Chapter 3. Integrity in the context of social interactions

Moral choices are not taken in isolation. Human decisions are often driven by social motives such as loyalty, trust building, returning favours or helping someone out of a tricky situation. People take decisions in their own best interest, but they also care what others think or do. Social motives can work in favour of integrity, but they can also function as reasons for corrupt decisions. Understanding why humans as social decision makers are swayed by their own preferences, as well as those of others, can provide insights into why and how integrity policies work or fail.

3.1. Transparency is not an end in itself

The moral wiggle room for a decision maker reduces when other actors are involved. An individual when alone will be the first to be convinced of the morality of his or her actions. Outside observers who are not biased might not be as easily convinced by the justifications offered. The opinion of outside observers usually matters to a decision maker: people prefer to act in a self-serving manner, but at the same time like to appear moral to others (Batson et al., 1999[46]). Transparency and accountability mechanisms could thus reduce unethical behaviour. The perception that one’s behaviour is visible, and potentially observed, introduces an element of accountability that makes it more difficult to apply justifications (see section 2.4), because potential observers could easily detect an excuse.

However, in the absence of interaction with an observer, such effects might dissipate. Transparency without interaction can be compared to a security camera without anyone to watch the tapes; those monitored may forget its existence and behave as if they were alone. Many integrity policies introduce transparency, for example, through open government and open data strategies; however, transparency is not an end in itself. From a behavioural perspective, approaches to promote transparency can have a significant impact if they trigger dialogue. Not all elements need to be subject to regular interaction, but regular attention to specific elements can increase the sense of social observability for all transparent actions. For example, a committee whose work is publicly observable and who, every once in a while, receive comments or complaints from citizens, might feel more liable to the public. The regular reminder that their decisions affect citizens reduces the perceived distance between action and harm, and thus limits moral wiggle room. This effect is not achieved by transparency, but rather through actively engaging with the respective stakeholders. Transparency is thus a necessary condition to achieve this desirable effect, but is not sufficient in and of itself.

The interaction between transparency and integrity is not always straightforward. In experiments, people who disclosed a conflict of interest afterwards gave biased advice in a more brazenly manner (Cain, Loewenstein and Moore, 2011[47]); (Cain, Loewenstein and Moore, 2005[48]); (Loewenstein, Sah and Cain, 2012[49]). This finding reveals a diffusion of moral responsibility in transparency: the person who disclosed the information feels that they can act upon their bias once it is openly admitted. They feel
less responsibility to balance their decisions, based on reflections such as “I didn’t hide anything” and “If this is a problem, someone will tell me”. Transparency may thus provide absolution for an unethical choice. Who granted the absolution? The other participants in the experiments usually did not expect biased advice. On the contrary, they mistook the disclosure of information as a sign of trustworthiness (Sah, Loewenstein and Cain, 2013[50]). In reality, those affected by the decision might not even be aware of the conflict of interest. In scenarios where significant volumes of information are disclosed under open government and open data policies, the onus to filter relevant information and discover a potential conflict of interest falls on the media and civil society. If a conflict of interest is hidden in disclosed information but remains undetected, responsibility may become lost somewhere between the disclosure and the reception of the information. In such cases, untargeted transparency without a clear recipient might even reduce accountability for integrity.

3.2. The role of reciprocity for integrity

The social determinants of ethical behaviour can be widely understood with one concept: reciprocity. Reciprocity is a social behaviour that can be observed – to different degrees – in every culture of the world, and is fundamental to the human species (Bowles and Gintis, 2011[51]); (Camerer, 2003[52]); (Henrich et al., 2005[53]). Someone who receives kindness will return the favour. If treated unfairly, people seek revenge. This simple reasoning lies at the core of human logic and constitutes the foundation of market exchange, trust and fairness (Fehr and Gächter, 2000[54]).

However, reciprocity also lies at the core of many corrupt interactions. Since corruption is clandestine, a corrupt deal cannot be enforced through formal institutions such as those provided by the justice system. To some degree, the actors have to trust that their counterparts will follow through as expected (Lambsdorff, 2012[55]). The persistence of human reciprocity allows them to trust that they will obtain the services that they paid for with bribes; there is such a thing as “honour among thieves” (Husted, 1994[56]).

Essentially, reciprocity can function as a motivation or excuse for engaging in corruption. Box 3.1 presents three settings where reciprocal norms might make corruption seem less immoral. For example, involving several actors in a decision can lead to a diffusion of responsibility. Therefore, the involvement of additional actors in decisions (e.g. a requirement of approval by at least two equally responsible individuals known as the four-eyes principle) might not necessarily be an effective approach to promoting integrity.
Box 3.1. Social preferences as justification

Hiding behind good intentions can deter people from admitting the actual moral implications of misconduct to themselves or others. Typical justification patterns involving social preferences include:

- **Self-serving altruism**: When someone else also benefits from misconduct, the other person’s interest is used to justify the action in place of one’s own. The benevolent motive, for example, helping a friend to get a job, overshadows the corrupt act, in this case, nepotism (Barkan, Ayal and Ariely, 2015[15]; Shalvi et al., 2015[57]).

- **Robin Hood logic**: A harm done to a stronger/powerful/richer entity is justified on the basis of a preference for equality.

- **Diffusion of responsibility**: When several people engage in misconduct, the chances of one individual speaking up against it are reduced. Each individual feels less responsibility for the action and does not want to limit the other persons’ freedom of choice or indicate their distrust (Moore and Gino, 2013[58]; Wiltermuth, 2011[59]; Schikora, 2011[60]).

Acting with integrity often means refraining from reciprocity – recruiters are expected to favour strangers over people they know personally based on merit, bidders should make their case with good proposals rather than gifts, and parliamentarians and judges should make impartial choices. Yet, in some settings, the patterns of reciprocity are more reliable than institutions. For example, a significant career step might not be possible without good personal connections or a public service might not be obtainable without a bribe. In such cases, it is rational to build personal connections and perform favours to expect favours in return. The fact that reciprocity is a strong human preference can in many cases explain why corruption occurs or why an integrity policy is not effective.

Strengthening integrity of course does not require eliminating reciprocity. On the contrary, it is in the interest of integrity to strengthen reciprocity, in the form of a more sophisticated **indirect** reciprocity. Instead of “you scratch my back and I’ll scratch yours”, indirect reciprocity follows the logic of “I’ll scratch your back and someone else will scratch mine” (Nowak and Sigmund, 2005[61]). However, such indirect reciprocity has to be supported by shared norms and formal institutions – a system of laws and regulations. Unlike a corrupt interaction, the reciprocity of norms and institutions is not bilateral. It takes place between people who do not necessarily have a personal relationship. Individuals agree to serve the public interest, trusting that the public interest also serves them. Truly moral behaviour does not pay any attention to personal benefits and is solely true to the norm. “Doing the right thing” without expecting praise signals to other people that one is trustworthy and believes in the shared norm (Yoeli et al., 2013[62]). Integrity thus means remaining truthful to the public interest in the face of opportunities to abuse entrusted power for personal gain.

Making choices in the public interest requires trust in the system that it supports: for example, trust in a recruitment system and certifying institutions, trust in a fair bidding process, and a belief in justice and fairness. Eventually, integrity also means trusting that everyone in society benefits from upholding certain norms, values and rules above their direct personal interest. An individual’s contribution to integrity is thus indirectly
rewarded. Strengthening trust in formal institutions and ethical norms and entrenching values in professional identities essentially strengthens indirect reciprocity.

3.3. Norms and identities shape integrity

People orient their actions in accordance with their perception of acceptable behaviour within their social context – whether this is a culture, society or peer group. Famous psychological studies such as Asch’s conformity experiment, the Stanford Prison Experiment or Milgram’s obedience experiment have demonstrated the extent to which people’s choices are impacted by the norms inflicted and the roles prescribed to them (Asch, 1952[63]); (Milgram, 1963[64]); (Zimbardo and Cross, 1971[65]). The Stanford Prison Experiment, for example, transformed students into cruel prison guards or submissive detainees simply through external prescription of these roles by the researchers. In all three experiments the social environment induced participants to engage in cruelty and untruthfulness to an extent that surprised them as much as outside observers.

More recent evidence from research into behavioural ethics confirms the relevance of social norms and identities for moral choices. If one person lies or cheats without facing consequences, this behaviour can spread among friends or colleagues (Gino and Bazerman, 2009[66]); (Robert and Arnab, 2013[67]). In particular, gradual divergences from an ethical norm tend to be more accepted by others, creating a slippery slope towards generalised dishonesty (Gino and Bazerman, 2009[66]). A key factor in creating and strengthening behavioural norms is social identity (Akerlof and Kranton, 2011[68]); (Tyler, 1999[69]), defined as the role individuals assign themselves in a group and the group in society.

The key outcome of such research for integrity policy makers is the importance of continuously strengthening and renewing their efforts to establish integrity as a social norm and to integrate integrity into professional identities. Policies shape norms and identities in many ways. A code of ethics, for example, can be used to emphasise integrity as a common norm of behaviour. Box 3.2 presents an experiment that showed how reference to a moral norm as part of social identity increased honesty.

Box 3.2. Honour and identity

Would students perform better in a test if supplied with the answers to all questions? Researchers gave a group of students a Maths test together with the answers. On average, the students did better than a control group who took the test without knowing the answers. Evidently, the first group cheated and looked at the answers to improve their results. Some test sheets, however, included the following sentence: “I understand that this short survey falls under [the university’s] honour system”. Students who received such a sheet were significantly less likely to cheat on the test. The honour system motivated them to behave more ethically. However, no honour system existed at the time at the schools where the experiment was conducted (MIT and Yale). Regardless, the effect of mentioning an honour system was the same as at Princeton University – where an honour code actually exists. The effect thus cannot be attributed to the honour system in itself, but rather to the reminder of a moral norm. The reference to honour connected with the social identity of the prestigious schools induced greater honesty in the students.

Source: (Mazar, Amir and Ariely, 2008[13]).
Integrity policies can acquire leverage where they acknowledge the power of social norms. From shaping the public debate to changing an organisational culture, social norms can play a key role in enhancing or hindering the effectiveness of integrity policies. Section 3.4 explores possible actions to strengthen integrity norms.

3.4. Enforcing norms

Behavioural research points to the conclusion that informal social norms cannot be strengthened through formal enforcement. Rewarding co-operative behaviour is often more effective than punishment in motivating co-operation (Rand et al., 2009[70]). Reducing an individual’s freedom to devise their own rules might even hinder the development of social norms and the collective resolution of social dilemmas (Ostrom, 2000[71]). As discussed above, an overly strict regime of enforcement and control can crowd out the intrinsic motivation for integrity. While it might be possible to achieve compliance by using fear, more beneficial results can likely be achieved by encouraging public officials to follow their intrinsic motivation towards integrity. Signalling distrust reduces the moral reference point (see section 2.6) as well as trust in the shared norm.

Furthermore, a fear-based enforcement approach could eventually be misused for extortion. Box 3.3 provides an example of how fear of punishment could reinforce corrupt relationships. In some cases, both the abuse of office and the action of receiving a bribe are punishable, leading to situations where the briber may threaten to report the recipient of the bribe if they do not receive an expected favour in return. An open organisational culture, anonymous reporting mechanisms and lenient punishment of self-reporters can help prevent such situations (Buccirossi and Spagnolo, 2006[72]); (Lambsdorff and Nell, 2007[73]). However, institutional oversight, whistleblower laws and prosecution need to be designed specifically to prevent their misuse and increase their effectiveness.

Box 3.3. The risk of extortion

In 1987, an inexperienced employee of the road construction authority in the city of Bochum, Germany, accidentally transmitted the names of competing firms in a public tender. Following this incident, he received an envelope containing 2 000 Deutsche Mark from the private firm that obtained the information. Out of fear of being reported he unduly favoured this firm in various tenders in the years to come. In court, he gave the following statement justifying his behaviour and explaining how he became entrapped in this corrupt relationship: “Suddenly I knew that I had begun to be at his mercy”.


At the same time, where formal norms (e.g. laws and regulations) are concerned, enforcement can serve to build trust. Enforcing sanctions for violation of a formal norm increases confidence that others will not violate the norm, even in cases where a person’s commitment to integrity raises doubts. Sanctions enable a group to trust that its members will uphold an agreed standard of behaviour, even when individual members might not have confidence in the ability of all to adhere to the agreed norms. Where misconduct is common and clearly in breach of an integrity norm, control and enforcement can be used
as tools to restore trust in the norm. Investigations of integrity breaches should therefore be thorough, and any sanctions imposed must be credible and visibly enforced in an impartial and timely manner.

At no point does trust render enforcement redundant. Even in cases where most people are committed to a certain formal norm and would still adhere to it in the absence of punishment, enforcement has a relevant behavioural function – it serves as negative indirect reciprocity. The wish to reward commitment to a norm by others (indirect reciprocity) is matched by the desire to penalise violation of a norm by others (negative indirect reciprocity). If all but one person in a group acted with integrity in the absence of sanctions, the group would still feel that this person should not profit from corruption. Accordingly, the other group members would urge the authorities to impose sanctions on the person violating the norm. The punishment of norm violations also functions as a valuation of the norm and the contributions that everyone makes to ensure it is upheld. Strict and consequent enforcement of integrity laws and regulations is therefore a tribute to integrity and all those committed to it.

Not all norms are enforceable. As argued above, efforts to enforce and control behaviours that cannot be controlled promotes distrust and crowds out the intrinsic motivation for integrity. Introducing the necessary controls separately from motivating and norm-building policies avoids undermining the intrinsic motivation for integrity (Lambsdorff, 2012).  

3.5. The collective action trap

A group of individuals may find itself caught in a low-integrity trap, which can be understood as a collective action problem (Persson, Rothstein and Teorell, 2013). The shared norm of indirect reciprocity is what economists call a public good: if everyone contributes, then everyone benefits. Where everybody cheats, however, those who do not are simply foolish. They might even be punished by others in the group. Cheating then may become accepted and sometimes even expected. It might still be recognised as an integrity breach, but there is no severe (socially enforced) moral stigma. Take, for example, an organisation where nearly everybody over-claims travel expenses. Claiming the true cost of travel might lead a person to feel they were missing out. At the same time, the claimant would risk the disapproval of colleagues, as a low expenses claim might expose the wrongdoing of others.

In cases where systems are failing and loyalty to an integrity norm no longer benefits those who adhere to it, more direct forms of reciprocity gain relevance again. People look after the interests of their peers and clientele and expect assistance only from those whom they help. Sacrificing direct reciprocity for the sake of integrity under these conditions seems unwise. For example, if an individual takes a test, knowing that all the other candidates are secretly helping each other, not cheating will put him or her at a disadvantage when the results are compared. Under these circumstances, the individual might feel less guilty about helping their neighbour. Similarly, an individual will feel less responsible for an integrity breach in a collective action trap. The individuals in such situations may even feel that they had little choice in the matter. A true sense of guilt becomes less likely at this point, while justification mechanisms are probably already present. An integrity breach thus becomes understandable, and people can more or less openly admit their participation. Such collective action traps explain in part why corrupt actors often have no sense of wrongdoing or why a group may be surprised that a
3. INTEGRITY IN THE CONTEXT OF SOCIAL INTERACTIONS

Individual organisations or services can all fall prey to collective action traps. Integrity policy makers therefore need to be aware of this risk, its underlying dynamics and ways to break the vicious circle. Appeals for more integrity can easily fall on deaf ears when they emphasise a norm that nobody sees applied in their daily lives. Policies that build on the shamefulness of integrity breaches (e.g. gift registry, asset declarations, citizen accountability through transparency) might lose their impact when a group commonly accepts misconduct. To avoid spending resources ineffectively, integrity policy makers need to know their target group well. They need to identify which norms are upheld, which behaviours are accepted and which integrity breaches are tolerated. Citizen or staff surveys can be used to gain a better understanding of the dynamics involved. Once a collective action trap has been diagnosed, targeted measures can be applied.

The vicious circle of the collective action trap needs to be disrupted by sanctioning misconduct in a visibly and timely manner and by fostering the collective action of those who are willing to do a good job, for instance by facilitating communication between them and making them visible. In this way, the vicious circle can gradually revert to a virtuous circle. Escaping a collective action trap using resources inside the system can be a tedious process, however. In extreme cases, intervention from outside or a complete change of leadership may be necessary to achieve the required shift. Usually, some type of outside support will be needed to strengthen change actors within the organisation.

3.6. Precepts for integrity in social interactions

Integrity can be defined as an individual’s choice to stand up for the public interest. Yet, ethical choices are not made in isolation – they result from social interaction. Integrity shifts in line with the dynamics of the moral identity of an individual, as well as with beliefs about what others think or do. Integrity policies can therefore benefit from a design approach that takes into account the social environment in which they will be applied. Which social norms are prevalent? Which motives are relevant to the people? Which relationships are impacted? These factors and their influence on integrity should not be underestimated. Table 3.1 highlights the ways in which social interaction dynamics can be taken into account when designing integrity policies.

<table>
<thead>
<tr>
<th>Precepts for strengthening integrity in social interactions in integrity systems</th>
<th>See section</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Provide purposeful transparency</td>
<td>3.1</td>
</tr>
<tr>
<td>b) Strengthen indirect reciprocity</td>
<td>3.1</td>
</tr>
<tr>
<td>c) Strengthen integrity norms and identities</td>
<td>3.2</td>
</tr>
<tr>
<td>d) Investigate and enforce strictly without creating a culture of fear and distrust</td>
<td>3.3</td>
</tr>
<tr>
<td>e) Separate motivating and enforcing functions</td>
<td></td>
</tr>
<tr>
<td>f) Identify and counteract systemic de-stigmatisation of corruption (collective action trap)</td>
<td>3.4</td>
</tr>
</tbody>
</table>

A key factor in social dynamics is responsibility. When the burden of responsibility is spread over many individuals, the resulting diffusion of ethical accountability can pose an integrity risk. When several people jointly engage in misconduct, each of them feels less responsible, because guilt decreases when shared. These dynamics can undermine policies that aim to increase integrity by involving additional actors (e.g. open data...
policies or the four-eyes principle). As such, it is important to avoid the diffusion of responsibility within an integrity system.

Additionally, transparency only serves integrity when those with responsibility have the capacity to act upon it. Efforts to share decisions with additional actors in a transparent manner need to be targeted and undertaken in a purposeful way. This is applicable to transparency efforts at all levels, whether they relate to opening up public data to millions of citizens or sharing a conflict-of-interest situation with a single colleague. If the relevant information is accessible, but the addressee is not aware or cannot act upon it, transparency does not ensure accountability.

Upholding the public interest can be understood as an act of indirect reciprocity, undertaken on the basis that others will do the same and the belief that integrity benefits all. Integrity policies can strengthen such behaviour by creating an institutional setting in which adherence to integrity is expected to be the norm. Setting a high standard of conduct is thus a key objective of integrity policies.

Enforcing this norm, in turn, is a balancing act between building trust in shared values and strictly investigating and sanctioning violations. Both functions gain credibility when they are separated (e.g. executed by different offices).

A group that has collectively diverged from a norm that supports integrity and has destigmatised a behaviour that enables corruption might find itself caught in a collective action trap. In this situation, an individual’s motivation to act with integrity is no longer sufficient to break the vicious cycle and moral appeals will fall on deaf ears. Such situations need to be identified and systematically counteracted within an integrity system.