Chapter 4. Applying behavioural insights to integrity policies

Research on human behaviour and moral decisions provides an inspiring perspective on integrity policies. Chapter 4. invites policy makers to explore how behavioural insights can help shape the design of modern integrity policies.

Essentially, there are two approaches to applying behavioural insights to integrity policies:

1. Rethinking existing integrity policies and the design of integrity systems in the light of behavioural evidence. The behavioural approach addresses some of the discontent and criticisms surrounding traditional anti-corruption policies, such as the negative side-effects of overly strong control and enforcement approaches, while making an argument for promoting trust and an organisational culture of integrity. The lessons presented in the previous chapters have implications for the design of integrity systems and related policies. Section 4.1 thus maps out the relevance of behavioural precepts for integrity policies in line with the OECD Recommendation on Public Integrity (OECD, 2017[5]).

2. Designing a specific intervention to fix a behavioural pitfall, improve an existing policy or activate a certain behaviour. The application of behavioural insights through innovation interventions is gaining relevance for public policy making, as featured in the policy streams of the OECD, as well as those of the European Union and the World Bank. A 2017 OECD survey collected more than 100 examples of behavioural interventions in public policies (OECD[3]). Today, 135 public institutions worldwide have institutionalised behavioural capacities. The design of behavioural interventions is based on knowledge of human behavioural biases, cognitive limitations and social preferences. Often the application takes the form of a “nudge”, a subtle change in the way a decision presents itself that helps individuals make a better choice to their own benefit. Section 4.2 examines how integrity policies can benefit from this trend and explores of the established behavioural interventions might be applicable to integrity policy making.

4.1. A human-centred perspective on integrity systems

The OECD Recommendation on Public Integrity provides a shared vision of the core elements of an integrity system. It shifts the focus from ad hoc integrity policies towards a comprehensive, risk-based approach with an emphasis on cultivating a culture of integrity across the whole of society (Figure 4.1). Examining integrity policy making through a behavioural lens places a different perspective on many elements of integrity systems.
4.1.1. Commitment: Demonstrate commitment at the highest political and management levels within the public sector to enhance integrity and reduce corruption

The tone at the top matters for integrity. An unambiguous commitment to integrity sets the scene for how integrity is debated and perceived across the public sector and society. Defining, supporting, controlling and enforcing public integrity sends a clear message that integrity is valued and protected. In terms of the logic of indirect reciprocity, this can function as a motivation for others in the integrity system to uphold those values. Meanwhile, a clear and repeated commitment to integrity can serve to emphasise and raise individuals’ moral reference points: it signals to all public officials that integrity is a crucial part of their professional identity.
Behavioural insights highlight the relevance of attitudes and perceptions for integrity. Integrity policy makers can shape these attitudes and perceptions by participating actively in the public debate. Timely and authentic communication can be a powerful complement to the policies of an integrity system. When a corruption scandal occurs, a public reaction from anti-corruption authorities can prevent discouragement and strengthen trust in the enforcement of integrity norms. Statements from political and institutional leadership create visibility for the actions undertaken in defence of integrity, while reassuring the public that the breach of an integrity norm is not silently accepted.

4.1.2. Responsibilities: Clarify institutional responsibilities across the public sector to strengthen the effectiveness of the integrity system

Shared responsibility can translate into shared guilt in cases of misconduct. As argued above, the diffusion of responsibility poses an integrity risk. Accordingly, clarification of responsibilities across the public sector not only increases the effectiveness of an integrity system, but can also strengthen the integrity of individual decision makers. The institutional design of an integrity system should ensure that responsibility for any type of choice and the related integrity risk is actively taken and perceived to its full extent. To this end, institutional and personal responsibilities need to be clearly defined. Policies that add another actor to a process without giving this actor a unique responsibility, such as the four-eyes principle, do not function as effective tools to control corruption. The same can be said for policies for procedural clearance. A *de jure* requirement does not guarantee *de facto* implementation. On the contrary, the existence of a *de jure* requirement can relieve feelings of responsibility among those providing it as well as among those demanding it (Lambsdorff, 2016[75]). Demonstrating ‘adequate procedures’ or sufficient efforts for integrity cannot replace acting with integrity. Procedural clearance, formal requirements and reviews can never be allowed to replace responsibility for *de facto* integrity. Responsibilities within an integrity system should thus be not only clearly, but meaningfully, assigned.

4.1.3. Strategy: Develop a strategic approach for the public sector that is based on evidence and aimed at mitigating integrity risks

The OECD Recommendation on Public Integrity recognises the relevance of a strategic and evidence-based approach within the public sector. The growing empirical evidence gathered by behavioural science researchers is already feeding into strategic approaches based on sound evidence. From a behavioural perspective, the fact that time pressures and unfavourable working conditions limit people’s capacity for self-control when confronted by opportunities for corruption (see section 2.2), argue for greater strategic planning and prioritisation within the public sector. Staff that are well-equipped, not overwhelmed and strategically guided are more resistant to the temptations of corruption.

However, empirical evidence shows that goals can also corrupt (Ordnez and Welsh, 2015[76]); (Welsh and Ordoñez, 2014[77]). Unrealistic goals create frustration and feelings of unfairness, which can serve as justifications for misconduct (see section 2.4). Faced with overly burdensome compliance rules, staff might cut corners to achieve performance goals and begin to tolerate non-compliance. Moreover, where performance evaluation is closely linked to the achievement of specific performance goals, staff might redirect all their efforts to achieving these performance goals rather than fulfilling their overall role. This often happens at the expense of overarching objectives – such as integrity.
A values-based strategic approach for the public sector –reflects integrity in the overarching strategy, as well as in management goals and performance appraisals. While the integrity-enhancing aspects of strategic management and performance evaluation are recognised, the potentially integrity-damaging aspects of over-compliance and frustration need to be acknowledged.

4.1.4. Standards: Set high standards of conduct for public officials

As can be concluded from Chapter 2, standards of conduct are a behavioural reference: they signal a high moral reference point and contribute to developing professional identities for public employees. To help achieve the behaviour-changing potential of standards of conduct, the following points could be considered.

Standards and guidance for ethical conduct are often derived from a commitment to overarching values. Such values are the frame against which everyday choices and actions can be evaluated. As the number of items humans can store in their working memory is limited, a memorable set of values or key principles ideally has no more than seven elements (Miller, 1956[78]). Box 4.1 gives the example of the Danish Code of Conduct, which defines seven central duties to guide civil servants.

<table>
<thead>
<tr>
<th>Box 4.1. Setting standards for integrity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meaningful and memorable public sector codes</strong></td>
</tr>
<tr>
<td><strong>The Australian Public Service (APS) Values</strong></td>
</tr>
</tbody>
</table>
| In 2010, the Advisory Group on Reform of the Australian Government Administration released its report, which recognised the importance of a robust values framework to a high-performing, adaptive public service, and the importance of strategic, values-based leadership in driving performance. The APS values aim to provide “a small[er] set of core values that are meaningful, memorable, and effective in driving change”. The model follows the acronym “I CARE”:
| • Impartial |
| • Committed to service |
| • Accountable |
| • Respectful |
| • Ethical |
| **The Colombian Integrity Code** |
| In 2016, the Colombian Ministry of Public Administration initiated a process to define a General Integrity Code. Through a participatory exercise involving more than 25,000 public servants through different mechanisms, five core values were selected:
| • Honesty |
| • Respect |
| • Commitment |
| • Diligence |
| • Justice |
| In addition, each public entity has the possibility to integrate up to two additional...
values or principles to respond to organisational, regional and/or sectorial specificities.  

Syv centrale pligter – Seven key duties


The “Kodex VII” describes the relevance and the implications of each duty for the Danish public sector. Moreover, MODST provides fictional case studies that can be used to practice the application of “Kodex VII”. Potential solutions for the case studies are available to public institutions, but are not published.


In addition, standards of conduct are ideally defined through a participative process to ensure that the result is meaningful and actionable for public employees. The overarching values, the language used and the examples provided in a code of ethics need to be able to activate and remind the user of their own moral reference point. The impact on the user will be even greater if the values, language and examples used are personally meaningful and relatable.

The dialogue on standards of conduct does not end after the development of a code. Regularly refreshing one’s commitment to ethical behaviour adds value to the dynamics of moral choice making. Within an integrity system, ethics training and regular dialogue in team meetings could serve this function. Integrity policy makers could also consider approaches inspired by behavioural insights to provoke ideas and discussions on integrity, such as e-mail reminders or competitive quizzes. Section 4.2 might provide some inspiration for such focused interventions.

Public employees who have familiarised themselves with the standards of conduct through training and discussions and have internalised the related values can be encouraged to mentally reactivate their commitment to these standards through regular reminders. References to core public sector values in the work place and in official communications can help to strengthen ethical reflection and serve to strengthen moral choices (see Chapter 2. ). While a mousepad or a pen with the word integrity on it might not seem likely to have a strong impact on a person’s behaviour, they can act as moral reminders and alter actions by establishing a link with previous discussions and a sense of personal commitment. The same can be said for the message “thank you for your honesty” or the honour code in the experiments described in Box 2.3 and Box 3.2.

Ethical reflections making a regular appearance in the everyday routines increase the chance that unethical behaviour is identified and spoken up against. Moral reminders are one way for policy makers to trigger ethical reflections. The research on reminders shows that they work best when introduced in a timely manner. Integrity policy makers could therefore aim to trigger ethical reflection during at-risk procedures. A signature, for example, can function as a tool to invoke stronger personal ethical engagement. Research
shows that a signature has a stronger impact on choices when it is required prior to the decision that poses an integrity risk (Shu et al., 2012[80]).

4.1.5. Whole of society: Promote a whole-of-society culture of integrity, in partnership with the private sector, civil society and citizens

Active dialogue with citizens and regular contact with those affected by public sector decisions amplifies feelings of ethical responsibility within the public sector. Breaking out of a limited sphere of interaction can also help public sector decision makers to feel that their work has implications for people’s lives and reduces the perception that corruption is a victimless crime (see section 2.5). Furthermore, once public sector staff can put a face or a story to the people who place trust in them, they might take more pride in protecting this confidence. An opportunity to engage with stakeholders is often also an opportunity to increase integrity. Otherwise, public employees who only interact with those affected by their work through complaints may become frustrated and eventually lose their motivation to uphold the public interest.

One specific behaviourally informed way to encourage feedback from citizens is the so-called ambient accountability mechanism. This takes the form of a physical element such as a poster or a screen, which is placed directly in the public office for citizens to leave comments (Zinnbauer, 2012[81]). Box 4.2 highlights a particularly innovative approach to obtaining direct citizen feedback, in this case regarding corrupt public officials. In addition to physical feedback, policy makers could embrace the digitalisation of services and procedures as an opportunity to integrate feedback questions.
**Box 4.2. The zero rupee note**

The non-governmental organisation (NGO) 5th Pillar created a zero rupee note in India with the inscription, “I promise to neither accept nor give a bribe”. The organisation has distributed more than 3 million such notes among the public, mostly through students in schools and colleges, all of whom have taken a verbal pledge against corruption.

Handing over the note is a surprising and revealing way to refuse co-operation in a corrupt transaction. Dissatisfied recipients would have to break the silence by initiating a debate in which they do not have the moral high ground. The NGO reports that thousands of citizens who have handed out the notes found that their counterparts immediately yielded their request without insisting on the bribe. One official was supposedly “so stunned to receive the note that he handed back all the bribes he had solicited for providing electricity to a village” (World Bank, 2015).

**Figure 4.2. Zero Rupee Note**

![Zero Rupee Note](https://5thpillar.org/programs/)


### 4.1.6. Leadership: Invest in integrity leadership to demonstrate public sector commitment to integrity

As mentioned above, the tone at the top matters for integrity. However, the top does not refer only to the highest political and management levels. The relevance of middle and lower management cannot be downplayed as their immediate impact on the staff under their direct responsibility may arguably be even higher and more direct. They can help to raise the moral reference point of their employees by promoting discussions about ethical dilemmas or by setting an example with respect to day-to-day conduct in accordance with the shared values of the public service.

Additionally, the preventive approach of building trust and encouraging integrity is ideally implemented by managers other than those responsible for the detection, enforcement and control of corruption (see section 3.4). Managers that have ideally
established a trusting relationship shaped by positive reciprocity are not always the most likely to detect improper conduct among staff. People tend to ignore information contradicting their existing beliefs, sometimes dismiss warning signals and are less likely to judge the behaviours of those close to them as unethical, behavioural research shows (Gino and Galinsky, 2012[83]); (Eil and Rao, 2011[84]); (Hildreth, Gino and Bazerman, 2016[85]). In addition, the enforcement integrity among staff often involves communicating distrust. Managers who implement strict measures of control, or more severe, sanction their staff, risk crowding out their intrinsic motivation and losing their trust (see section 2.7).

4.1.7. Merit based: Promote a merit-based, professional public sector dedicated to public service values and good governance

Integrity fundamentally concerns human behaviour. As such, the civil service and human resource management play a key role in its promotion. Per definition, integrity is indirectly part of every job description in the public sector. Accordingly, “I’m only doing my job” is never an excuse for unethical behaviour on the part of any public official. Behavioural research has long understood that people tend to act according to the roles they are given. Priming a certain social identity can lead people to act according to the characteristics of this identity (Benjamin, Choi and Strickland, 2010[86]).

Public employers need therefore to clearly communicate to their staff that they are expected to uphold the public interest and give them actionable guidance on how to do so, for instance through regular integrity training. Where integrity is clearly linked to the professional identity associated with a position in the public sector, acting with integrity becomes an act of pride and duty. The prominence of integrity for the role can be signalled to new applicants for public sector positions already in job postings and form an integral part of the selection and recruitment process. The Office of Government Ethics in the United States requires all written offers of employment in the executive to include a statement regarding the agency’s commitment to ethics and to set role-specific expectations early. New supervisors are alerted of their special responsibility for advancing government ethics. Box 4.3 provides an example from Australia on how to integrate integrity into selection processes.
Box 4.3. Recruitment processes and integrity – experience from Australia

“Filters” can be built into a recruitment process to ensure applicants are suited to an organisation’s requirements. In Australia, one agency analysed disciplinary issues among new recruits after 12 months on the job and identified a need to better manage integrity indicators earlier in the selection process. Interventions were then instituted at key stages:

- A question and answer survey was incorporated into the general information for potential applicants. It asked questions about how candidates felt about certain working conditions and interactions. Based on an indicative score, potential applicants were encouraged to proceed to the next stage or to speak about the role with people who knew them well before proceeding to the next stage. This approach encouraged self-filtering by applicants.
- The application asked more targeted integrity questions about the background and experiences of applicants (e.g. dealing with authority, diverse cultures and financial management). These provided base data for comparative purposes.
- Successful applicants in the technical assessment phase were asked to retake the integrity questions. Experts were then asked to identify discrepancies or anomalies between the datasets and to follow these up individually with applicants. The delay between administering the questions increased the validity of the data.
- Only applicants who passed both the technical and the integrity phases were invited for face-to-face interviews, which included a practical role play.

The outcome was a considerable decrease in disciplinary issues and increased retention rates for new recruits.

Source: Australian Merit Commission, June 2016.

However, the understanding that public employees have with respect to their own role and professional identity can diverge widely from that of the management. Conducting regular public sector staff surveys and encouraging ethical reflections can contribute to maintaining awareness of integrity as part of the identity of public employees.

The relationship between a public employer and its employees is more than an exchange of labour for payment; it is shaped by reciprocity. Employees who receive a generous wage are willing to work harder, even where this is not observable (Akerlof and Yellen, 1990[47]; (Gilchrist, Luca and Malhotra, forthcoming[88]); (Van Veldhuizen, 2013[89]). Their behaviour is shaped by the way they are treated. Non-monetary incentives are even more effective in activating intrinsic motivation (Gneezy, Meier and Rey-Biel, 2011[90]); (Kosfeld, Neckermann and Yang, 2014[91]); (Kosfeld and Neckermann, 2011[92]). Employees who feel they are trusted, appreciated and recognised will be motivated to reciprocate through strong performance (Falk and Kosfeld, 2006[44]); (Fehr and List, 2004[93]). Employees who are treated with integrity are more likely to uphold this value in their daily work.

Individuals who are failed by an institution, a norm or their superior might eventually become frustrated and stop contributing to the public good of integrity. Employees who feel treated unfairly could feel licensed to act unethically (see Chapter 2. ). Feeling
mistreated not only reduces the obligation of reciprocity towards the employer and the employer’s norms, but acts as a gateway to justifications for misconduct (see section 2.4). For example, an individual disadvantaged during a hiring process because another applicant exploited personal connections, might lose trust the integrity of the system and decide to rely on connections themselves in the future. If the number of people who have lost trust rises, the integrity norm will start to fall apart. Organisational fairness is a crucial condition for integrity (OECD, 2009[94]).

Modern managers therefore seek to establish a trusting and appreciative relationship with their employees. Integrity should be a visible part of this relationship. Internal communication, goal setting and role descriptions could include integrity as a core objective and subject of the employer-employee relationship. Consequent communication of ethical standards makes clear that integrity is expected of the members of the organisation, who can in turn expect the organisation to respect integrity as an organisational value. Evidently, such claims must be applied in practice. Employees can be discouraged to uphold norms and principles that they feel are not respected throughout the organisation (see Chapter 3).

In an open organisational culture that aims to prevent integrity breaches, it is thus crucial that staff feel they are treated fairly, that errors are tolerated and that they have the opportunity to voice their concerns should that change. Ensuring that integrity is appreciated and that speaking up is not penalised promotes integrity as a reciprocating norm. Settings in which those who adhere to the rules have to bear additional costs (e.g. increased paperwork, conflict with colleagues, longer waiting times) should be avoided. As noted above, organisations could monitor the perceptions of staff through regular employee surveys. Employee complaint mechanisms could also be in place to ensure conflicts are addressed. Any suppressed feelings of sacrifice or under-appreciation among employees will pose an integrity risk.

4.1.8. Capacity building: Provide sufficient information, training, guidance and timely advice for public officials to apply public integrity standards in the workplace

Integrity training offers a unique opportunity to emphasise and reinforce shared values, and can thus emphasise and contribute to raising the moral reference point of participants. Induction trainings, for instance, which take place at the very start of a public servant’s career, communicate the values of the respective public organisation and provide guidance on the expected behaviour in the social context. Training well aligned with the day-to-day reality of the organisation face a lower risk to be perceived as naïve or to be seen merely as a formal requirement. For sustainable capacity building, training needs to be repeated over time, and may be adjusted to the needs of specific target groups such as at-risk positions or management.

Interactive components where participants are confronted with realistic situations are more likely to generate a personal mental commitment to integrity than mere presentations by trainers (Bazerman and Tenbrunsel, 2011[20]). Box 4.4 provides the example of dilemma training implemented by the Flemish Government in Belgium. The design of training can also build on behavioural insights. For instance, training can benefit from confronting participants with past situations where they likely behaved in an unethical way. The participants then need to resolve this moral dissonance by performing a positive action.
Box 4.4. Dilemma training in the Flemish Government (Belgium)

The Agency for Government Employees in the Flemish Government offers dilemma training to public officials. The training provides the participants with practical situations in which they face an ethical choice with no clear path to resolving the situation with integrity. The facilitator encourages discussion between the participants about how to resolve the situation and helps them explore the different choices. The focus of the training is the debate rather than possible solutions, as the objective is to help participants identify how different values might act in opposition to one other.

In the majority of training sessions the facilitator uses a card system. The participants receive four “option cards” with the number 1, 2, 3 or 4. A set of “dilemma cards” are then placed on the table. Each dilemma card describes a situation and gives four options for resolving the dilemma. In each round, one of the participants reads out the dilemma card and the options. Each participant indicates their choices using their option cards and explains the motivation behind their choice. The participants then discuss the different choices. The facilitator remains neutral, encourages the debate and suggests alternative ways to look at the dilemma (e.g. sequence of events, boundaries for unacceptable behaviour).

One example of a dilemma situation could be:

“I am a policy officer. The Minister needs a briefing within the next hour. I have been working on this matter for the last two weeks and should have already been finished. However, the information is not complete. I am still waiting for a contribution from another department to verify the data. My boss asks me to submit the briefing urgently as the Chief of Cabinet has already called. What should I do?

1. I send the briefing and do not mention the missing information.
2. I send the briefing, but mention that no decisions should be made based on it.
3. I do not send the briefing. If anyone asks about it, I will blame the other department.
4. I do not send the information and come up with a pretext, and promise to send the briefing tomorrow.”

Other dilemma situations could cover themes such as conflicts of interest, ethics, loyalty, leadership and so on. The training sessions and situations used can be targeted to specific groups or entities.


Beyond training, the OECD Recommendation on Public Integrity also emphasises the importance of the availability of guidance when public officials are faced with specific doubts or ethical dilemmas. Although integrity is ultimately the responsibility of all organisational members, dedicated “integrity actors” can play a particularly important role in stimulating integrity and shaping ethical behaviour (OECD, 2009[94]). In addition, guidance on ethics and conflict of interest in cases of doubts and dilemmas can respond on an ad hoc basis when public officials are confronted with a specific problem or doubts. Written guidance, such as the REFLECT model provided in Australia (Box 4.5), together with complementary trainings can support public officials in making the correct moral choices.
Box 4.5. Guiding public officials facing ethical dilemmas in Australia

To support the implementation of ethics and its integrity regime, the Australian Public Service Commission has enhanced its guidance on the Australian Public Service (APS) Values and the Code of Conduct. This includes integrating ethics training into learning and development activities at all levels.

To help public servants address ethical dilemmas during the decision-making process, the Australian Public Service Commission developed a decision-making model, known as REFLECT:

1. **REcognise** a potential issue or problem. Public officials should ask themselves:
   - Do I have a gut feeling that something is not right or that this is a risky situation?
   - Is this a right vs right or a right vs wrong issue?
   - Recognise the situation as one that involves tensions between APS Values or the APS and their personal values.

2. **F**ind relevant information
   - What was the trigger and circumstances?
   - Identify the relevant legislation, guidance, policies (APS-wide and agency-specific).
   - Identify the rights and responsibilities of relevant stakeholders.
   - Identify any precedent decisions.

3. **L**inger at the “fork in the road”
   - Talk it through, use intuition (emotional intelligence and rational processes), analysis, listen and reflect.

4. **E**valuate the options
   - Discard unrealistic options.
   - Apply the accountability test – public scrutiny and independent review.
   - Be able to explain your reasons/decision.

5. **C**ome to a decision
   - Come to a decision, act on it and make a record if necessary

6. **T**ake time to reflect
   - How did it turn out for all concerned?
   - Learn from your decision.
   - If you had to do it all over again, would you do it differently?

*Source: (Godwin, 2009)*

4.1.9. **Openness: Build an open organisational culture within the public sector responsive to integrity concerns**

Integrity breaches often remain unchallenged in their immediate environment due to the strong behavioural forces of reciprocity, the power of unconscious biases, and the
tendency towards conformity of judgement and behaviour within groups. Corruption is often the end of a slippery slope consisting of small misjudgements and gradual acceptance. Moreover, research points to the fact that leaders and those who identify strongly with their organisation present an even higher risk of letting integrity breaches go unchallenged. (Barkan, Ayal and Ariely, 2015[15]); (Kennedy and Anderson, 2017[96]). Managerial oversight is therefore insufficient as the only defence of integrity.

The ability to voice concerns in the face of misconduct is key to upholding an integrity system. Public employees who witness wrongdoing are expected or required to report it. However, employees may be tempted to turn a blind eye to irregularities to avoid trouble for their colleagues or organisation. Behavioural insights can help to understand and overcome this dilemma.

Speaking up requires a strong organisational culture. In organisations lacking a culture of openness to discuss ethical concerns or dilemmas, whistleblowers put their personal relationship with their colleagues and even their job at risk to protect the public interest. A team member who voices concern about the behaviour of another team member questions this colleague’s personal integrity. Regardless of whether the disclosure is based in fact, it may affect mutual trust among team members (Reuben and Stephenson, 2013[97]). Whistleblowers are sometimes also suspected of personally benefiting from the situation, for example, by wanting to inflict harm on the person they are reporting. In cases where direct and indirect reprisals are not prevented or mitigated, employees will need courage and a strong moral compass to speak up in the light of misconduct.

However, protecting the public interest may require speaking up against a respected colleague, an established practice or an accepted system. Public sector employers emphasise integrity hoping that their staff will voice their concerns, but it would be naïve to rely solely on the categorical morality of public officials under such circumstances. As noted in Chapter 2, moral choice is a dynamic process in which people also consider their own self-interest. Whistleblower policies stand to gain by acknowledging this and aim to reduce the burden a whistleblower has to carry to the extent possible.

From a behavioural perspective, the following two aspects can facilitate speaking up:

- **Certainty**: Potential whistleblowers will evaluate the act in light of what follows. Uncertainty about the procedure or the consequences of reporting makes it harder to take this step. Whistleblowing policies should thus clearly and credibly state the process of reporting, make this information known to all staff, and make those who are responsible for handling disclosure of misconduct accountable for their actions. In particular, whistleblowers should be provided with as much certainty as possible regarding whether they will be granted protection and whether their case will be investigated. Furthermore, whistleblowers will speak up with more confidence if they know which behaviours constitute misconduct. A broad definition of the types of misconduct eligible for reporting and knowledge and awareness of this definition could increase use of the whistleblowing system.

- **Confidentiality**: Just as control can act as a burden on the trust between managers and their team, voicing an integrity-based concern can burden a public employee’s relationship with their managers and colleagues. Public officials might choose to avoid risking longstanding work relationships, especially on the basis of suspicion. Protecting the identity of whistleblowers, for example, through a third-party reporting mechanism or confidential digital reporting system, allows them to seek advice on an issue without disrupting their team relationships. Once they have
gained confidence that the incident in question is indeed unethical, and have greater certainty regarding the consequences, they might choose to reveal their identity.

Once again, a dedicated integrity contact point, either a unit or a person, can be conducive both to certainty and confidentiality. An external person serving as contact point could provide ethical guidance and procedural advice to public officials, while keeping their identity confidential. With this support, the public official can make an informed decision about disclosing misconduct. When an external contact points is well-known, trustworthy and clearly separated from enforcement functions, this increase the likelihood that people will turn to it.

Protection from reprisal or effective remedies, interim relief, legal support or honorary awards can be other ways to show appreciation of a whistleblowers dedication to integrity. Moreover, thorough investigation of the reported case and sanctioning of misconduct honour integrity. If there is no certainty that the reported cases will be consequently followed-up with, then taking on the cost of reporting is not reasonable for a potential whistleblower.

The logic of reciprocity (see Chapter 3. calls for whistleblowing to be rewarded. By exposing behaviour that can potentially harm the organisation as well as the public interest, whistleblowers mitigate financial and reputational risks within an organisation. However in an organisation where the most members are generally committed to integrity appreciation for whistleblowers does not necessarily need to be expressed through a monetary reward. In fact, a monetary reward can in some contexts even discourage speaking up. Incentives can have a hidden cost, that is, a counterproductive effect on a trust-based relationship (Fehr and List, 2004[93]). Monetary incentives might make the actions of the whistleblower appear to be based on self-interest rather than principle. People’s estimation of a moral choice is generally increased when the individual in question acts without direct reward (Hoffman, Yoeli and Nowak, 2015[98]) see also section 3.1). Nonetheless, an organisation within a collective action trap (see 3.5) where norms are already deterred and whistleblowers fear severe vengeance can use monetary rewards as an incentive to break the silence in corrupt networks.

As Chapter 3. emphasised, integrity policy makers could seek to strengthen indirect reciprocity. In this sense, an integrity system encourages whistleblowers to act with integrity not for personal monetary reward (direct reciprocity), but to protect the shared norm. Nonetheless, a person who places much at stake to protect integrity will usually hope that others value this dedication to the public interest (indirect reciprocity). Such appreciation can be integrated in whistleblowing systems (e.g. by awarding whistleblowers) and wider human resource management policies (e.g. through positive recognition of whistleblowing in performance appraisal).

Ways of showing appreciation of a whistleblower’s dedication to integrity include protection from reprisals or effective remedies, interim relief, legal support and honorary awards. Integrity is also honoured through thorough investigation of the reported case and punishment of misconduct. If there is no certainty that the reported case will be investigated, the cost of reporting might not be perceived as reasonable by a potential whistleblower.

Sometimes employees do not report misconduct simply because they do not perceive it as such. Conformity bias (Asch, 1952[63]); (Asch, 1951[99]), self-serving social identity and justifications can prevent people from becoming aware of ethical misconduct within their own work environment (see sections 2.4 and 0). A behaviour that is accepted by
colleagues or managers is less likely to be identified as unethical. Integrating regular ethical reflection into public officials’ work routines can counteract such effects. This could take the form of a moral reminder message, a refresher ethics workshop or a regular discussion of integrity topics with an external third party. Most importantly, these activities need to create salience for ethical standards and invite public officials to reflect on whether these standards are being upheld in their own work.

Finally, whistleblowers who are themselves involved in the misconduct in question might be deterred by the prospect of sanctions. Will the team’s reputation be destroyed or all past work declared void? Will they personally be under scrutiny? Box 3.3 in Chapter 3 describes a case where an actor was scared to report misconduct because he had unintentionally become involved. Fear of sanctions can create confidence among corrupt partners that no-one involved will sever the reciprocal arrangement. Asymmetric liability in whistleblower law can break this dynamic of extortion. Mitigation or limited sanctions for confidents who speak up can encourage such insiders to become whistleblowers (Abbink et al., 2014[100]); (Apesteguia, Dufwenberg and Selten, 2007[101]); (Buccirossi and Spagnolo, 2006[72]); (Lambsdorff and Nell, 2007[73]).

4.1.10. Risk management: Apply a control and risk management framework to safeguard integrity in public sector organisations

Supporting ethical conduct without surrendering control safeguards is a central challenge for any integrity system. Internal controls, audits and risk assessment policies are key policies concerned with monitoring and managing integrity risks. Sometimes these controls can be automatised and thereby delinked from human behaviour. In many cases, however, those controlling and inducing the risk are humans. Behaviourally informed control and risk management policies designed to anticipate human ethical blind spots can help reduce errors.

Human error can be a challenge for assessments that rely solely on human judgement. However, humans often err systematically; thus, certain misjudgements can be anticipated and avoided. For example, it has been established that humans struggle to correctly assess probabilities (Kahneman and Tversky, 1979[103]). Estimating both the likelihood and consequences of an event at the same time is cognitively challenging. Faced with such complex estimations, people tend to let one or the other guide their judgement. Focusing on the probability, an assessor might underestimate an integrity risk, because they think the event is unlikely to occur. When judging the probability of an incident, risk assessments could be deterred by the so-called representative heuristic (Kahneman and Tversky, 1979[102]): people tend to associate higher probabilities to events they can imagine or have experienced. A person might assume an integrity risk is very unlikely to occur, simply because they have never personally experienced it or because it has apparently never happened in the organisation before.

In response to this, the subject matter of any assessment could be broken down and stripped of all redundant information that could bias the decision maker. A risk assessor could, for example, estimate the impact of an event independent from its probability of occurring. Furthermore, auditors and other individuals that regularly conduct independent assessment could be made aware of potential bias and be trained to counteract it.

Instead of asking people directly about their own integrity risk, it might be better to pose general questions or ask questions about a similar person in a similar organisation. When people are asked to assess their own integrity risks, they are bound to be biased in favour of themselves (see Chapter 2 ). Positive moral self-assessment clouds their judgement,
leading them to underestimate integrity risks related to their own role, as compared with the judgement of an outside observer. Along the same line, leaders might overestimate their own ability to prevent an integrity breach. People with high trust in their organisation might underestimate the likelihood of an integrity breach among their colleagues. Posing self-assessment questions more indirectly can reduce overconfidence and social desirability bias. For example, an interviewer could extract sensitive information by asking, “How likely do you think it would be for someone in a similar position to have done this?”

Certain roles have higher integrity risks and are therefore more closely examined in risk assessments and internal audits. Behavioural research can provide a few insights to inform the criteria for identification of these roles:

- People with power tend to judge their own integrity breaches more leniently (Barkan, Ayal and Ariely, 2015[15]).
- Higher ranks statistically identify more with their organisation. This weakens their ability to identify an integrity breach among their own colleagues (Kennedy and Anderson, 2017[96]).
- People who are at risk of exclusion from groups are more likely to act in an unethical manner for the sake of the group (Derfler-Rozin, Pillutla and Thau, 2010[103]); (Thau et al., 2015[104]).
- As discussed in section 2.5, the perceived immorality of an action decreases with distance from the potential victims of an action. Individuals whose corrupt choices impact people they are never likely to meet thus have a higher integrity risk than those who interact closely with those whose lives they impact regularly.

4.1.11. Enforcement: Ensure that enforcement mechanisms provide appropriate responses for all suspected violations of public integrity standards by public officials and all others involved in the violations

Detection and sanctioning of corruption help build trust (see section 3.4). Enforcement of formal norms gives those committed to integrity a sense of fairness. Separating such enforcement from encouragement allows both functions to achieve their full psychological strength (Lambsdorff, 2012[55]): Human resource policies that are by nature distrustful (e.g. reporting requirements, auditing, checking attendance) are more convincing when communicated separately from policies built on a trusting relationship (e.g. performance measurement, code of ethics). Ideally, such policies are introduced by different institutions or individuals with different roles. Integrity as a subject of dialogue between manager and staff can then be framed in a positive, motivating manner. For example, integrity can be communicated as part of organisational identity and as a performance goal.

In turn, it is important to prevent a perception of impunity as this would undermine the credibility of rules and procedures, promote cynicism among staff and provide individuals with arguments to justify their own integrity breaches (see section 2.4). As emphasised in section 3.5, organisations entrapped in a collective action trap, where it is rational for individuals to adapt to the corrupted context, require strict enforcement in order to change the organisational culture. Thorough investigations of integrity breaches, credible sanctions and impartial enforcement signal a valuation of integrity and provide certainty that laws and regulations cannot be circumvented without punishment (see sections 3.4 and 3.5). Indeed, it is important to emphasise that sanctions are not necessarily limited to criminal sanctions. According to the severity of the infringement, a
4. APPLYING BEHAVIOURAL INSIGHTS TO INTEGRITY POLICIES

4.1.12. Oversight: Reinforce the role of external oversight and control within the public integrity system

External auditors examine facts and numbers in accordance with clear-cut guidelines and standards. They are also subject to strict controls to ensure they do not falsify their work. However, sometimes “good accountants do bad audits” (Bazerman, Moore and Loewenstein, 2002[105]): Despite having the external appearance of a hard science, auditing leaves considerable leeway for leniency and ambiguity. Wherever facts depend on interpretation there exists the possibility for unconscious bias.

Auditors can be made aware of and trained to counteract biased judgement. Sections 2.3 and 2.4 describe how mental shortcuts and misperceptions can lead humans to make biased judgements and yield to unethical behaviour without even being aware of their actions. This phenomenon is often referred to as “ethical blindness” (Bazerman and Tenbrunsel, 2011[20]); (Sezer, Gino and Bazerman, 2015[106]). Auditors and other individuals responsible for oversight functions could make use of behavioural insights and evidence to compensate for such unintentional blind spots. Although knowledge about behavioural biases alone is not sufficient to ensure they will have no impact on others, an awareness of ethical blindness allows decision makers to strengthen their resilience against typical sources of bias, such as overconfidence, conformity, justifications and representativeness bias.

Individuals performing auditing roles can protect their impartiality by performing only auditing functions. Providing consulting or standard-setting services in addition to auditing can challenge the objectivity of auditors, an issue implicitly recognised by the International Organization of Supreme Audit Institutions (International Organization of Supreme Audit Institutions, 2010[107]). In one role they flag problems, while in the other they design solutions. To avoid conflict of interest, auditors could refrain from performing both tasks for the same organisation or assessing a solution they designed or implemented themselves. Even where conflict of interest is well-managed, the psychological challenge of a double role can remain and potentially result in unconsciously biased judgements.

Consequently, clearly separating internal control activities from internal and external audit functions could strengthen their objectivity. While the responsibility for management of integrity risks and functions of internal control is normally shared within an organisation, internal audits are clearly a checking and monitoring function. Autonomy and impartiality – also mentally – are essential to these functions. In addition, perceived close links between audits and the internal control environment could harm the sense of ownership of internal control and risk management among public managers.

4.1.13. Participation: Encourage transparency and stakeholder engagement at all stages of the political process and policy cycle to promote accountability and the public interest

Implicit biases and the intuitive tendency to reciprocate can create an unconscious bias (see section 2.2). Judgments considered to be objective by those who make them might actually distorted by misperceptions. The risk of such unconscious bias increases with psychological closeness, resulting in a tendency to favour people similar to oneself or
with whom one interacts frequently. Where unconscious bias functions as an entry point for undue influence, it becomes an integrity risk. This is particularly relevant for individuals expected to make impartial decisions, as is often the case for policy makers and regulators.

Frequent interaction with different stakeholders is often essential for the work of key decision makers. Regulators, for example, interact with ministries, parliaments, regulated industries and citizens, who are the ultimate beneficiaries of governmental and regulatory actions. “These interactions are inevitable and desirable” (OECD, 2016[108]), but they can also impact decision making. People use mental shortcuts, such as the “availability heuristic” (Tversky and Kahneman, 1973[109]), according higher relevance to things that they can recall, put an example to or have recently experienced. Thus, without any exertion of undue influence, a frequent exchange with a particular group can bias decisions towards this group’s interests.

A balanced consideration of interest cannot be achieved by prohibiting such exchanges, but rather by ensuring the equal representation of interests in the minds and perceptions of decision makers. It is therefore important to ensure a level playing field in terms of access to the policy-making process and the policy cycle for all interest groups (OECD, 2017[31]).

4.2. “Nudging” for integrity

Behavioural insights and their use in policies have been popularised through low-cost, small interventions, often called “nudges”. Such interventions are now being applied in public policies around the world (OECD, 2017[3]). Their use is usually intended to overcome a particular behaviour-related challenge within a policy design (e.g. when people forget to sign up for a programme).

However, integrity is too complex and broad an issue in behavioural terms to be defined as the intended outcome of a “nudge”. Nonetheless, behavioural interventions can be used to increase the effectiveness of integrity policies or procedures within an integrity system. The first step is to identify challenges related to behavioural concepts within an integrity system. Behavioural interventions can then intervene wherever the success of a policy or procedure is compromised by subjective perceptions and judgements on the part of decision makers.

The way in which choices present themselves can alter the outcomes of a decision. For example, humans are generally averse to losses. As a result, most people behave differently in the face of a potential loss compared to a potential gain. This preference also affects ethical choices: experiments showed that people were more likely to act unethically to prevent a loss than they would to make a gain (Kern and Chugh, 2009[110]). The amount that could be won or lost was the same in both cases – the difference was merely a matter of phrasing.

A nudge alters the choice architecture – the way in which people are presented with a choice – without limiting the options involved. It builds on behavioural research mapping the various biases in human perception and decision making. By considering all the ways in which human choices differ from rational predictions, the choice architecture can help people make a good choice in their own interest. Typical alterations to the choice architecture include (OECD, 2017[111]):

- changes to wording, naming and communication
changes to the physical environment
changes to procedures and forms
use of feedback and interaction
rewards, goal setting and commitment.

The identification of a potential area of application for behavioural interventions requires a detailed analysis of the problem to be solved. Figure 4.3 illustrates this process. Not every integrity challenge can be fixed with a behavioural intervention, but behaviours are involved in almost any integrity policy. Whenever this behaviour is “suboptimal” – meaning that the choices of the individuals might be falling short of achieving their own best interest – a behavioural intervention can help.

**Figure 4.3. Targeting a behavioural intervention**

For example, assume that an organisation has a confidential integrity contact point, but that this person or unit is not often contacted regarding advice about potential integrity breaches. If the underlying reason is that the contact point does not provide competent advice, it is in the interest of people not to approach them. Instead of a behavioural intervention, this situation requires a policy re-design to rethink or strengthen the capacities of the contact point. However, it is also possible that people may not be aware of the existence of the contact point. In such cases, a behavioural intervention could be used to remedy the situation by raising awareness of the existence of the contact point.
Where the concerned behaviour is central to the policy and a strong bias or blindness is affecting the choice, the impact of a small and cost-effective behavioural intervention can be significant. Although behavioural interventions are no panacea, it is worth investigating their potential applications. In cases where behavioural insights cannot necessarily be used to design a whole new policy, they can still make conventional policies more effective. For example, behavioural insights will probably not yield a whole new approach to asset declarations, but they could be used to nudge people to declare their assets on time or to make declaration forms more user-friendly.

Behavioural policy making has developed a considerable repertoire of policy responses to encourage certain choices and counteract biases. Common tools for choice architecture changes include the following:

- **Default settings**: People show a certain inertia with regard to making active choices. This is why one of the most powerful nudges is to change the default (e.g. default inclusion of travel insurance when buying flight tickets).
- **Reminders**: The human mind is a messy place in which various thoughts are constantly battling for attention. Introducing a piece of information at just the right moment can provoke people to act upon it.
- **Framing**: The same piece of information is often perceived very differently depending on how it is presented. Framing nudges often make use of this fact to make information more actionable.
- **Simplification and increased convenience**: Complexity frustrates and demotivates. Making something easier increases its chances of being completed.
- **Gaming**: Humans like to play. Curiosity and ludic urges are great motivators for participation and making efforts.
- **Commitment**: People sometimes fall short of their own intentions. Creating a commitment beforehand increases the chance that they will follow through.

Table 4.1 provides some examples of how these common approaches could be used within an integrity system. However, the application of behavioural insights in specific interventions needs to be adapted to the context and respond to the individual behavioural problem. The list presented in Table 4.1 is thus just a starting point for making integrity policies more efficient, more effective and more human.
### Table 4.1. Use of tools for choice architecture change in an integrity system

<table>
<thead>
<tr>
<th>Tool</th>
<th>Typical application</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>Setting default in procedures</td>
<td>At the beginning of a new project each team member automatically receives a form to declare any conflict of interest</td>
</tr>
<tr>
<td>Reminders and framing</td>
<td>Moral messaging</td>
<td>“Remember, you should report and surrender the gifts you receive throughout the year. Show your honesty! Report your gift here.”</td>
</tr>
<tr>
<td></td>
<td>Personalised messaging</td>
<td>“Dear Tina, do not forget to take the ethics survey.”</td>
</tr>
<tr>
<td></td>
<td>Identity prime</td>
<td>“As a good police officer, you are committed to integrity…”</td>
</tr>
<tr>
<td></td>
<td>Value prime</td>
<td>“Receiving gifts compromises your impartiality as a public official.”</td>
</tr>
<tr>
<td></td>
<td>Social prime</td>
<td>“96% of your colleagues have already submitted their asset declarations…”</td>
</tr>
<tr>
<td>Simplification and convenience</td>
<td>Reducing the number of steps in a process</td>
<td>Submission of asset declaration through the electronic system</td>
</tr>
<tr>
<td></td>
<td>Chunking (remembering no more than 7 items at a time)</td>
<td>Group the paragraphs of the code of ethics</td>
</tr>
<tr>
<td></td>
<td>Visualisation</td>
<td>Visual comparability of tenders in a graphic</td>
</tr>
<tr>
<td></td>
<td>Colour priming</td>
<td>Use of a red button to submit a sensitive decision</td>
</tr>
<tr>
<td></td>
<td>Convenience</td>
<td>Asking for the asset declaration at the same time as the tax filings are due</td>
</tr>
<tr>
<td>Gaming</td>
<td>Lottery</td>
<td>“You have been selected…”</td>
</tr>
<tr>
<td></td>
<td>Games, quizzes</td>
<td>… to participate in an ethics quiz on the elements of the Code of Ethics…</td>
</tr>
<tr>
<td></td>
<td>Competitions</td>
<td>… players from each division compete…</td>
</tr>
<tr>
<td></td>
<td>Micro-incentives</td>
<td>… and the winning division receives a delicious breakfast at the office</td>
</tr>
<tr>
<td>Commitment device</td>
<td>Self-commitment</td>
<td>“I hereby declare that all information entered below in this online recruitment form will be truthful”</td>
</tr>
<tr>
<td></td>
<td>Actionable dilemma training</td>
<td>Considering ethical dilemmas in advance and committing beforehand to handling them well (e.g. preparing a diplomatic response to a dinner invitation that one should not accept).</td>
</tr>
</tbody>
</table>

Source: Inspired by (Sunstein, 2014[112]), as well as (European Commission, 2016[2]); (OECD, 2017[111]); (Van Bavel et al.,(n.d.))[113]).