As the view of normative economics is that the economy should ensure the maximal satisfaction of individuals' preferences, this failure to act in accordance with one's preferences creates 'behavioural market failures' which necessitate a regulatory response in the view of nudging advocates (Sugden, 2017; Bubb and Pildes, 2014).

However, despite their rejection of the Neoclassical positive description of behaviour, advocates of nudges maintain some form of rational behaviour as a normative criterion for policymaking. As the key success metric of a nudge is the degree to which the divergence between the optimal and observed behaviour is minimised, a definition of this ideal behaviour is necessary. The maintenance of this postulate raises serious issues, both ethically and practically in the measurement of 'unrevealed preferences'. Instead of regarding themselves as responsible for inferring and satisfying the latent preferences of the individual, obesity policymakers must detach themselves from the notion of directly interfering with outcomes and instead concentrate their efforts on designing optimal policy responses that enable individuals to make their own 'optimal' choices. The first step in designing such policy responses is understanding the role of various cognitive processes that underpin decision-making and learning, such as executive function.

Obesity Nudges

The prevalence of obesity presents serious threats to the health and well-being of societies and incurs substantial economic costs. In the developed world, there is a socioeconomic gradient to obesity, with individuals from lower socioeconomic (SES) backgrounds possessing a greater probability of being obese than their higher SES counterparts (Bickel et al., 2014). As obesity can act as a poverty trap, obesity nudges have been particularly focused on eliminating childhood obesity by aiming to make exercise for children more enjoyable, providing nutrition information to parents, and introducing a 'fat tax' (Seeman, 2011; Oliver and Ubel, 2014).

While obesity is a complex and multifactorial issue, it fundamentally occurs due to the consistent overconsumption of calories and consistent physical inactivity. In the nudging literature, the most common cognitive bias reported in obese populations is the hyperbolic delay discounting of rewards. Also known as present bias, hyperbolic delay discounting refers to the over-valuation of immediate rewards and the under-valuation of delayed rewards. A high discount rate γ , along with a high reinforcement value of food, has been positively correlated with obesity (Carr et al., 2011).

Despite the reliance of nudges on the existence and prevalence of cognitive biases and heuristics as grounds for their justification, the related literature says remarkably little about the variance of the strength of these biases across various cohorts, or indeed which variables mediate the strength of such biases.

Cognitive Scarcity and the Competing Neurobehavioural Decision Systems Model

One potentially pivotal factor that has been overlooked by the advocates of obesity nudges is cognitive scarcity, or the constraints on cognitive capacity due to the imposition of a cognitive load. Cognitive load (CL) refers to the amount of information that must be held and simultaneously manipulated at a given moment. A high CL places a tax on the cognitive bandwidth of the individual, which can result in negative spillover effects in other cognitive domains – it causes changes in judgments and decisions, and harms associative learning (Schilbach, Schofield and Mullainathan, 2016; Frank and Claus, 2006).

The Competing Neurobehavioural Decisions Systems model (CNDS) frames the individual decision-making process as a competition between the impulsive and executive function decision systems (Bickel et al., 2014). The impulsive decision system, located in the limbic and paralimbic regions of the brain, is responsible for the selection of immediate reinforcers, while the executive function (EF) system favours long-term outcomes and is responsible for planning and adjusting behaviour. A high CL, or reduction in working memory capacity, acts to dysregulate the balance between the dual systems. This impairs EF and constrains one's capacity to both establish and persevere with decisions that have healthy outcomes in the present and the future. When the balance is tilted towards the impulsive system, the present bias is exacerbated; the discount rate γ rises and more impulsive choices are selected. High CL is positively correlated with the risk of obesity and other negative health behaviours such as smoking and the use of alcohol and illicit drugs (Appelhans, 2009; Carr et al., 2011; Bickel et al., 2014).

CL is also linked to the SES gradient of obesity; lower SES individuals tend to have more scarce cognitive resources due to the stress, emotional distress and fatigue associated with the scarcity of various resources (Byrd-Bredbenner and Eck, 2020).

An Integrated Neurobehavioural and Neurocomputational Perspective on Obesity

While a high CL exacerbates present bias and leads to the selection of suboptimal health choices, obesity does not occur after the selection of one nor a small number of such poor choices; it is representative of an aggregation of the consistent selection of these poor choices over time. It follows then that obese individuals must experience a breakdown in the action-outcome associative learning process.

The Reinforcement Learning Framework

Reinforcement learning (RL) offers a valuable framework for conceptualising reward-decision choice processes as its algorithms harness neural insights and can explain many sophisticated aspects of human behaviour (Rmus, McDougle, and Collins, 2021). RL is an experienced-driven autonomous learning model where agents look to detect an optimal stochastic policy by approximating the expected utility of various actions. In this framework, the agent undertakes an action $a \in A$ within a state $s \in S$, transitioning them to a new state where the reward is received. The policy π gives the probability that a particular action is taken within the current state space. The reward function is a function of the current state and the action undertaken, given by $R(s, a)$. The goal of the agent is to maximise the discounted cumulative reward. The optimal state-action value function (Q-function) is given by $q_*(s, a) = max_\pi q_\pi(s, a)$. This must satisfy the Bellman optimality condition:

$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha(r_t + \gamma(max\ Q(s_{t+1}, a)))$$

This represents the immediate value of the move from state s_t to s_{t+1} and the discounted value of the state-action value of the state s_{t+1} , given the selection of some action a (Blackburn, 2020).

The RL computations have been shown to update estimates of expected values via reward prediction errors (RPEs). An appreciation of this computational framework for behaviour coupled with an understanding of the interactions between the EF and RL neurocognitive domains is vital for conceptualising the inhibited learning that is observed in obese populations, and subsequently designing solutions to improve it (Rmus, McDougle and Collins, 2021).

The EF-RL Relationship

According to Rmus, McDougle and Collins (2021), the executive functioning (EF) system contributes to the reinforcement learning computations in the brain. Thus, the imposition (or sustained imposition) of a high cognitive load may impair an agent's ability to learn from their behaviour. More specifically, EF plays a central role in 'setting the stage' for RL computations in defining the reward function and the value and probability estimation process.

The definition of the reward function $R(s, a)$ is underpinned by higher cognitive processes such as the computational and attentional components of EF. The dimensional computations of the state space, the assessment of transitional probabilities, and both the encoding and retrieval of action-reward associations require huge efforts from the EF system. In the case of probability estimation processes in particular, heuristics are relied upon in order to reduce the complexity of probability assessment. Although heuristics are treated, perhaps paradoxically, as both causes and symptoms of bounded rationality, heuristics in computational settings can indeed be seen as functions of resource rationality; the brain is choosing to optimise its choices given its computational constraints and preferences for trade-offs in accuracy over time (Korteling, Brouwer and Toet, 2018; Lieder, 2013).

However, this tendency to minimise computations and cognitive expenditure will lead to reductions in the accuracy of probability assessments; in choosing to perform fewer interactions and relying on sampling posterior distributions, results will be biased towards the initial value (Lieder, 2013; Tversky and Kahneman, 1974; Courville and Daw, 2007). This is known as the anchoring bias. This failure to update beliefs, and consequently change actions, is a failure in Bayesian reasoning. This is particularly harmful in the pursuit of tasks that involve continuous decisions of a conjunctive character, such as maintaining a healthy and active lifestyle. It has been shown that Bayesian reasoning is indeed reliant on EF functions and is harmed under the imposition of a high cognitive load (Yin et al., 2020). It should also be noted that EF taxation can act to increase the discount rate in the value estimation process, due to a loss in inhibitory control.

Thus, this taxation on EF that is so prevalent in obese and low SES populations may increase the reliance on or the strength of heuristics in probability and value estimation contexts, through an increased reliance on sampling posterior distributions. As the encoding of associations between a reward and its associated state space-action pair is also dependent on EF, the accuracy of such associations is diminished under high CL. Evidence of breakdowns in these reward associations can be seen in the neurobehavioural literature. The ventromedial and orbitofrontal cortices (OFC) are responsible for adapting behaviour, or learning, and maintain reward associations in working memory. The OFC is more flexible compared to the basal ganglia-dopamine (BS-DA) system as it weights the magnitudes of rewards and punishments more accurately, and can influence responses almost immediately (Frank and Claus, 2006). Hypometabolism of the OFC has been implicated in obese populations* (Volkow et al., 2008, 2009).

Model-based and Model-free RL

This increased reliance on heuristics and use of posterior probability distributions is consistent with model-free RL. Humans make decisions both based on prior experiences and forward planning through the cognitive mapping of tasks. Unlike in model-based RL, model-free RL (MF RL) agents do not engage in forward planning and instead rely on a set of stored value estimates. MF RL algorithms have lower computational and working memory demands, but are less responsive to change strategies and are slow to learn (Collins and Cockburn, 2020). Therefore, individuals with lower budgets of executive control, such as some obese individuals and low SES individuals, will substitute towards MF RL due to its low EF demands. The MB RL system is consistent with Kahneman's (2011) fast-thinking but error-prone System 1 and the habitual stimulus-response mechanism of the BG-DA system in Frank and Claus (2006).

Conclusion

This essay has presented clear evidence of a potential cognitive basis for obesity and its socioeconomic gradient. As the executive function system underpins the judgment, decision-making and reward associative learning processes involved in obesity, it is intuitive that policymakers should aim to tackle these underpinning factors and treat the problem at its root. Marginal improvements in this complex policy issue should not be overlooked, but nudges should not be a main policy tool for tackling obesity; in looking to simplify the process of optimal choice selection, advocates of obesity nudges have over-simplified their regulatory responses. The goal should be to relax the cognitive constraints on the individual in order to facilitate better autonomous learning. A relaxation of these constraints would not only have positive implications for obesity outcomes and

* Glucose metabolism is a key marker of normal brain function. Hypometabolism of the medial orbitofrontal cortex (mOFC) in obese individuals refers to a sub-normal level of neural activity in this region for these individuals. As this region is responsible for the assessment of rewards, and the OFC more generally is responsible for maintaining reward associations in working memory, hypometabolism would suggest a down-regulation of this area's functions and a reduction in the accuracy of reward assessments and the retention of such associations in obese individuals.

social mobility; it may have positive spillover into non-economic but important areas, such as daily life functioning and self-care activities.

As an alternative to the current nudging policymaking paradigm, this essay calls for the examination of the 'boosting' paradigm, which concerns itself with the protection of collective cognitive capital (Murphy, 2021). A full treatment of this topic is beyond the scope of this essay, but this framework provides a robust outline for the integration and application of brain and behavioural science into behavioural law. It defines collective cognitive capital, comprised of cognitive development, plasticity, reserve, and resilience, as a key policy metric for autonomy and the fulfilment of the potential of the individual and of society as a whole. The need to mobilise obesity policy beyond the existing paradigm of mere information provision and myopic nudges is urgent; we need durable solutions, not behavioural band-aids.

References

1. Appelhans, B.M. (2009) 'Neurobehavioral Inhibition of Reward-driven Feeding: Implications for Dieting and Obesity', *Obesity,* 17(4), pp. 640–647.
2. Bickel, W.K. et al. (2014) 'A Competing Neurobehavioral Decision Systems model of SES-related health and behavioral disparities', *Preventive Medicine,* 68, pp. 37–43.
3. Blackburn (2020) 'Reinforcement Learning : Markov-Decision Process (Part 2), Medium'. [online] Available at: https://towardsdatascience.com/reinforcement-learning-markov-decision-process-part-2-96837c936ec3 [Accessed: 28 February 2022].
4. Bubb, R. and Pildes, R.H. (2014) 'How Behavioral Economics Trims Its Sails and Why'. *SSRN Scholarly Paper ID 2331000*. Rochester, NY: Social Science Research Network.
5. Byrd-Bredbenner, C. and Eck, K.M. (2020) 'Relationships among Executive Function, Cognitive Load, and Weight-related Behaviors in University Students', *American Journal of Health Behavior,* 44(5), pp. 691–703.
6. Carr, K.A. et al. (2011) 'Reinforcement Pathology and Obesity', *Current Drug Abuse Reviews,* 4(3), pp. 190–196.
7. Collins, A.G. and Cockburn, J. (2020) 'Beyond simple dichotomies in reinforcement learning.', *Nature reviews. Neuroscience,* 21(10), pp. 576–586.
8. Courville, A.C. and Daw, N. (2007) 'The rat as particle filter', in Advances in Neural Information Processing Systems. Curran Associates, Inc. [online] Available at: https://proceedings.neurips.cc/paper/2007/hash/fba9d88164f3e2d9109ee770223212a0-Abstract.html [Accessed: 28 February 2022].

9. Elahi, K. (2015) 'Homo Economicus in Neoclassical Economics: Some Conceptual Curiosities about Behavioural Criticisms', *Homo Oeconomicus,* 32, pp. 23–51.

10. Frank, M.J. and Claus, E.D. (2006) 'Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal.', *Psychological Review, 113(2),* pp. 300–326.

11. Kahneman, D. (2011) 'Thinking, fast and slow'. Macmillan.

12. Korteling, J.E., Brouwer, A.-M. and Toet, A. (2018) 'A Neural Network Framework for Cognitive Bias', Frontiers in Psychology, 9 [online]. Available at: https://www.frontiersin.org/article/10.3389/fpsyg.2018.01561 [Accessed: 28 February 2022].

13. Le Grand, J. (2022) 'Some challenges to the new paternalism', *Behavioural Public Policy,* 6(1), pp. 160–171.

14. Lieder, F. (2013) 'Burn-in, bias, and the rationality of anchoring', p. 9.

15. Murphy, E.R. (2021) 'Collective Cognitive Capital'. *SSRN Scholarly Paper ID 4001849.* Rochester, NY: Social Science Research Network.

16. Oliver, A. and Ubel, P. (2014) 'Nudging the Obese: A UK-US Consideration Special Section', *Health Economics Policy and Law,* 9(3).

17. Rmus, M., McDougle, S.D. and Collins, A.G.E. (2021) 'The Role of Executive Function in Shaping Reinforcement Learning', *Current Opinion in Behavioral Sciences,* 38, pp. 66–73.

18. Schilbach, F., Schofield, H. and Mullainathan, S. (2016) 'The Psychological Lives of the Poor', *American Economic Review,* 106(5), pp. 435–440.

19. Seeman, N. (2011) 'Move If U Wanna: Obama and the weight loss nudge', CMAJ : *Canadian Medical Association Journal,* 183(1), p. 152.

20. Simon, H.A. (1990) 'Bounded Rationality', in Eatwell, J., Milgate, M., and Newman, P. (eds) *Utility and Probability.* London: Palgrave Macmillan UK (The New Palgrave), pp. 15–18.

21. Sugden, R. (2017) 'Do people really want to be nudged towards healthy lifestyles?', *International Review of Economics,* 64(2), pp. 113–123.

22. Sunstein, C.R. and Thaler, R.H. (2003) 'Libertarian Paternalism Is Not an Oxymoron', *The University of Chicago Law Review,* 70(4), pp. 1159–1202.

23. Thaler, R.H. and Sunstein, C.R. (2008) 'Nudge: Improving Decisions about Health, Wealth, and Happiness'. Yale University Press.

24. Tversky, A. and Kahneman, D. (1974) 'Judgment under Uncertainty: Heuristics and Biases', *Science,* 185(4157), pp. 1124–1131.

25. Volkow, N.D. et al. (2008) 'Low dopamine striatal D2 receptors are associated with prefrontal metabolism in obese subjects: possible contributing factors', *NeuroImage,* 42(4), pp. 1537–1543.

26. Volkow, N.D. et al. (2009) 'Inverse association between BMI and prefrontal metabolic activity in healthy adults', *Obesity (Silver Spring, Md.)*, 17(1), pp. 60–65.

27. Yin, L. et al. (2020) 'The Effects of Working Memory and Probability Format on Bayesian Reasoning', Frontiers in Psychology, 11 [online]. Available at: https://www.frontiersin.org/article/10.3389/fpsyg.2020.00863 [Accessed: 28 February 2022].