Ethics and Behavioral Policy Interventions in the Labor Market

SUMMARY

This paper offers a guide for policymakers developing behavioral policy interventions in the labor market to help them apply ethical standards. The core ethical concerns about behavioral policy can be met by adherence to **two guiding principles**: policy should promote the **welfare** of citizens and policy should treat citizens with **respect**.

Implementing these two principles requires taking two practical steps:

(1) The assessment of evidence to ensure policy is **welfare improving**. Research assessing the needs of individuals directly affected by the policy should be combined with a testing-based approach that evaluates the broader welfare impacts prior to scale-up.

(2) The **provision of transparency** to eliminate the possibility for deception or manipulation and ensure people are treated with respect. Policy should be discussed publicly and justified based on its impacts on the welfare of those affected, allowing people to consider whether their best interests are being served.

Behavioral policy interventions can be implemented at different points in the labor value chain, from educational preparation and job search to retirement. A brief review indicates that interventions affecting career choices and hiring decisions are likely to pose the largest ethical challenges, requiring careful assessment of evidence and clear transparency.

The paper provides a checklist for putting ethical standards into practice. It includes steps that establish the capacity for ethical reviews, assessment of evidence, and provision of transparency, procedures for the design, testing, and implementation stages of behavioral policy, and general ethical safeguards that assure legality, privacy, and legitimacy.

1. INTRODUCTION: BEHAVIORAL POLICY AND ETHICAL CONCERNS

Behavioral policy interventions – sometimes referred to as policy "nudges" – have become more common over the past decade as governments have applied insights developed in the fields of behavioral economics and psychology to policy design. Behavioral economics has highlighted the ways in which people are not like the perfectly rational actors featured in traditional economic models but instead have limited attention and cognitive abilities, are prone to a range of biases that affect their judgements and decisions, struggle at times with self-control, and are powerfully affected by social norms and context. A key finding is that people often fail to align their actions with their intentions. Given these human psychological tendencies, relatively small changes in the choice environment or "architecture" can have large impacts on people's behavior and the outcomes they experience, even when the financial incentives associated with different actions remain unchanged.

A policy "nudge" or behavioral policy intervention can be defined as a deliberate alteration to a feature of the choice architecture created by policy – exactly how and when choices are presented to the citizens affected and how policy is administered. A nudge does not mandate or prohibit certain choices, change the financial incentives associated with those choices, or alter the information about the choices that is available. In contrast to traditional types of policy levers, behavioral interventions take place in the design details, which can make some types of choices by citizens or administrators more likely than others. Nudges include such things as simplifications of government forms and messages, checklists, reminder messages and warnings to avoid mistakes, reordering and filtering of choice menus, changing default settings, and sharing information about choices made by others.

Many governments around the world have established units or teams to apply behavioral insights to help improve the design of their policies, programs, and administrative procedures. Applications of behavioral insights to labor market policies have included new features of programs supporting unemployed individuals seeking work, new ways of assisting people to acquire new skills through vocational training, changes to programs assisting small businesses with compliance with laws regarding wages and working conditions, and the introduction of new processes aimed at reducing the impacts of discrimination in hiring.

The major concern raised about the ethics of behavioral interventions is that "nudging" can appear to involve manipulating citizens, influencing their choices without their knowledge, perhaps shaping their behavior in ways that do not align with their own best interests. From this perspective, nudging could be opposed as a tool that enables a dangerous type of paternalism in which governments exercise increasing control over the behavior of unknowing citizens. It could also be objected to on the basis that it does not treat people with respect and preserve individual autonomy.

These are important concerns, but they can be addressed in direct fashion by establishing sound principles for ethical behavioral policy and using these to create a practical guide for governments to apply behavioral insights in an ethical way.

2. TWO GUIDING PRINCIPLES FOR ETHICAL BEHAVIORAL POLICY

As a starting point we should recognize that "nudging" is unavoidable. Every policy or program a government adopts has a choice architecture. This architecture includes all the words used to describe its purpose, processes, and components in all communications about the program, along with the specific procedures designed for applications, participation, reporting, and administration, and the features of the social contexts in which all the relevant individual choices are made. Each element in this architecture – every aspect of the forms, choice menus, messages, social settings – affects the behavior of those involved, including officials administering the program as well as citizens engaging with the program, and thus affects outcomes. There is no neutral (nudge-free) design for the choice architecture that allows for perfect rational autonomy for all individuals involved.

This is a critical realization because it means that the choice for policymakers is not *whether* to nudge, the issue is how to be sure that we are being responsible, careful, and ethical when we design the choice architecture and the various kinds of nudges unavoidably incorporated within it. To put this another way, the issue is not whether nudging is ethical, it is how to nudge ethically.

What ethical criteria should we use when designing behavioral policy? There are two guiding principles: policy should promote the welfare of citizens and policy should treat citizens with respect. These are guiding principles for policy of *any* type, but we should carefully examine how they can be applied to behavioral interventions specifically.[1]

a. Welfare

Ethical behavioral policy is welfare improving. This principle requires that the individuals whose choices are directly influenced by the policy, and other citizens who might be affected indirectly, are made better off, in aggregate. Analysis of the welfare impacts can be quite straightforward when nudges help people align actions with intentions in simple ways, but in some cases a more challenging assessment of aggregate social welfare is required to be sure a policy meets the ethical standard.

Direct effects
When those whose actions are directly influenced by the policy are the only individuals affected, the assessment of welfare impacts can be reduced to a simple question: are these people made better off? For this assessment we should rely upon what people themselves identify as a better outcome for them. An ethical nudge will help people align their actions with what they themselves see as the best choice for themselves when they have a chance to think about their options, focus their attention, and consider all the relevant information in a reflective, deliberate way.

---

[1] These two principles represent a combination of, or a compromise between, two distinct approaches to ethics: a utilitarian approach that focuses on consequences and a rights-based approach. Note that Sunstein (2016) discusses four ethical criteria for policy: welfare, autonomy, dignity, and self-government. The discussions of autonomy and dignity center on a concern for respect. The discussion of self-government focusses on democratic principles for policymaking.

+For many types of behavioral policy, this will be easy to assess. Most people in a training program for job seekers will quickly agree that they would rather not forget an important appointment or deadline, for example, and that a reminder that helps them remember in time will make them better off. In other cases, it may be more difficult to identify people's true preferences and be sure that a nudge is helping them align intentions with actions. Do people want to acquire skills that will prepare them for a very specific set of jobs? Do people want to transition to a completely new career path? These are big decisions that may be affected by behavioral interventions that will have large welfare impacts (we will discuss these examples more below).

People affected by a behavioral policy may differ in their preferences. They may also differ in their need for help with aligning actions with intentions. Some eligible individuals may prefer not to apply for a government job-seeker assistance program, for example, or they may have no trouble remembering the deadline for the application or filling in the required forms, even if many of their counterparts would like to participate in the program and do sometimes lose track of deadlines and find it difficult to complete complex forms. In assessing the welfare effects of a nudge, such as reminder message about an application deadline or a simplified application form, we do need to be sure it would not impose any undue burden on those who prefer a different choice or do not need any assistance. Since nudges are by definition "choice preserving," unlike more onerous types of government policies such as mandates, the costs for those who prefer a different choice are likely negligible.[2]

It is possible to think of some instances in which one group of those affected directly by a behavioral policy feel that their welfare would be significantly *reduced*. Workers who have physical disabilities, for example, may feel that workplace message campaigns that encourage people to stretch and exercise to prevent injuries on the job make them feel unhappy or marginalized. In such cases, some assessment of the impact on the aggregate social welfare for the full set of those affected may be required. While no one has a welfare meter that would enable us to compare welfare gains made by some individuals with welfare losses incurred by others, a reasonable proxy is cost-benefit analysis that compares estimates of total monetized benefits with monetized costs.[3]

Indirect effects
What if the behavioral policy affects not just those whose actions are directly influenced by the nudge, but other members of society too? In fact, these indirect effects may be the motivating reason for adopting the policy. This applies for nudges designed to address market failures by discouraging individual behaviors that impose harm on others. Some employers may prefer to avoid compliance with certain workplace health and safety regulations, for example, but any welfare loss they experience if nudged into greater

---

[2] This attribute of nudging is emphasized by Thaler and Sunstein who have coined the term "libertarian paternalism" to describe the policy approach: see Thaler and Sunstein (2003).

[3] For a detailed discussion of the application of cost-benefit analysis by government to inform policy decisions, with detailed examples, see Sunstein (2018). For a comparison of social welfare and cost-benefit analysis, see Adler (2012). Adler points out that if income and other determinants of individual utility are symmetrically distributed among winners and losers from a given policy choice, cost-benefit analysis is a reasonable proxy for social welfare analysis based on equal weighting of individual utilities

compliance is likely to be overshadowed by the general benefits for workers and other members of society from reductions in injuries and chronic health issues that may also raise healthcare costs. In cases such as this, the fact that traditional policies are already applied by most governments to outlaw and fine businesses with unsafe work conditions helps establish the social welfare case for discouraging this behavior more via nudging if possible.

Another possibility that may be more problematic for welfare assessment involves the potential for market-level effects of behavioral policies that broaden impacts beyond those whose actions are directly influenced. A hypothetical example might be a policy that improves take up of a government program among one group of eligible participants so much that it crowds out other eligible participants. Or we might imagine policy that encourages so much participation in a skills training program that it increases the supply of workers with those specific skills so much that wages fall. When such impacts are possible, due diligence requires some attempt to establish whether the aggregate impact of the policy is an improvement in social welfare, using cost-benefit analysis and, where appropriate, need-based weighting of effects felt for different groups.

This final point is an important one. In defining and assessing the implications of policy for social welfare, when different individuals or groups are affected differently, we have a choice to make. We can weight all individuals equally or we can apply some type of weighting that prioritizes the impacts for some groups over others, perhaps in line with broader societal preferences for fairness or justice that imply giving more attention to the welfare of those in lower-income or historically disadvantaged groups.[4]

Implementing the welfare principle in practice: the case for rigorous and carefully designed evaluations

The best safeguard for ensuring that a proposed behavioral policy is welfare improving is the careful assessment of evidence. We should begin by conducting careful diagnostic research to assess the preferences and needs of those individuals who would be directly affected by the policy. Evidence gathered from interviews and surveys can be critical in each case to understanding what people want for themselves and whether (and why) they may not be able to align their actions with their intentions in the moment.[5]

We can go further by carefully evaluating the impact of the proposed nudge in small-scale pilot studies and randomized controlled trials conducted prior to scale-up. Rigorous, carefully designed evaluations can allow us to measure average effects, variable effects for different groups, indirect impacts for others whose actions are not directly influenced by the policy, and unintended impacts on behavior.[6] The evidence gathered in these evaluations can include quantitative assessments of the impacts on choices and outcomes as well as self-assessments from the individuals directly affected.

---

[4] One may think of this as including another ethical principle – fairness or justice – in the assessment of policy, but because it will depend on societal-level agreement about the appropriateness of prioritizing the needs of some groups in society over others, it can be incorporated into the broader concept of social welfare.

[5] Behavioral economics allows that choices may not accurately reveal preferences. A person's stated preferences may differ from their actions due to a variety of issues, such as limited attention and imperfect self-control.

[6] For an excellent example in which a randomized controlled trial is used to generate an accurate assessment of the aggregate social welfare impacts of a behavioral policy, see Allcott and Reed (2017).

Instituting a testing-based approach can help to ensure that policy satisfies the ethical requirement that it improves welfare. Put a slightly different way, outside some unusual circumstances, it may be difficult to make the case that untested policy is ethical.

b. Respect

Ethical behavioral policy treats citizens with respect. This principle requires that the people whose choices are influenced by behavioral policy should not be deceived or manipulated. Manipulation occurs when people are influenced by someone in ways that are undisclosed and unknown to them, even when there is no outright misrepresentation. Both deception and manipulation expose people to the risk that their choices will be shaped to fit the interests of the influencer rather than their own best interests.[7] Both create a justified sense of betrayal at being treated as a mere instrument in the hands of the more-informed influencer.[8]

Safeguarding autonomy
Treating people with respect is associated with safeguarding their autonomy, their ability to control their own fate. Concerns about the erosion of respect and autonomy lie at the heart of some fears that behavioral policy represents a dangerous form of paternalism.[9]

It is important to note that attaining complete autonomy in all individual decision-making would be an impossible standard to set. One of the core insights from behavioral economics is that humans have imperfect control of their choices, for reasons related to very human limitations in attention and cognitive abilities, implicit biases that affect judgements, and lack of self-control. People cannot be fully engaged, informed, reflective, and deliberative when making all the choices they make each day. Indeed, our minds seem to be adapted to enable us to make many choices using quick shortcuts and unconscious habits to expend as little cognitive effort as possible.[10] A nudge will not displace anyone from a state of full autonomy. A well-designed nudge is more likely to increase autonomy by helping people focus their limited attention on the things most important for making good choices and align their actions with their best intentions.

Promoting personal responsibility
A related concern about behavioral policy that is sometimes raised in connection with autonomy is the fear that nudges promote passivity and avoidance of personal responsibility.[11] Here the complaint is that nudges take our cognitive and other limitations

---

[7] Addressing respect in these terms conforms well with the treatment of respect in the pathbreaking Belmont Report, that has set the now internationally recognized ethical standards for human subjects' research (see National Commission, 1979). The report makes clear that treating individuals with respect requires that they not be deceived and not be exposed to risk without informed consent.

[8] See Sunstein (2016, 83). Note that Sunstein defines manipulation in somewhat broader and more ambiguous terms as something that influences people and "does not sufficiently engage or appeal to their capacity for reflection and deliberation."

[9] See Glaeser (2006).

[10] See Kahneman (2011), pp.19-97. Note too that people are in general happy to delegate many decisions they could make themselves to others we trust (or hope) will act in our best interests, including government officials and policymakers, who can be held to account for their performance in some way.

[11] See Waldron (2014).

as given rather than aiming to improve our abilities as choosers via education or training of some type that might improve our attentiveness, for example, or our self-control. But it is unimaginable that people could become experts in all the fields in which they are required to make fully informed, optimal choices (e.g., health and medicine, finance, environmental science, engineering, and law) and be trained also to suppress human tendencies to rely on short-cuts, develop habits, and give in to momentary temptations. Again, a well-designed nudge is more likely to improve choosing ability by helping people focus their limited attention at the right time on the things that are critical for making choices aligned with their best intentions.

It is clear immediately that many types of behavioral policy can be classified as consistent with the respect principle in straightforward fashion, as they are neither deceptive nor manipulative. Nudges that consist of simplifications made to government forms, reminder messages, safety warnings, informational labels, checklists, re-ordered menus of options, altered default settings, and the provision of information about choices made by others are all clearly out in plain sight. Indeed, these nudges would not be effective if they were hidden and unnoticed.

Other types of behavioral policy may be less obvious to the people being influenced, at least in terms of conscious recognition. The use of particular words (or colors, images, even sounds) in the choice architecture created by policymakers may not be registered consciously by affected individuals, for example, even when designed deliberately to influence their choices.

Implementing the respect principle in practice: The case of full disclosure and transparency

The critical safeguard to ensure ethical behavioral policy is the **provision of transparency**. We should fully disclose what we are doing and why when we introduce a new behavioral intervention. Public discussions of the policy when it is introduced, and separate disclosure statements made at the point at which nudges are delivered, should make it clear how and why an important feature of the choice architecture has been altered or set in a specific way.

For example, when sending reminder messages, we can explain that, because we know that many people find it difficult to remember important deadlines at times, we think that such messages can be helpful. When setting default options in enrolment forms, to take another common example of nudging, we can note that many people find it difficult to make a choice in the limited time they have available for this task, and we have selected an option that is typically the best choice for most people.

Transparency ensures that we treat people with respect, by eliminating the possibility for deception or manipulation, while also providing an additional ethical safeguard against potentially harmful or misguided applications of nudging that might reduce welfare. [12] Publicly explaining and justifying policy based on its impacts on the welfare of those
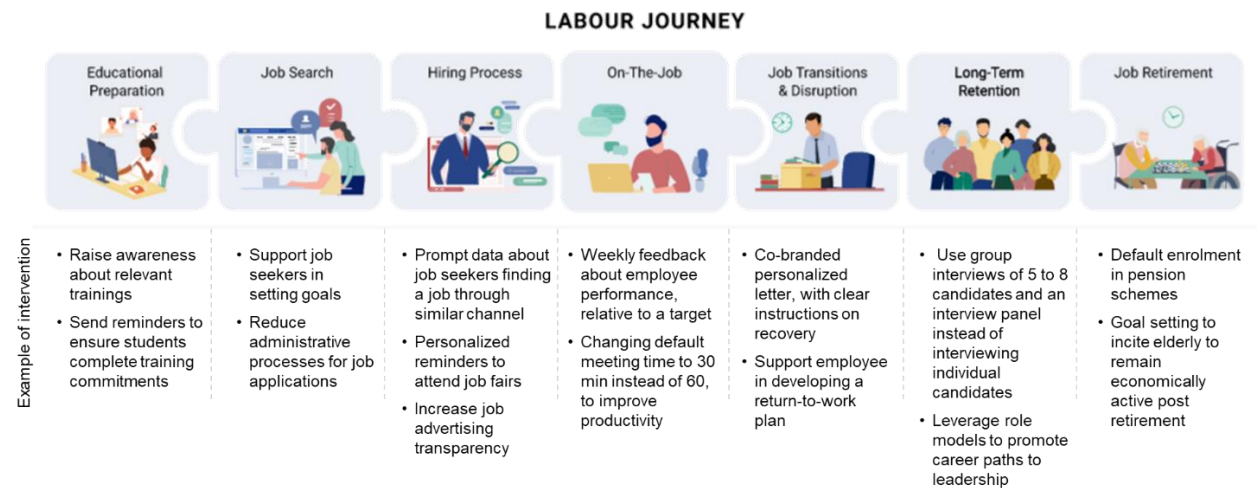
---

[12] Sunstein (2016, 42).

affected allows people to scrutinize nudges to be sure that they are serving their best interests. Transparency like this thus serves to limit policy to forms that are convincingly welfare-improving, buttressing adherence to the first ethical principle as well as the second principle. Assessing the evidence and testing to be sure that a policy is welfare improving means we can be confident that transparency will not imply any reduction in policy effectiveness; being made aware of a nudge will not undermine its impact if people understand how it is designed to help them. [13]

3. APPLICATIONS TO BEHAVIORAL LABOR MARKET INTERVENTIONS

Behavioral policy interventions can be implemented at different points in the labor value chain, from educational preparation and job search, to hiring processes, on-the-job issues and workplace standards, job transitions and employment disruption, long-term retention issues and lifelong learning programs, and finally retirement (See exhibit). A brief review indicates that interventions affecting career choices and hiring decisions are likely to pose the largest ethical challenges, requiring careful assessment of evidence and clear transparency.

Exhibit: Behavioral applications across the labor journey

LABOUR JOURNEY

Educational Preparation	Job Search	Hiring Process	On-The-Job	Job Transitions & Disruption	Long-Term Retention	Job Retirement
• Raise awareness about relevant trainings • Send reminders to ensure students complete training commitments	• Support job seekers in setting goals • Reduce administrative processes for job applications	• Prompt data about job seekers finding a job through similar channel • Personalized reminders to attend job fairs • Increase job advertising transparency	• Weekly feedback about employee performance, relative to a target • Changing default meeting time to 30 min instead of 60, to improve productivity	• Co-branded personalized letter, with clear instructions on recovery • Support employee in developing a return-to-work plan	• Use group interviews of 5 to 8 candidates and an interview panel instead of interviewing individual candidates • Leverage role models to promote career paths to leadership	• Default enrolment in pension schemes • Goal setting to incite elderly to remain economically active post retirement

a. Getting started. Supporting students and job seekers

Welfare: clear positive impact
Respect: guaranteed

To date, perhaps the most common types of behavioral labor market interventions have been programs designed to assist students and employees to undertake education and vocational training courses that provide in-demand skills, and programs designed to help unemployed individuals to find jobs. Examples of nudges in these areas have included the use of plan-making prompts, commitment devices, and goal-setting tools aimed at helping individuals overcome present bias and the problems of self-control. Individuals may intend

---

[13] Note also that the evidence indicates that awareness of a nudge does not appear to reduce its effectiveness in terms of influencing choices. See Thaler and Sunstein (2021, 324), Bruns et al (2018), and Loewenstein et al (2015).

to diligently study and complete course assignments until they graduate from a program, or to diligently search and apply for positions until they find a good job, but they often have trouble staying on track. Behaviorally designed support programs appear to help.[14]

For these types of behavioral policies, applying the two guiding ethical principles is straightforward and unproblematic. It is simple to assess whether they have positive impacts on the welfare of individuals directly affected (they do) and these policies carry negligible conceivable risk of welfare losses for others. These interventions are also provided in transparent ways without deception or manipulation, treating all individuals with respect.

b. Fair and square: Reducing implicit bias

Welfare: testing required to confirm
Respect: guaranteed

Different types of behavioral interventions have been developed to help reduce discrimination and promote diversity and inclusion in hiring practices in the public and private sectors. Implicit bias is an unconscious process that may affect judgements and evaluations of candidates by employers even when they are committed to eliminating discrimination. De-identifying job applications, by removing information about gender, race, and age, is a way to prevent implicit bias from influencing decisions by employers, at least in the early, pre-interview stage of the hiring process.[15]

The ethical review here is perhaps more challenging. The intervention is not based on deception or manipulation and so meets the standard for respect. For employers who are sincere in wanting to reduce implicit bias in hiring, it is welfare improving. For employers who want to practice discrimination, including those who would like to practice a form of positive discrimination (or "affirmative action") that favors traditionally underrepresented groups, the intervention could reduce welfare. Aggregate social welfare might be increased significantly by reducing discrimination and improving efficiency, with better matching of skills with positions. But it is possible that diversity among employees may decrease if positive discrimination is reduced overall and this may reduce aggregate social welfare if diversity improves team productivity or is valued for other reasons. Welfare impacts may vary considerably across different groups of job applicants. Reducing discrimination will harm those benefiting from explicit or implicit bias and benefit those being discriminated against; this latter group could include traditionally underrepresented groups in some scenarios in which affirmative action is being practiced.[16] In a complicated case such as this, evidence is critical and only rigorous testing of the impacts of the intervention prior to full-scale implementation will allow assessment of the welfare effects.

c. Keeping the law: improving compliance

---

[14] See Briscese and Tan (2018).
[15] See BETA (2017b)
[16] Note that this scenario was the case in most of the agencies that participated in the BETA trial that tested the impact of de-identification in the Australian Public Service. A similar finding in a private sector setting in France was reported in Behaghel, Crépon, and Le Barbanchon (2015).

Welfare: positive at the aggregate level
Respect: guaranteed

A little further along the value chain, dealing with on-the-job issues, behavioral interventions have been applied to address compliance with labor laws. Here the focus is on assisting managers of small and medium enterprises to ensure that the wages and conditions of work (including hours of work and workplace health and safety standards) provided for their employees meet the requirements set by regulations. Limited attention and cognitive capacity among managers, often overwhelmed by the daily cognitive demands of running small businesses, may mean that they fail to notice noncompliance, especially when regulatory standards are altered or updated. This can be a particular issue when employees are foreign citizens for whom conditions of work are set by visa status. A behaviorally designed program that provides managers with reminders, checklists, planning prompts and a simplified online audit process appears to be effective in boosting compliance.[17]

Applying the ethical principles in this case also seems quite straightforward, although with some minor complications. This type of intervention is fully transparent and involves no deception or manipulation, so it satisfies the ethical requirement that people are treated with respect. On the issue of welfare, managers who intend to comply fully with labor laws but sometimes fail to do so due to inattention and cognitive overload, presumably welcome the intervention and would see it as serving their interests well. There may be other managers, however, who prefer not to comply fully with labor laws if they can do so without being detected (and fined) in order to lower costs and maximize profits. It is possible that their welfare would be lowered by an intervention that reduced noncompliance overall and made detection easier for authorities. But the existing laws mandating standards provide a clear basis for regarding such losses as second-order compared to the gains in aggregate social welfare (and particularly, the welfare of employees) that follow from compliance with such laws; policymakers would not have made these standards mandatory otherwise.

d. Career guidance: Influencing life-defining choices

Welfare: ambiguous, require rigorous testing
Respect: guaranteed

Finally, we might contemplate other potential behavioral labor market policies that, in theory at least, could pose much greater challenges for ethical assessment. Governments might nudge individuals towards choosing particular types of jobs, for example, or nudge people towards acquiring particular types of skills. Such interventions could be justified as efforts to help people with complicated judgements about the probabilities of desirable outcomes (high earnings, say) associated with choices between job and skill types. Most people have difficulty making reliable predictions about uncertain outcomes that will occur in the distant future. Accurate predictions about future labor market outcomes require analysis of a great deal of information about economic, technological, and regulatory

---

[17] See BETA (2017a).

changes that is beyond the capacities of most people. Instead, people are likely to substitute best guesses shaped by availability bias (e.g., recent news about job prospects in a specific sector) and optimism bias (e.g., wishful thinking about earnings for a job preferred for other reasons). Governments might design an intervention to provide more accurate estimates of likely future earnings for specific jobs and skills, which may guide people towards growth sectors and associated skills that are likely to be in high demand in the future. Would this be ethical?

Interventions of this type would be fully transparent and thus would satisfy the respect principle, but because they may influence career choices that have large and long-lasting impacts, they would pose some large challenges for welfare analysis. Career choices are self-defining and transformative in ways that most other choices are not. They can place a person in a new social category that affects how they are regarded by others, how they see themselves, and over time can also shape values and preferences. When contemplating such transformative choices, it may be extremely difficult for a person to anticipate just how much he or she will enjoy things as the "new" self.[18] If individuals themselves are not reliable assessors of which career choice is best for them, we have an interesting puzzle. An intervention that provides reliable predictions of likely earnings for different types of jobs seems helpful. But if, by making earnings especially salient it thereby steers people toward some sectors or occupations, without being able to say with confidence that those careers are what these individuals themselves know will make them better off, it is difficult to be sure that the policy is welfare improving for those directly influenced.

## 4. PUTTING THE PRINCIPLES INTO PRACTICE: A CHECKLIST

Applying the guiding principles allows us to evaluate behavioral policy on ethical grounds and rule out the potential for harmful nudging and misuse. Implementing the principles in practice requires that we establish the capacity for ethical reviews, assessment of evidence, and provision of transparency, along with clear procedures for the design, testing, and implementation stages of behavioral policy, and general ethical safeguards that assure legality, privacy, and legitimacy.

---

[18] See Paul (2014) and Ullmann-Margalit (2017). This issue has been explored in detail in regard to the choice to become a parent.

A preliminary ethical assessment of the designed behavioral intervention		

Are clear guidelines followed when studying the behavior, including:
   ○ Collecting and documenting consent
   ○ Revealing the purpose of the study
   ○ Securing legal permission if required
   ○ Ensuring participants are voluntarily participating
   ○ Safeguarding vulnerable population?
	Yes / no	If no, re-consider the behavioral assessment

Does it identify the direct target group?	Yes / no	If no, re-define the intervention

Does the intervention's design respect the target group; i.e., ensure full disclosure and transparency of the intervention? In other words, does it:
   ○ Clearly communicate the purpose and nature of the intervention?
   ○ Safeguard individual's autonomy?
   ○ Promote personal responsibility?
	Yes / no	If no, engage with stakeholder groups to redesign the intervention to safeguard individual's autonomy and personal responsibility

Does the intervention have a clear positive welfare on?
   ○ Individuals
   ○ Target groups
   ○ Society?
	Yes / no	If no, conduct a more rigorous evaluation to assess individual, group, societal, and indirect impacts

Are other variables and / or other indirect effects on welfare taken into consideration?	Yes / no	
Do all aspects of the policy comply fully with applicable national laws, existing ethical guidelines, standards and codes including privacy provisions?	Yes / no	If no, engage with stakeholder groups to redesign the intervention to be more compliant with national laws, existing ethical guidelines, standards and codes
Is the policy consistent with official government directives and goals?	Yes / No	
Are evaluation findings published and discussed transparently with stakeholder groups?	Yes / No	If no, findings need to be published and discussed prior to proceeding with implementation

An ethical governance and implementation		
Is there a dedicated ethical review committee to assess key ethical principles of the intervention, monitor compliance with standards and provide approvals?	Yes / no	If no, establish a dedicated ethical review committee
Is there an ethics supervisor for data collection, use, and storage?	Yes / no	If no, appoint an ethics supervisor for data collection, use, and storage

Have the people conducting observations, data analysis and experimentation received sufficient training to safeguard ethical guidelines?	Yes / no	If no, conduct trainings on compliance with ethical guidelines
Is there a dedicated research team that can conduct rigorous evaluations to assess welfare and cost-benefit analysis?	Yes / no	If no, assign a research team that can evaluate the impact of the project
Is there a dedicated communication team that can engage with stakeholders and manage publication and public discussions on the intervention's impact?	Yes / no	If no, assign a communication manager to manage discussions
Does the intervention's implementation adopt the welfare-maximizing policy design based on testing evaluations?	Yes / no	If no, provide justification and / or re-design the intervention in line with evaluation outcomes
Does the team closely monitor behaviors and outcomes post-launch to ensure positive impacts are realized?	Yes / no	If no, set-up a monitoring stream to ensure target impacts are achieved
Is data collection, analysis and publishing handled in a secure way that ensures confidentiality?	Yes / no	If no, revert back to guidelines
Does the team publicly disclose and explain the reasons for the design choices, including statements at the point at which nudges are delivered?	Yes / no	If no, ensure findings are regularly published to promote transparency
Does the intervention include monitoring of long-term effects?	Yes / no	If no, design monitoring processes to track long term effects

░░ Specific to the ethics of behavioral interventions

REFERENCES

Adler, M. 2012. *Wellbeing and Fair Distribution: Beyond Cost-Benefit Analysis*.

Allcott, Hunt, and Judd B. Kessler. 2019. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." American Economic Journal: Applied Economics, 11 (1): 236-76.

Behaghel, Luc, Bruno Crépon, and Thomas Le Barbanchon. 2015. Unintended Effects of Anonymous Résumés. *American Economic Journal: Applied Economics*, 7 (3): 1-27.

BETA. 2017a. Building Consistent Compliance with Labour Law: Evidence from a Randomized Controlled Trial. Australian Government, Department of Prime Minister and Cabinet.

BETA. 2017b. Going Blind to See More Clearly: Unconscious bias in Australian Public Service shortlisting processes. Australian Government, Department of Prime Minister and Cabinet.

BETA. 2020. The Workplace Experience of Hospitality Apprentices. Australian Government, Department of Prime Minister and Cabinet.

Briscese, G. and C. Tan. 2012. Applying Behavioural Insights to Labour Markets: How behavioral insights cab improve employment policies and programmes. London: Behavioural Insights Team.

Bruns, H., et al. 2018. Can Nudges be Transparent and Yet Effective? *Journal of Economic Psychology* 65: 41-59.

Glaeser, E. 2006. Paternalism and Psychology. *University of Chicago Law Review*, 73 (1):133-156.

Loewenstein, G. et al. 2015. Warning: You Are About to be Nudged. Behavioral Science and Policy Association 1 (1): 35-42.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. United States Department of Health and Human Services.

Paul, L.A. 2014. *Transformative Experience*. Oxford: Oxford University Press.

Sunstein, C. 2016. *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge University Press.

Thaler, R.H., and C.R. Sunstein. 2003. Libertarian Paternalism. *American Economic Review*, 93 (2): 175-179.

Ullman-Margalit, E. 2017. *Normal Rationality*. Oxford : Oxford University Press.

Waldron, J. 2014. It's all for your own good. *New York Review of Books*.