

Research Briefing:

Ethical Challenges and Frameworks for Behavioural Public Policies

Research Summary

This briefing is provided through the research activities of the following project “Expertise and Ethics in Times of Crisis: Political controversies and ethical dilemmas of applying behavioural insights in the COVID-19 pandemic”. The project is funded by the University of Birmingham and the University of Bielefeld, 2020-21. It explores the question of how governments are engaging with ethical debates during the Covid-19 pandemic. The project builds on our previous research on the normative foundations and practical implementation of Behavioural Public Policies (BPP).

Investigators

Dr. Jessica Pykett, School of Geography, Earth and Environmental Sciences, University of Birmingham; Prof. Dr. Holger Straßheim, Faculty of Sociology, Bielefeld University (Principal Investigators); Dr. Sarah Ball Postdoctoral Fellow, School of Social and Political Sciences, The University of Melbourne; Dr. Robert Lepenies, Department of Environmental Politics, Helmholtz Centre for Environmental Research, Leipzig (Co-Investigators).

Suggested Citation: Ball, S., Pykett, J., Straßheim, H., Lepenies, R., (2021) *Research Briefing: Ethical Challenges and Frameworks for Behavioural Public Policies*. Working Paper, University of Birmingham and Bielefeld University.

The views expressed in this publication are those of the authors and not necessarily those of our institutions.

Contents

Overview of Ethical Challenges and Frameworks for Behavioural Public Policies	3
Figure 1. Summary of Ethical Frameworks proposed for Behavioural Public Policies and suggested responses	4
Discussion: Commonalities, Gaps and Implementation	9
Ethical Challenges	12
Figure 2. Distribution of Ethical Responsibilities	12
Figure 3: The DATA Framework for Ethics of BPP.....	13
Emerging Questions:.....	14
Review of Ethical Frameworks.....	15
A Bill of Rights	15
BASIC Toolkit.....	17
Nudge FORGOOD	19
Behavioural Scientist’s Ethics Checklist	22
Dilemmas in Applying Behavioural Insights.....	24
Ethics of Nudges.....	27
Constitution for Behavioral Policy-making	29
Nudging in Public Health.....	30

Overview of Ethical Challenges and Frameworks for Behavioural Public Policies

This briefing details eight recent frameworks that have been developed with the aim of addressing the ethical concerns about the use of Behavioural Public Policies in the global academic and policy literature. It provides details of core principles and practical guidance. It also serves to highlight where there are potential gaps. This is intended to assist in demonstrating what has already been done to manage the ethical concerns, but also provide grounds to address gaps and highlight new research priorities. The frameworks are summarised in figure 1 and a model for thinking through the distribution of ethical responsibility is proposed in figure 2. The research briefing includes a discussion and a series of 'emerging questions' before providing the full review of the frameworks.

Figure 1. Summary of Ethical Frameworks proposed for Behavioural Public Policies and suggested responses

		Ethical Frameworks							
		Bill of Rights: Sunstein & Reisch (2019)	BASIC: OECD (2019)	FORGOOD: Lades & Delaney (2019)	BS Ethical Framework: Jachimowicz et al (2017)	Ethical Dilemmas: Pykett & Johnson (2015)	Ethics of nudges: Clavien (2018)	Constitution: Fabbri & Faure (2018)	Public Health Framework: Department of Health, Irish Government (2015)
Ethical Challenges	Transparency	Nudges should be transparent rather than hidden	Prioritise transparency, offer a way out	Practitioners should consider 'Openness'	Transparent research process	Practitioners should ask "Is the intervention open to challenge?"			
	Public acceptability	Nudges must be consistent with people's values and interests	Refrain from targeting/changing behaviours that cannot be defended as being in the public interest or aligned with government priorities	Practitioners should explore citizen's 'Opinions'	Maintain data privacy/security			Meaningful public participation	
	Risk of manipulation of citizens/misuse	Nudges must not manipulate people	Protect individual rights, values and liberties when targeting behaviour change				Practitioners should ask "What are the side effects? Can this be measured? If so, how?"		Practitioners should consider the risk of discrimination. Policymakers should strive to minimise any identified burdens to

									individuals or specific communities
	Scientific pluralism		Use multiple forms of analysis. Triangulate if possible and use observational/ in situ methods over other forms of qualitative data collection			Practitioners should ask “Is what works the best thing to do?”			
	Legitimacy	Public officials must promote legitimate ends		Practitioners should consider the ‘Goals’ of the intervention Practitioners should consider whether they have the ‘Delegation’	Aligned interests (but problematic)		Practitioners should ask “Are there conflicts of interest? Are the goals justifiable? Are they selfish, social or nudge-driven? Are they qualified?”		Goals generally ought to be framed in terms of public health improvement As a rule of thumb, the greater the burdens posed by a specific intervention the stronger the evidence must be to demonstrate that the intervention will achieve its goals
	Risk of expert/government bias		Always evaluate the existing evidence for targeting a given behaviour change			Practitioners should ask “What is the legitimacy of the behavioural expert?”	Practitioners should ask “What behavioural biases and heuristics are influencing the nudger?”		

	'Technocracy'					Practitioners should ask "Is what works the best thing to do?"			
	Role of deliberation/ coproduction					Practitioners should ask "What is the legitimacy of the behavioural expert?"		Meaningful public participation	
	Welfare trade-offs – individual vs social		Ensure citizens are not being held responsible for consequences that they did not consciously select	Practitioners should consider the 'Fairness' of the intervention	Cost benefit analysis	Practitioners should ask, might there be "Unintended consequences?"			Practitioners should consider the proportionality of the intervention i.e. cost-benefit analysis If public health nudging policies aspire to reduce health inequalities, then they should include measures specifically aimed at improving the quality of life of particular individuals and/or groups who are disadvantaged and/or at risk
	Incrementalism/ Structural level policy tools vs 'trivial' nudges					Practitioners should ask, might there be "Unintended consequences?"			

	Autonomy	Nudges must respect individual rights	Protect individual rights, values and liberties when targeting behaviour change	Practitioners should ensure the interventions are 'Respectful'	Ease of opt out	Practitioners should ask "Is the intervention open to challenge?"	Practitioners should ask "How will autonomy be respected?"	Meaningful public participation	Practitioners should consider autonomy
	Segmentation/ Power relations/ inequality		Ensure justice, fairness and distributional impacts are considered.			Practitioners should ask, might there be "Unintended consequences?"			Practitioners should consider the issues of justice and equity Human vulnerability should be taken into account when considering the introduction of nudging policies The populations should share the benefits, risks and burdens of public health policies
	Myopia. Nudge vs alternative methods		Assess the avoidability of intervention.	Practitioners should make sure to consider other 'Options'		Practitioners should ask "Is behaviour the real problem to be tackled?"	Practitioners should ask "Is it an/the most effective way to achieve the aforementioned goal? What are the alternatives?"		
	Accountability		Follow principles of accountability.						Practitioners should ask "What role can the

			Publish results broadly.						concepts of trust and accountability play in the field of nudging, which works better “in the dark”?”
--	--	--	--------------------------	--	--	--	--	--	---

Discussion: Commonalities, Gaps and Implementation

Figure 1 outlines the different ethical challenges that these frameworks address. It also highlights several areas of focus and several gaps.

What are the commonalities between the ethical frameworks?

The concept of **transparency** was commonly shared across the frameworks. Most addressed the two different levels behavioural interventions could act upon – the reflective and automatic systems – and encouraged the use of reflective decision making. However, many also made allowances for less reflective interventions under ‘warranted’ circumstances usually premised on research (Department of Health, Ireland - Public Health Framework 2015; OECD 2019) or public acceptability (Lades & Delaney 2019; OECD 2019; Sunstein & Reisch 2019).

There was a strong alignment between transparency and autonomy. **Autonomy**, as a concept, took many forms, and some were less explicit than others about what it might actually mean in practice. Some simply talked about being respectful (Lades & Delaney 2019), others used the language of respecting individual rights (OECD 2019; Sunstein & Reisch 2019). Clavien (2018) and the Department of Health, Ireland - Public Health Framework (2015) simply suggests that policymakers consider autonomy.

There are also areas of alignment in how these frameworks respond to questions of **legitimacy** and **public acceptability**. This raises questions about whether they are, in fact, similar or whether government/policy makers use public acceptability as a litmus test for legitimacy. Sunstein and Reisch’s surveys about the acceptability of nudging serve as an example of how the acceptability of nudges is provided as justification for their legitimacy (2019).

What is missing from the ethical frameworks?

None of the frameworks explicitly raised questions about the risk of **manipulation** or misuse by decision makers. Perhaps because this touches on the complex issue of corruption. Based on the responses provided by these frameworks, by maintaining transparency and doing rigorous research you can minimise the risk of accidental misuse (Clavien 2018; Department of Health, Ireland - Public Health Framework 2015; OECD 2019; Sunstein & Reisch 2019). Explicit misuse would be a very different question. This speaks to a deeper ethical question about legitimacy. Who defines misuse and manipulation? By some ethical perspectives it may be acceptable to intervene for social good,

while others would see this as illegitimate. These types of **value-based questions** cannot be answered through research or transparency, but speak to deeper, ideological concerns.

This brings us to perhaps the most significant issue with these frameworks. For the most part, they simply raise questions that policy makers and practitioners *should* consider. There is limited engagement with the concepts of **accountability** or practical strategies to manage deeper debates or complexities. What happens if practitioners *don't* consider these things? Are there any real consequences or even risk of consequences? How could these be meaningfully implemented in a way that is transparent and creates accountability? Would policymakers be as interested in implementing BPP to the same extent if these boundaries were put into practice?

Some issues remain underdeveloped and under explored. The question of how BPP navigates questions regarding **individual vs social good** is critical, particularly in regards to public health, but the frameworks rarely go beyond the assertion that policy makers 'consider' fairness and proportionality. This also speaks to questions of **segmentation and inequality**. If BPP is designed for the many, how do policy makers and practitioners manage the complex issue of balancing the costs for the few? There is a risk of 'unintended consequences', but how can they be addressed (Pykett & Johnson, 2015)?

The importance of **deliberation and coproduction** was rarely addressed (cf. Pykett & Johnson, 2015). It was more common that this issue would be captured under ideas of public acceptability, with limited consideration about how this acceptability was to be explored.

How are the ethical frameworks to be implemented?

Compared to the other frameworks, Fabbri & Faure's Constitution approach is specifically focused on practical strategies to managing ethical issues, rather than the issues themselves. Their proposed independent decision making body and use of sunset clauses could be used, theoretically, to address many of the ethical challenges raised. A similar academic advisory panel was used in the UK for the Behavioural Insights Team (BIT), although **little is known** about this besides comments made by Peter John (2018)¹. It appears from early research a similar ethics panel existing for German decision makers as well. How are these used?

¹ John, P. (2018) How Far to Nudge. Assessing Behavioural Public Policy. Edward Elgar Publishing, Cheltenham.

The OECD BASIC framework is not specifically focused on the ethical ramifications so provides broader advice rather than specific recommendations. The value in this may be that it is a 'how-to' guide to using BPP in general which includes specific and often practical recommendations. However, similarly to Jachimowicz (2017), the **recommendations are often more closely aligned to basic expectations of ethical research such as the use of review boards, rather than the underlying complex ethical debates that accompany BPP.** One of the ongoing challenges with the ethical frameworks reviewed is that they assert the importance or necessity of ethical consideration, but practitioners and decision makers are very unlikely to ever encounter any ongoing scrutiny. As long as the nudge is publicly justifiable with reference to evidence and expertise then it is unlikely to raise significant concern.

A further finding is that the debates and responses regarding the threat of technocracy, the absence of **scientific pluralism** and the **risk of bias** in expert/policy maker decision making are all closely aligned. They all speak to issues of **authority**. Who decides whose preferences count? RCTs alone cannot answer that question as it is a question of appropriateness. Broader methods need to be used to support this, however, to date, **the idea of RCTs as a cornerstone of BPP has resonated far more than the arguments for greater citizen involvement and deliberation.** This commitment to quantitative, strict causal methods is well documented in policy making and it appears the practice of BPP is likely to have been shaped by this.

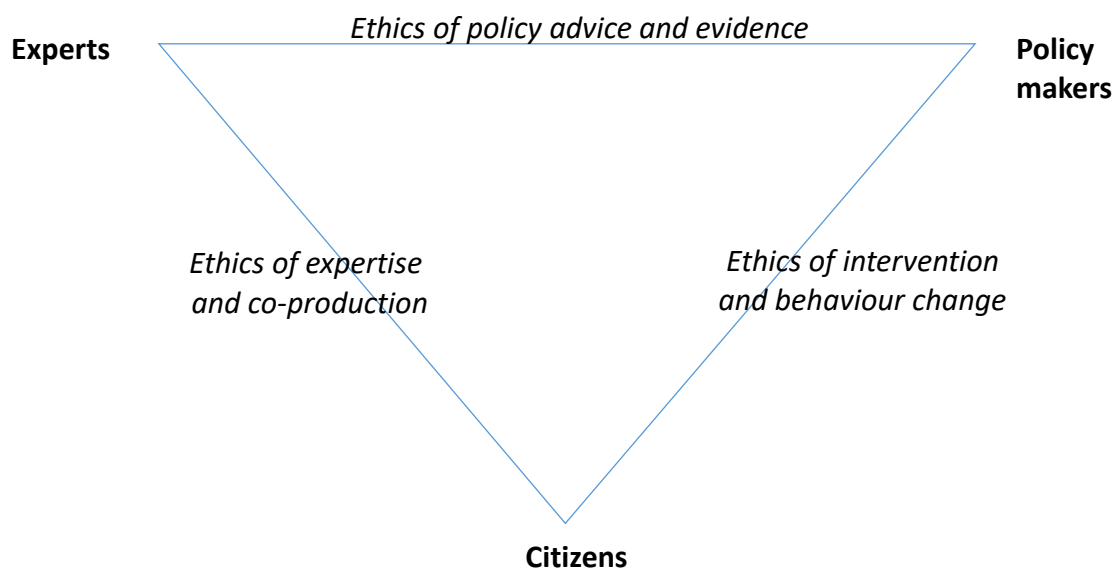
Many of the frameworks address the issue of **myopia**, namely that practitioners may push for a behavioural intervention when an alternative approach may be preferable. This issue sticks out for being a concern predominantly for government and policy makers. A practitioner, especially one seeking funded contracts, is unlikely to suggest alternatives. There are practical concerns to consider here. First, is the reliance on contracting and the 'cost effectiveness' of nudge in that environment. Second, capacity and capability within civil service to assess the 'best' approach to a problem. Third, the issue of **solutions chasing problems**. With the growth of nudge units, design teams etc there is value in considering how much decision makers are influenced by entrepreneurial "instrument constituencies" which refers to how policy instruments or tools are made up of: "knowledge about specific modes of governing is made and actualized by specific actors in concrete practices" (Simons & Voss, 2018:14)².

² Arno Simons & Jan-Peter Voß (2018) The concept of instrument constituencies: accounting for dynamics and practices of knowing governance, *Policy and Society*, 37:1, 14-35, [DOI: 10.1080/14494035.2017.1375248](https://doi.org/10.1080/14494035.2017.1375248)

Ethical Challenges

The design and implementation of BPP involves a specific distribution of ethical responsibilities. This can be summarised as in Figure 2.

Figure 2. Distribution of Ethical Responsibilities



Based on the review of ethical frameworks as summarised in Figure 1, key ethical challenges were raised in regards to behavioural policy. These should be considered for the activities represented at each side of the triangle – activities involving: **provision of policy advice and evidence; gathering expertise and co-production**, and **implementing behaviour change interventions**.

The main ethical challenges can be summarised as:

- Transparency
- Public acceptability
- Risk of manipulation of citizens/misuse
- Scientific pluralism
- Legitimacy
- Risk of expert/government bias
- Technocracy
- Role of deliberation/coproduction
- Welfare trade-offs – individual vs social

- Incrementalism/ Structural level policy tools vs ‘trivial’ nudges
- Autonomy
- Segmentation/ Power relations/inequality
- Myopia. Nudge vs alternative methods
- Accountability

One possible way to reconsider this is to condense the multiple challenges into a typology of four broad ethical dimensions. The metaanalysis of existing ethical frameworks can be summarised in the DATA framework:

Figure 3: The DATA Framework for Ethics of BPP

Diversity (or the absence of it) captures the challenges of myopia, scientific pluralism and the potential failure to consider alternatives to nudge. In BPP there is a risk of focusing exclusively on cognitive and affective behavioural science. This can be the source of significant ethical concerns.

Authority: Technocracy, the role of deliberation and the risk of government bias all point to a similar question: Which voices are involved in determining the suitability and legitimacy of the intervention? If the answer is experts or government alone raises important ethical questions. Do they have authority to define the problem? To determine the solution? To judge what is best for (a specific segmentation of) people?

Transparency should be considered across all dimensions, from problem definition, intervention design and evaluation. This would then capture discussions not only of the transparency of the intervention itself and, therefore, the ability for citizens to exercise autonomy it would also capture the transparency of the knowledge production and decision making process. This would include transparency regarding discussions of the welfare trade-offs and issues of segmentation.

Appropriateness captures broader questions to do with these ethical challenges. Policy makers and the experts that support the design and evaluation of BPP have, to date, directed most of their attention towards effectiveness of interventions. Appropriateness would capture all of these ethical questions and can easily be supported by measurements of effectiveness. It may work, but is it transparent, give citizens the freedom and autonomy to self-govern and meet broader social goals?

Emerging Questions:

There is currently a lack of research and evidence on the use and value of ethical frameworks too inform BPP. The following questions are proposed to inform their future exploration:

- Do practitioner's or policy makers use any of these resources? How?
- How is public acceptability understood?
- How are decisions about legitimacy made? What is a legitimate vs illegitimate goal?
- Are ethics panels routinely convened where BPP experiments are designed?

A Bill of Rights

[Sunstein, C., & Lucia, R. \(2019\). Trusting Nudges: A Bill of Rights for Nudging. Routledge.](#)

One of the original authors of Nudge (2008), Cass Sunstein, has invested a significant amount of energy in the discussion on the ethics of nudging. He has written several papers and books on the subject. In 2019, with co-author Lucia Reisch, he developed a Bill of Rights for Nudging.

Public officials must promote legitimate ends

All instruments, including nudges can serve legitimate or illegitimate ends. The authors recognise that what counts as legitimate is debatable and that their goal “is not, of course, to reach conclusions about what count as legitimate ends” (p131).

Nudges must respect individual rights

Because nudges respect freedom of choice (inherently according to Sunstein) rights violations can be less obvious. The example given, a nudge framing the design of ballot forms in order to prioritise the incumbent, demonstrates the importance of *judgment* in regards to what is and isn't acceptable.

Nudges must be consistent with people's values and interests

This appears to be predominantly about compliance – Sunstein and Reisch note that nudges are more likely to be effective when they are consistent with peoples values. Interestingly, they do also give public officials *leading* rather than *following* public opinion on this. The example points to nudging people not to discriminate against other ethnic groups or religions but one can imagine less benign examples. They note that “the question whether and when public officials should be able to depart from what (some or many) people believe to be their values or their interests is obviously a delicate one, turning on the grounds on which the departure might be justified” (p132).

Nudges must not manipulate people

The authors note here that manipulation is not easy to define, and that there is a long history of philosophical debate on the subject. They do note that, based on their research/survey, “in most of

the nations explored here, manipulation creates serious concern, and there is at least a presumptive principle against it” (p133). This is interesting given that many framing nudges could potentially be seen as manipulative.

Nudges should not take things from people, and give them to others, without their explicit consent

This points to the results of their research finding that default charitable donations or organ donation received widespread disapproval. People do not like to have things taken from them without the ability to provide explicit consent. Sunstein and Reisch point to tax and military service as interesting examples of when this principle does not seem to apply and therefore this principle “should be treated with considerable caution as we do not know its boundary conditions”(p133).

I wonder whether there is something here about tax and military service being universally applied or it being a strong mandate for social good? I imagine it may be possible for a country to mandate organ donation – with the possibility of exemption – if it wanted to. This may point to the following ‘right’.

Nudges should be transparent rather than hidden.

Sunstein and Reisch state that nudges, like other interventions, should not be hidden or covert. This is not only through the visibility of the nudge but also through explicit justification. They do caveat this by noting that during the testing of the intervention this may need to be waived. However, once the study is complete the results and the decision should be justified and transparent.

Additionally, they argue that policy-makers should consider the **welfare and autonomy** implications of nudges. This section points to some interesting ethical questions regarding autonomy i.e. Does autonomy require freedom of choice? Always? Does it require us to attend to the background conditions under which people form their preferences and values? If so, what exactly does that mean? Does it mean that we should dismiss or fail to respect judgments that are based on a lack of information or that come from a problem of self-control? Are some judgments non-autonomous? (p134-5) In response, they argue that defaults may need to be carefully monitored as they may be able to undermine autonomy even while preserving freedom of choice.

Ethical questions on welfare include, does welfare refer to utilitarianism, narrowly conceived? How shall we define “utility”? Does it refer to pleasures and pains? Does it make distinctions among

qualitatively different goods (a beach, a house, a dog, a friendship)? If we can answer such questions, how do we measure welfare? If it is a broader concept than utility (as many think), does it capture everything that ought to matter in human life? If wealthy people lose more than poor people gain, has welfare been reduced? Is that decisive? Cost benefit analysis may not be the most effective way to communicate welfare as monetizing various welfare effects can be a serious challenge (p135-6).

One way to address these concerns, say Sunstein and Reisch, is to engage the public to help in making the judgements (p137).

BASIC Toolkit

[OECD \(2019\) Tools and Ethics for Applied Behavioural Insights: The BASIC Toolkit.](#)

Moving away from complex academic debates about preferences, the OECD have developed the more practitioner friendly BASIC framework. This was also developed in public consultation with practitioners and government.

BASIC stands for Behaviour, Analysis, Strategies, Intervention, Change. The framework itself is intended to support the overarching design of a BI project. Each step of the framework has its own ethical considerations section. These guidelines are “intentionally both practical and aspirational – while some guidelines may not be implementable in every setting, they are intended to give the policymaker high standards to consider throughout a BI project” (p53). These include:

- Engaging, where possible, with an ethical review board from the beginning
- When undertaking analysis, follow recommended ethical guidelines i.e regarding consent, study purpose, safeguards for vulnerable populations
- Only collect data that is necessary
- Consider whether legal permission is required and demonstrate the necessity of the experiment
- Appoint someone with responsibility for data collection, use and storage

- Refrain from targeting/changing behaviours that cannot be defended as being in the public interest or aligned with government priorities
- Always evaluate the existing evidence for targeting a given behaviour change
- Always consider how to minimise potential side effects and protect individual rights, values and liberties when targeting behaviour change
- When assessing the transparency and “avoidability” of policy intervention, keep in mind the following considerations:
 - Prioritise transparency. Is your intervention clearly communicated, including being transparent about its purpose and nature?
 - Offer a way out. Can citizens avoid the intervention? Does the intervention offer easy pathways to objections and complaints?
 - Ensure the policy intervention serves the public interest. Is it in line with public sentiments? Does it prevent harm against others?
 - Ensure citizens are not being held responsible for consequences that they did not consciously select. In your context, are they able to fully understand the implications of their choices? Are they considered legally accountable for these?
- Always consult experience – but notably this refers not to citizens but experts in trial design and implementation
- Ensure justice, fairness and distributional impacts are considered
- Adhere to principles of proper stakeholder engagement. Make sure to *involve* public bodies, staff, citizens, businesses and other affected parties
- Follow principles of transparency and accountability. Results of experiments and consultations should be shared with executive and legislative branches, as well as with broader society
- Report on what works and what does not
- Monitor long-term and side effects
- Use multiple forms of analysis. Triangulate if possible and use observational/in situ methods over other forms of qualitative data collection. “No policy can truly be said to be behaviourally informed if the informant has not been there herself to observe through the lens of BI” (p68).

Despite the engagement with practitioners there is limited practice focus in regards to the ethical concerns. Who is responsible for what and when? How might these principles shape the intervention design? Where are the boundaries for when something should or shouldn't progress?

Nudge FORGOOD

[LADES, L., & DELANEY, L. \(2020\) Nudge FORGOOD. Behavioural Public Policy, 1-20.](#)

The FORGOOD framework is explicitly designed for the purpose of guiding policymakers. It is designed in a similar fashion to many of the products of BPP practitioners – using a mnemonic and focusing on creating something easy to understand and put into action. It aims to synthesize the debate on nudging to help practitioners design nudges “for good”, as advised by Thaler (2015). It presents 7 core elements for practitioners to consider: Fairness, Openness, Respect, Goals, Opinions, Options and Delegation. These are detailed below.

Fairness

- Does the behavioural policy have undesired redistributive effects?
- Does the behavioural policy focus too much on one group and neglect another group that is in more need of an intervention?
- Does the behavioural policy lead a subset of the population to behave against their preferences and best interests?
- Does the behavioural policy lead to a reallocation of resources?

Openness

- Is the behavioural policy open or hidden and manipulative?
- Is the automatic or reflective system triggered? i.e. is this a Type 1 or a Type 2 nudge?
- Does the behavioural policy have the potential to be manipulative?
- Does the public have the chance to scrutinize the behavioural policy?
- Is it possible for the person under the influence of the behavioural policy to identify the policy and its influence and impact?

Respect

- Does the policy respect people’s autonomy, dignity, freedom of choice and privacy?
- The authors note that Type 2 nudges are inherently more reflective and are therefore more likely to be respectful

- The authors provide some details on how to assess autonomy, dignity, freedom of choice and privacy.

However, these are not easy questions to answer. They require a clear ethical framework about the expectations regarding autonomy, dignity, freedom of choice and privacy in the first place. Given the ongoing debates about privacy and autonomy in the data sharing and digitisation of service provision this is far from clear for many practitioners

Goals

- Does the behavioural policy serve good and legitimate goals?
- The authors encourage policy-makers to “engage in honest cost–benefit analyses considering the effects on the welfare of all actors potentially influenced by the behavioural intervention” (p14)
- Does the behavioural intervention serve goals that are ethically acceptable? Lades & Delaney (2019) note that the “FORGOOD framework *does not provide an answer as to whether a certain nudge is ethically permissible or not*. Rather, it gives guidelines as to where to look for potentially problematic issues” (italics added, 2019, p14)

A key consideration here is that for behavioural interventions that aim to improve people’s lives, do these interventions really make people better off and how is this ‘better off’ defined?

Opinions

- Do people accept the means and the ends of the behavioural policy?
- Not everyone will agree on the ethical boundaries of what is and isn’t acceptable. Where are the acceptable boundaries?
- They suggest using surveys or opinion polls to better understand the public opinion about the intervention itself, its goals and the means used

Options

- Do better policies exist and are they warranted? Are there other policy options?
- Is the behavioural intervention the best policy amongst all of the policy options?
- Does the behavioural policy divert attention and/or political will away from better political decisions?

- Is the behavioural policy more or less cost-effective than other policies?

Delegation

- Do the policy-makers have the right and the ability to nudge using the power delegated to them?
- This section encourages self-reflection. Is there sufficient trust and public acceptability to use a nudge, i.e. do they have authority to nudge? Do they also have sufficient competency?

Another important consideration in implementing this framework is to acknowledge the internal power dynamics that may be faced by policy-makers who were either directed to design a behaviourally informed policy intervention or nudge, or a practitioner within a nudge unit. To what degree would they be able to or have the opportunity/remit to question the policy proposals?

The primary challenges in respect to this framework appear to be the assumptions that these questions can a) be addressed in a straightforward manner and that b) policy makers see these discussions as relevant to their work. The authors worry it may be too simplified, but in respect to the amount of ethical detail, not in the prioritisation of ethical concerns. They had hoped:

“to keep the framework as simple and memorable as possible, but complex enough to capture most of the dimensions in the nudge debate. We chose the terms in FORGOOD in the hope that they are easy for policy-makers to make intuitive sense of. We hope that FORGOOD is easily understandable, attractive for policy-makers to use, will be used by many social groups and comes at the right time.” (p15)

The framework could be further strengthened by addressing the institutional, personal or cultural limitations that policy makers and practitioners may face in using them.

Behavioural Scientist's Ethics Checklist

[Jachimowicz, J., Matz, S., & Polonski, V. \(2017\) The Behavioral Scientist's Ethics Checklist - Behavioral Scientist.](#)

This checklist is very practitioner focused, and is based on a report on BehaviouralScientist.Org. It begins with a story about how one of the authors, Jon Jachimowicz, encountered an ethical challenge while working with a large multinational company, while:

“there was nothing illegal about how the company planned to use the data... [it] would have failed even basic ethical conventions in university-based research. Jon tried to push back against the company's plan. But without a specific set of ethical guidelines in place, it was his professional opinion against that of the vice president.”

The checklist is focused primarily on private sector issues, but ultimately the challenges remain the same. An ethical behavioural practitioner may approach the issue to be addressed, or the intervention to be designed, in a very different way to those who have sought out their services. When the practitioner or expert sits outside of the organisation, whether as a consultant or academic for example, there is far greater scope to push back than if the practitioner or expert is employed by the business or government department.

It is proposed as a living document, open to change in response to feedback from academics and other practitioners. Interestingly, it focuses almost exclusively on the testing of interventions – rather than the ethics of the intervention itself. This highlights how interchangeable some practitioners see behaviourally-informed interventions and trials.

There are – currently – 6 key principles that they recommend practitioners use in negotiating on projects to “ensure that the work we conduct lives up to the highest ethical standards” (np). These principles are:

1) Aligned interests

The interests of the practitioner and the company should be aligned. This is an interesting perspective as it allows for the idea that some practitioners may be perfectly happy to use behaviourally-informed design in a way that is unethical or run in the best interests of the company as opposed to participants. However, later principles do provide clearer ethical guidance.

2) Transparent research process

This section focuses less on the transparency of the behavioural intervention and more on ensuring that the participants are aware of their participation in a study. This is predominantly focused on how data is used. The ramifications are presented as reputation risk and loss of trust. It is interesting that it seems to ignore the transparency of the intervention itself.

3) Rigorous evaluation

This is focused on the 'ethics' of not wasting people's time on poor quality research.

4) Data privacy and security

This is particularly important given the strong focus in this framework on the use of big data and A/B trials for intervention design and learning. Companies using trials and data need to have a clear data privacy and security plan. This is presumably because, in these instances, the interventions and trials are not necessarily cleared through an ethics board.

5) Ease of opt out

Again, this is focused on the ability to opt out of trials and research. If it were run through an ethics board, this would be a non-issue, but given much of this work is being done by private industry there are different expectations.

6) Cost benefit analysis

The checklist recommends considering sample size, stratification and power and selecting participants on this basis. They also note that companies should consider "how the benefits and costs are distributed. For example, if all of the benefits go to the organization running the study and all of the costs go to the population under study (such as employees or customers), then it may be questionable whether the benefits truly outweigh the costs. Companies may therefore want to consider for whom benefits offset costs, and how a positive tally may be achieved for each group." (np).

This document is clearly designed for a specific audience, but was referenced in the FORGOOD document as an example of an alternative framework. The way in which the principles are proposed arguably lack the rigour and detail of the basic ethical standards of a university ethics board, and there remain questions around accountability.

Dilemmas in Applying Behavioural Insights

[Pykett, J., & Johnson, S. \(2015\). *Silver bullets need a careful aim: Dilemmas in applying behavioural insights*, University of Birmingham and Collaborative Change.](#)

Pykett and Johnson (2015) outline some of the principal ethical and political dilemmas raised by the influence of behavioural science on public policy. Rather than acting as a framework as such, this paper is intended to act as a conversation starter and an opportunity for policy-makers and practitioners to “collectively consider what might be the best course of action in shaping behaviour change interventions” (p4). It provides a series of ethical prompts to be considered in the application of behavioural interventions.

Is the intervention open to challenge?

Can people ‘see’ the nudge? Is it reliant on the automatic or reflective systems? Different kinds of behavioural insights carry with them different degrees of consent, openness, transparency and accountability. This reflects much of the work on deliberative nudges (John and Stoker 2019³).

Some ethical considerations include:

- Are citizens aware of the behavioural intervention?
- Should nudge approaches which bypass the conscious agent be permitted?
- Is the citizen able to consent to the intervention?
- Do different types of nudges have different ethical considerations?

³ John, P., Stoker, G. (2019) Rethinking the Role of Experts and Expertise in Behavioural Public Policy. *Policy and Politics*. 47 (2): 209-226. <https://doi.org/10.1332/030557319X15526371698257>

- What are the transparency measures that need to be put in place?
- Have the end goals of the intervention been publically debated?
- Does the intervention allow people to pursue their own end goals?

Pykett and Johnson (2015) recommend making behavioural interventions more transparent and deliberative in response.

What is the legitimacy of the behavioural expert?

Choice architects need to have sufficient training and development to make complex assessments. What is the accountability mechanism for practitioners and policy makers? What is the best way to prevent threats to autonomy?

Ethical considerations:

- Is the moral autonomy of the citizen maintained?
- Can the organisation designing the behavioural intervention be held to account?
- Is the citizen given opportunity to learn about their own biases or from their mistakes?
- Who decides the end goals/public goods to be pursued?
- How can citizens themselves be involved in understanding their own behaviour?

The authors recommend involving citizens/users as experts in their own behaviour when designing an intervention.

Is behaviour the real problem to be tackled?

By focusing on individual decisions and key points in time behavioural interventions can fail to take into account the context/structural barriers. Situations may not be equal for all.

The spaces in which we make decisions are not limited to our immediate environs and choice architectures – this narrow view of space as trigger for individual decision-making is problematic because it denies the cultural, social, political and economic driving forces which shape those very spaces and the social interactions which they cultivate (p10)

Ethical considerations:

- Does the emphasis on Behavioural Economics and nudge blinker policy professionals and behaviour change practitioners to other approaches and paradigms?
- Does the political appetite for Behavioural Economics create the risk of issues less conducive to such tactics being overlooked or de-prioritised?
- What forms of knowledge and understanding are obscured in the enthusiasm for a behavioural science approach?
- Can Behavioural Economics tactics create sustainable behavioural change as well as influence choice or trigger one-off behaviours?
- Does the concept of the 'social' forwarded in Social Psychology shed any light on structural drivers of behaviour?

Unintended consequences?

This section highlights the risk of obscuring context and of inviting an overly paternalistic view of role of the state to help people make 'better' decisions. Can the proposed intervention create social divisions?

Ethical considerations:

- Behavioural Economics focuses on universal cognitive dynamics. How can we ensure it responds to the diversity of our communities and their needs?
- Can 'nudge' approaches have negative unintended consequences—such as crowding out or diminishing intrinsic motivation, or producing emotionally vulnerable subjects?
- How might future technological developments (e.g. big data; wearable biosensors, smart cities) affect the application of behavioural insights and raise new ethical dilemmas?
- Can 'nudge' create long-term, sustainable change across a range of contexts?
- How can we avoid behavioural interventions that might be divisive or stigmatising?

The main recommendation is to ensure the interventions produce results that are empowering, respectful and enable the reflective system.

Is what works the best thing to do?

It is important to acknowledge that science is not value free. Discussions about morals, ethics and politics are critical. The report notes that:

“the drive for certainty and statistical truth can detract from the fact that social interventions necessarily always take place within dynamic systems, with people responding in unpredictable ways. Broader understandings of how societies change should not therefore be dismissed in the current enthusiasm for behavioural science”. (p15)

Ethical considerations:

- What kinds of evidence are being used to justify the behavioural intervention?
- Does this evidence tell the whole story?
- Can proof of ‘what works’ really be transferred to other contexts?
- What is the basis for ‘success’ for the intervention, and who has been involved in determining this?
- Can the effects of the intervention be isolated, and cause and effect identified?
- Is scientific evidence a sufficient justification for policy?

The recommendation is to introduce diverse types of evidence and insight. Policy-makers and practitioners need to embrace alternative perspectives on success and how ‘what works’ is defined.

Ethics of Nudges

[Clavien, C. \(2018\). Ethics of nudges: A general framework with a focus on shared preference justifications. *Journal of Moral Education* 47 \(3\),366-382](#)

Clavien (2018) makes the argument that, given that behavioural factors and choice architecture have been demonstrated to have an impact on decision making, it would be “problematic—at least not ethically neutral—to ignore or reject the possibility to use nudges” (p367). However, Clavien also acknowledges that their use may lead to unwanted side effects. To manage these ‘side effects’ she suggests an evaluation procedure be undertaken. Four main questions should be addressed:

Goals

Are the goals justifiable? Are they selfish, social or nudge-driven?

Effectiveness

Is it an/the most effective way to achieve the aforementioned goal? What are the alternatives?

Trustworthiness

Are there conflicts of interest? Are they qualified? What behavioural biases and heuristics are influencing the nudger?

Ethics

What are the side effects? Can this be measured? If so, how?

On the question of how autonomy will be respected, Clavien presents the argument that this is an unavoidable side effect of nudge. Some people will also resist being nudged. She suggests there are four broad arguments presented for doing it anyway.

1. nudger's goals are important
2. everyone is being nudged already
3. nudgers share the goals of nudges
4. nudgees agree or would agree if asked.

Clavien focuses on # 3 and #4. However further empirical justification and more contextualised claims about appropriateness – with reference to ethical or philosophical foundations – would help support the principle of shared goals. Clavien refers to #3 as *Shared Preference Justification (SPJ)* and #4 as *Content Justification (CJ)*. SPJs and CJs are similar, but they are grounded on different empirical data. She argues that CJs are much harder to achieve, as “empirical evidences for CJs involve nudgees’ informed consent which is a demanding condition” (p372). This is debatable, but leads to the remainder of the paper focusing on how to best justify the use of SPJs.

SPJ's require that an investigation is undertaken into nudgee's preferences and preferences can be hard to determine:

“[E]thically relevant preferences are difficult to define and identify. Individual ‘preferences’ may refer to different phenomena such as action-motivating states – e.g. drives, wants,

actual choices – or more cognitive states – e.g. simple ‘likes’, or more complex comparative evaluations. Moreover, preferences may be about different types of states, for example, long-term versus short-term states, global versus specific states. Deciding which of these phenomena are relevant for evaluating nudges is a matter of debate (Rebonato, 2014)” (p373).

The logic that follows is complex but essentially, for each proposed nudge the following details should be provided: “(1) the degree to which nudgers’ goals align with the preferences of individuals impacted by the nudge; (2) the proportion of individuals whose preferences align with nudgers’ goals; and (3) the empirical validity of the data” (p378). The complexity is highlighted by the suggestion that policymakers develop a software to use in the assessment process (p380). Clearly, when moving away from revealed preferences – the standard neoclassical approach – significant complications arise. Why a policy maker would feel motivated to do this is not explored.

Constitution for Behavioral Policy-making

[Fabbri, M., & Faure, M. \(2018\). Toward a “constitution” for behavioral policy-making. *International Review of Economics*, 65\(3\), 241–270.](#)

This framework is built upon one overarching issue that behavioural policy represents: “If agents are subject to behavioral biases leading them to make “wrong” choices, the policy-maker can no longer rely on the revealed preferences approach (e.g., what people choose is what people prefer) for defining a welfare criterion” (p241). This then leads to the absence of a defined normative criteria for justifying intervention as an alternative for the revealed preference approach, according to the authors, has not yet been developed in the literature in a satisfactory manner.

They recommend formulating a constitution, a set of guiding principles, which are collectively agreed and which policy makers would be accountable to. An ex-ante set of rules that specify the necessary conditions for behavioural policies. This is composed on four key components:

1) Independent decision making

An independent group of experts in the form of an agency should provide first a risk assessment and second a broad assessment which explores noneconomic considerations such as efficacy and acceptability.

2) Meaningful public participation

In order to avoid reliance solely on experts, the public should be included. There are some concerns that this may increase transparency and risk the effectiveness of the nudge.

3) Defining intervention domains

This component refers to the parameters assigned to the risk and broader assessment. These include:

- are there plausible traditional justifications (externalities, information failure, inadequate competition) for the measure, operating independently of paternalism?
- if not, and taking account of the insights of social psychology, is the regulated activity one with regard to which a significant proportion of the agents make decisions that are unlikely to reflect their real preferences?
- if so, are the likely costs of the regulatory measure proportionate to the likely benefits and/or could the same be reached at lower cost by an alternative instrument? (p264)

4) Sunset clauses

In order to be renewed, a behavioural policy must be explicitly re-discussed and re-approved (p264). However, if the policy is having a positive welfare effect, this does risk allowing it to lapse or be 'gutted'.

Nudging in Public Health

[Department of Health, Irish Government - National Advisory Committee on Bioethics. \(2015\). Nudging in Public Health – An Ethical Framework.](#)

This document is valuable firstly because of its focus on public health and secondly because it was drafted specifically by a Bioethics committee. The authors acknowledge that nudge has pros and cons, but that “the novelty of nudging requires policymakers to consider the main ethical concerns that may arise” (p2). This document presents specific ethical considerations for the proposed interventions proposed as part of the Institute for Government/UK Cabinet Office MINDSPACE framework: Messenger, Incentives, Norms, Default, Salience, Priming, Affect, Commitments, Ego (Dolan et al 2010)⁴. These are all captured under the following ethical principles:

Autonomy

The principle of autonomy refers to one’s right to make independent choices on the basis of one’s own beliefs and values without undue external influences.

Proportionality

The principle of proportionality requires that those considering limiting personal rights for the purpose of public health policy must balance the level of the interference with the intensity of the social need for such interference. It must be determined that the positive effect will outweigh the negative effect and that the negative effect is only permissible if there is a proportionate reason for permitting the foreseen negative effect.

Justice and equity

Justice recognises that different groups in society may be advantaged/disadvantaged, e.g. on the basis of socio-economic status, gender, ethnicity, and this leads to diminished health and wellbeing. Justice and Equity require the fair distribution of the likely benefits and burdens of public health policies amongst the population.

Non-discrimination

No individual or group should be discriminated against, disadvantaged or stigmatised on any grounds, in violation of human dignity, human rights and fundamental freedoms.

Protection of Vulnerable Populations

⁴ Dolan, P., Hallsworth, M., Halpern, D., King, D., Vlaev, I. (2010) MINDSPACE: Influencing Behaviour through Public Policy. Institute for Government/Cabinet Office, London.

Human vulnerability should be taken into account when considering the introduction of nudging policies. Individuals and groups of special vulnerability should be protected and the personal integrity of such individuals respected.

Respect for Cultural Diversity and Pluralism

The importance of cultural diversity and pluralism should be given due regard.

Solidarity

Solidarity requires that populations share the benefits, risks and burdens of public health policies. At a basic level, solidarity reflects a collective commitment to carry e.g. financial, social, emotional or other “costs” in order to assist others.

Trust and Accountability

Public trust is essential in ensuring the success of public health initiatives. In order to maintain public trust, public health initiatives involving nudging need to be evidence-based and have specific and reasonable objectives.

To meet these ethical obligations the report suggests that policy makers also ask themselves the following set of questions:

1. What are the public health goals of the proposed intervention?

These goals generally ought to be framed in terms of public health improvement, e.g. the reduction of morbidity or mortality, the reduction of smoking or obesity rates or the improvement of a particular service. For example, an HIV screening programme should have fewer incidents of HIV as its ultimate goal, not merely that a certain number of individuals will decide to be tested (p20).

2. How effective will the proposed intervention be in achieving its stated goals?

As a rule of thumb, the greater the burdens posed by a specific intervention—for example, in terms of cost, constraints on liberty, or targeting particular, already vulnerable or disadvantaged sections of the population — the stronger the evidence must be to demonstrate that the intervention will achieve its goals. If the evidence does not support the assumptions, then proceeding with the intervention in question would be ethically problematic (p21).

3. What are the known or potential burdens of the proposed intervention? Can burdens be minimised? Are there alternative approaches?

Policymakers should strive to minimise any identified burdens to individuals or specific communities. If there is a prospect that an intervention may carry potential or actual burdens, policymakers are ethically required to determine whether the intervention could be modified in ways that minimise the burdens while not greatly reducing the intervention's efficacy. They should also consider whether an alternative, less burdensome, intervention might achieve similar outcomes (p24).

4. Is the programme implemented fairly?

If public health nudging policies aspire to reduce health inequalities, then they should include measures specifically aimed at improving the quality of life of particular individuals and/or groups who are disadvantaged and/or at risk. Nonetheless, it is imperative that such measures do not result in the stigmatisation of people who are already marginalised or disadvantaged (p24). Policy-makers should consider what characteristics define a target group. They should investigate whether there is evidence to suggest that those characteristics correlate with the problem that is sought to be addressed (p25).

5. How can the benefits and burdens of a program be fairly balanced?

This section is focused on the importance of minimising burdens, rather than how.

6. Which should take precedence: individual freedoms or the common good?

In public health, this argument may be different than in other policy areas. Harms to health caused by unhealthy lifestyle choices, not only affect the individual concerned. The healthcare costs resulting from unhealthy behaviour can impact society more broadly.

7. What role can the concepts of trust and accountability play in the field of nudging, which works better "in the dark"?

It is yet to be seen if the "in the dark" principle is in fact necessary for BPPs.

Overall this report recommends that "communicating policy decisions and the rationale behind them in an open and transparent way has been identified as one of the crucial factors in increasing the acceptance and co-operation of those who will be affected by these decisions" (p29).