Mapping moral pluralism in behavioural spillovers : A cross-disciplinary account of the multiple ways in which we engage in moral valuing

Vincent, Michael and Koessler, Ann-Kathrin

Mapping Moral Pluralism in Behavioural Spillovers: A cross-disciplinary account of the multiple ways in which we engage in moral valuing

Michael Vincent

University of Queensland and Australian Catholic University, Australia

Ann-Kathrin Koessler

University of Osnabrück, Germany

Abstract

In this article, we reflect critically on how moral actions are categorised in some recent studies on *moral spillovers*. Based on classic concepts from moral philosophy, we present a framework to categorise moral actions. We argue that with such finer gradation of the moral values, associated behaviour is better understood, and this understanding helps to identify the conditions under which moral licensing takes place. We illustrate our argument with examples from the literature on pro-environmental behaviours. Moral spillovers are frequently observed in this behavioural domain and to understand what causes their occurrence is highly (policy) relevant if we wish to promote sustainable behaviour.

Keywords: moral values, behavioural spillover, spillover effects, moral licensing, moral cleansing**,** behavioural change, sustainable behaviour, pro-environmental behaviour

This version was compiled in October 2018.

Corresponding Author: Michael Vincent, School of Historical and Philosophical Inquiry, University of Queensland, E-mail: m.vincent2@uq.edu.au.

1) INTRODUCTION

Past good deeds can 'liberate' individuals to engage in behaviours which they otherwise would have considered to be inappropriate or immoral (Merritt, Effron, & Monin, 2010) . But there is still great uncertainty about when and how this occurs.

Imagine that you are engaged in two fairly effort-intensive projects – adapting to a diet with substantially less red meat to reduce your environmental impact, and working *pro bono* to help a charity with its new fundraising program. At the end of business one Friday, you feel like stopping at a nearby restaurant for a big steak dinner to unwind from a hard week's work, and reward yourself for a successful week. This is not really in line with your intentions, but it has been a big week and it doesn't hurt to reward oneself now and then, does it? And anyway, after all this good work you will still be confident after your meal that you are a good person who is making a positive contribution to your society. Now consider an alternate scenario: you are in the same position viz. the low meat diet and the craving for a steak, but rather than having worked *pro bono* on a new fundraising program for the charity, you have refrained from stealing from the charity while working for them as an IT contractor. Have you still earned that steak dinner?

The former scenario is probably much more plausible to readers than the latter. Usually, no rewards can be expected for refraining from theft. Is this because there is little value in refraining from theft? Certainly not. In our example, the pay-offs to the NGO can be comparable; we can imagine a missed opportunity for a $500 fundraising campaign, on the one side and a theft of a $500 windfall on the other side. But importantly, there seems to most people to be something particularly bad about theft. So, there can be great value in refraining from theft, but in doing so we do not earn any 'brownie points', whereas other good deeds might indeed have licensed a relaxation of some of our other moral strivings.

The story of costly charitable work resulting in lower commitment to other 'good' behaviour is an example of a 'spillover' effect, in this particular case a 'moral spillover'. Behavioural spillover effects describe how an action in one domain can have an influence on actions in another behavioural or temporal domain. In particular, we are concerned with 'moral spillovers' – that is, spillovers resulting from beliefs or feelings about what is morally right and wrong. The literature has so far argued that contextual factors play a dominant role in whether a behavioural spillover effects arises. In this paper we, however, argue that one's sense of the specifics of the moral values in play is also relevant. We are not the first to attempt to emphasize the role of moral values in this context (Cornelissen, Bashshur, Rode, & Le Menestrel, 2013; Dolan & Galizzi, 2015; Jordan, Mullen, & Murnighan, 2011; Sachdeva, Iliev, & Medin, 2009), but we believe that the discussion could be improved if some ambiguities in the descriptions of moral values are cleared up. Based on our background in moral philosophy and behavioural economics we offer a framework to better structure the discussion of moral spillover effects. In the following, we present ideas that are not new, and indeed many have a very long history of use in moral philosophy, but we feel that the social science literature on behavioural spillovers has not fully taken advantage of them yet. Acts or outcomes are regularly described as (morally) 'good' and 'bad' in the literature. From the point of view of the moral philosopher, such terms obscure important distinctions and without the help of these distinctions a convincing story about why observed behaviours tend to occur cannot, in our opinion, be told. We hope to convince the reader that using such classifications can help to structure the discussion and unravel misunderstandings.

We intend for the principles discussed to be applied to the study of all kinds of pro-social behaviours. However, in this paper we have focused specifically on studies in the domain of pro-environmental behaviour. Aside from being of particular relevance to our research interests, these

studies have two advantages for our project: First, a fairly clear split between self-interested and moral motivations exists in many pro-environmental actions. In at least some situations (such as recycling), acting in an environmentally friendly way has no obvious individual benefits. (In others of course, such as reducing costly electricity use, pro-environmental and self-interested behaviour can clearly coincide, and no appeal to morality is required to explain the behaviour.) So, we can identify more easily some cases which seem to clearly be cases of moral behaviour. Second, there is enormous controversy about the nature of pro-environmental values (Schmidtz, 2015) – a controversy unlike that in other domains of moral behaviour. So, we expect that it is easy for researchers to be open-minded about what exactly the motivations and attitudes are of people engaging in pro-environmental moral behaviour.

What actually counts as a moral value is itself a fraught question, but we do not need to take a stance on the subtleties of the answer here. It will suffice to differentiate reasons of 1) narrow self-interest (typically material, status-seeking, or hedonistic) from 2) other-regarding (perhaps altruistic) preferences and 3) preferences for what is perceived as being just or righteous or intrinsically good, and to locate morality in the vicinity of 2) and 3). While pro-environmental behaviour may be motivated by type 1 reasons (say, if the price per gallon of water is so high that conserving water is the most cost-effective option available), our interest here is in pro-environmental behaviour of type 2 (say, conserving water for the good of others in one's community) and type 3 (conserving water because wastefulness is seen as intrinsically vicious or immoral).

The key lessons which we hope to demonstrate are that a) humans value in many different ways, b) there is an extensive, though fallible, vocabulary with which to discuss these moral values already; and c) this vocabulary should be a helpful complement or alternative to the current practice

of appealing to commonsense stories in explaining the workings of spillover effects. It is important to note that our project is a conceptual contribution. We do not claim to have proven that our classifications of moral judgments is the best available for research on moral psychology – to prove this could require a whole academic sub-discipline worth of empirical work. Rather, we aim to demonstrate that there is conceptual confusion in the existing literature, but also that there are resources within the literature to begin to clear this confusion up.

2) BEHAVIOURAL SPILLOVERS

The term 'behavioural spillovers' describes the fact that a modification in one behavioural domain can influence behaviour in the same or another behavioural domain, either positively or negatively. Campaigns targeting a reduction of residential water have, for example, been found effective in promoting water conservation, but studies testify to the existence of non-intended negative spillovers. Jessoe et al. (2017) found that households which reduced their water consumption subsequently increased their energy consumption.

Spillovers are typically divided into *consistency effects*, in which an initial action begins a trend of similar action, or *balancing effects*, in which a good [bad] *behaviour 1* is followed by a bad [good] *behaviour 2* (Dolan and Galizzi 2015 provide a good review, and we have imitated some of their terminology here; Truelove et al. 2014).[1] Our particular concern here is the study of 'moral spillovers', that is, spillovers in which moral belief or motivation seems to play a role (Nilsson, Bergquist, & Schultz, 2017 give a good taxonomy of spillovers, both moral and non-moral). Within this field, balancing effects are often referred to as cases of moral 'licensing' and moral 'cleansing'.

---

[11] For the purposes of this paper, we focus on behavioural cross effects, i.e. spillovers to other behavioural domains. Behavioural effects within the same domain such as crowding effects are disregarded. For further readings on this matter we would like to direct the reader to e.g. Deci, Koestner and Ryan (1999), Frey and Jegen (2001); and for the environmental domain to Rode, Gómez-Baggethun, and Krause (2015).

Moral licensing has been attributed variously to self-perception (Khan & Dhar, 2006), a sort of moral-accounting 'credits' system (Miller & Effron, 2010), 'resting on one's laurels' (Mazar, Amir, & Ariely, 2008), or a 'single-action bias' (Weber, 2006). Obviously, consistency and balancing effects are incompatible with each other, so identifying the particular conditions under which one arises and the other does not is of great interest. This and similar problems are particularly pressing for those who study moral spillovers, as conflicting results and failures of replication have led to disagreement regarding the strength and prevalence of the phenomenon (Blanken, van de Ven, & Zeelenberg, 2015; Blanken, van de Ven, Zeelenberg, & Meijers, 2014). Sorting out the subtleties of human valuing is, we believe, absolutely fundamental to determining what can now be said with confidence on the issue, and to making any plans for future replications or novel research.

In the case of a purported licensing trend, a 'good' act tends to license a 'bad' act, but in the case of a consistency trend, a good act tends to motivate another good act. Whatever the difference is between the two initial behaviours, it does not seem to be whether they qualify as 'good'. So we must examine the factors linked to the acts (like framing, or effort required, or social expectations), or the degrees of 'goodness' of the acts, to determine what the relevant difference is. But, in this quest we should be aware that terminology like 'good' and 'bad', or 'moral' and 'immoral', is a highly simplified categorisation which might not provide the nuance we need to detect the determining differences.

3) WAYS OF VALUING

Humans always have and probably always will have different sorts of moral values. Consider the following analogies: Researchers are doing a study in physiology where subjects are ranked by size, but for some subjects 'size' means 'weight', for other subjects it means 'height', for others 'shoe size', and the differences between the sorts of measurements are not acknowledged. So,

someone could be ranked 60 – but it is not clear whether that means 60 inches tall or 60 kilograms, etc. Similarly, it is not clear that the person with the biggest number is the biggest person – both because someone who ranked 62 for height could still be *shorter* than someone who ranked 61 for weight, and because 'bigness' is ambiguous here – we are not sure how to commensurate the different measures into a single measure of who is 'big'. Similarly, a study could set out to rank countries in terms of their 'macroeconomic strength', but then measure the relative GDP size of some (America is ranked 1st), the relative growth rate of others (Iraq is ranked 1st), the relative political stability of others – yet never try to explain how each relates to 'macroeconomic strength' or how the different values are to be compared or commensurated. Obviously, the above research would be deeply problematic. But 'morality', we will try to show, is a notion at least as troublesome as the uses of 'bigness' and 'macroeconomic strength' here.

a) How to Value: a Scalar model and a Categories model

'Good' versus 'bad' is a distinction which gives us some means for conceptualising spillover effects. But there are certainly other distinctions at play in ordinary moral judgments. First, we should acknowledge that some actions are taken to be not so much good, as 'acceptable' or 'permitted'. Such actions contrast with those which are both bad and 'unacceptable' or 'impermissible'. To say that an action is acceptable is not to suggest that it is a good one, but simply to say that it does not count as a bad one. Brian McElwee, in a recent paper, asks us to consider:

> "two ways of morally classifying acts: (a) a division into *deontic categories* of morally obligatory, morally forbidden, and morally optional (neither morally obligatory nor morally forbidden), and (b) an arrangement within an *evaluative scale*, ranging from morally best to morally worst." (2017, p.505)

It is fairly easy to visualise the latter scale, it might look like the two sided arrow displayed in Figure 1. We call this a 'Scalar model'.
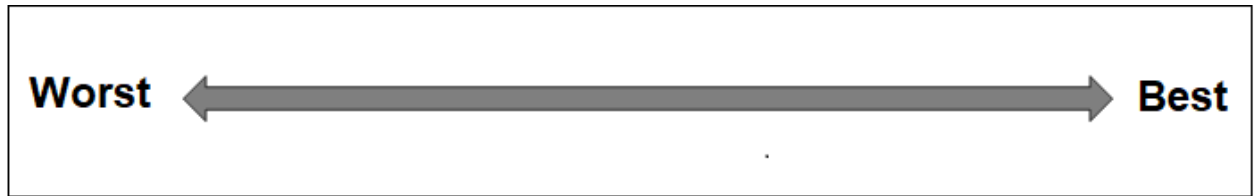
Figure 1: The Evaluative Scale

The earlier categories, of obligatory, forbidden, and optional, can be represented as in Figure 2.1. or in Figure 2.2. We will call these 'Categories models'.
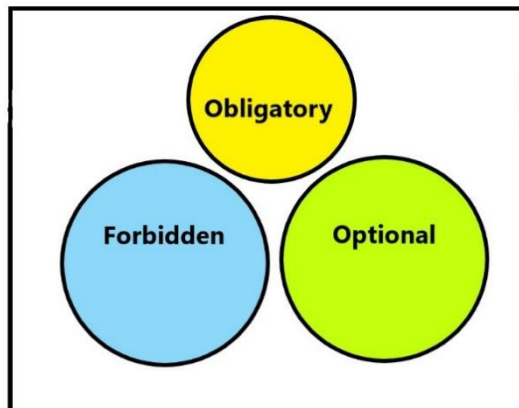

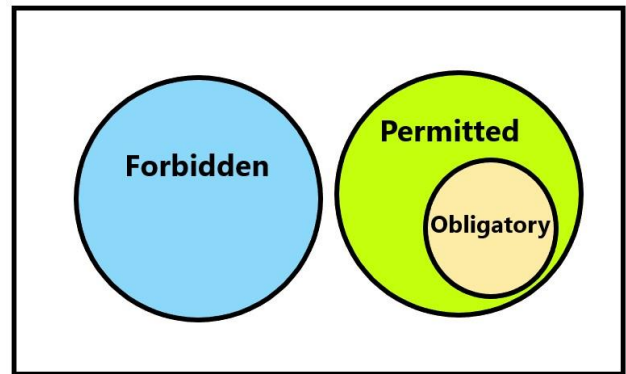Figure 2.1: The Deontic Categories 1

Figure 2.2: The Deontic Categories 2

Where do the deontic categories (Figure 2.1 and 2.2) fit onto the evaluative scale (Figure 1)? It seems intuitive that the 'forbidden' acts will be towards the 'worst' end of the scale. But where do the 'obligatory' acts fit? Should they be represented as around the middle of the spectrum then? This area is also occupied by many acts which are also not governed by obligations – like deciding which colour of t-shirt to wear today, or whether to eat an apple or an orange. Many of the 'optional' acts seem to be neither particularly good nor particularly bad. Working within common sense morality, we find that there is no neat way to graft the two ways of classifying acts onto one spectrum. It does not seem that one model is reducible to the other.

b) The Substance of these Models: Maximalist and Deontic Values, and Supererogatory Acts

One common way (it is not the only way that people employ) of assessing where an act falls on a scalar model of ethics is to evaluate its consequences – either the actual consequences, or those that could reasonably have been predicted. This relies on a <u>'maximalist'</u> account of valuing. 'Maximalist' valuing is the sort of valuing which fits most neatly into an economistic account of human nature – for example, the figure *Homo economicus* is a maximiser, and so is the utilitarian. In each case, ideally, the option which maximises good outcomes (of whatever sort) is always chosen. While there is no uncontroversial account of what humans do or should aim to maximise, archetypal contenders include income, consumption, happiness or preference satisfaction. So there are different sorts of maximalist values, and they need not be compatible with each other. 'Maximalist' is broad enough to encompass both selfish valuing (I want to maximise my income) and altruistic valuing (I want to maximise global welfare).[23]

Maximisation can serve as a normative goal, and a person who fails to maximise good outcomes for themselves might be judged lazy or irresponsible.[4] A person who could easily maximise good consequences for others but refrains from doing so may be judged even more harshly than this. Many goods that we pursue seem evidently maximalist. If I want money, then presumably the more money I can get the better it is for me. If I volunteer to help clean up litter on

---

[2] But the different sorts of ideals are only conceptually related, and may not even be compatible in practice. If person A wants to maximise global happiness and person B wants to maximise global suffering, then clearly these are incompatible maximalist goals!

[3] The concept is also broad enough to encompass non-linear increases in value, such as cases of diminishing marginal returns.

[4] Note that both of these are value-laden terms.

the beach, I presumably want to see the beach *maximally* clean. I will of course have to stop at some point, but any reduction in litter will translate fairly directly into greater success.

However, this is not the only way of assessing value, particularly with regard to moral action. Consider the two examples above, of money and litter. I may feel that the more money I can acquire, the better the outcome is for me. But I may also feel, at the same time, that it is particularly important that I get the money which I *deserve*: the money that I have worked for, or which I am entitled to for some other reason. Similarly, while I may regret having to leave the beach with some litter on it, I am likely to feel particularly bad about the situation if I am the one who put the litter there in the first place: if it is *my* litter which I am *responsible for*. Many judgments like these are captured by the label of 'deontic' valuing.

The term 'deontic' suggests a sense of strict obligation, of moral *duty*. A duty implies a strong reason to engage in or refrain from certain acts. Sometimes this is presented as a limit on one's set of options: 'You shall not do x' might suggest that a certain course of action should be unthinkable, regardless of the potential consequences. More plausibly, it could be held that a duty should only make us extremely averse to certain courses of action (actions at odds with our duties), such that we will do our best to avoid ever having to engage in them. It is common to suppose that the most important duties should be 'negative duties' or duties to refrain from certain forms of behaviour ('thou-shalt-nots'); this is usually held to be because it is harder to demand consistent active behaviour from people – for example, while we can plausibly demand that a person live their entire life without killing another, it is harder to demand that they intervene at every opportunity to prevent another from dying. So while it is commonplace to talk of a duty to take action, compliance with such 'positive' or 'active' duties tends to be harder to monitor and enforce, and the sense of 'duty' tends to be weaker.

'Deontology' has come to denote the discussion of a range of values which do not always provide strict obligations. While one feature of it is talk of duty and fundamental rights (and other notions which might be identified as 'sacred values' by psychologists (see for example Tetlock, Kristel, Elson, Green, & Lerner, 2000), the term is also invoked for the discussion of rule-based moralities in general. As we saw above, not all moralising is concerned with maximising the good; rather, some is concerned with setting out the bounds of the acceptable. And it is here that rules (whether laws, or conventions, or social norms, or the edicts of moral leaders like Moses or Immanuel Kant) are helpful, because they reliably differentiate the acceptable from the unacceptable in a wide range of different circumstances. So in modern parlance, despite its etymology, 'deontology' has come to be seen as just as much the study of rules and rights as the study of moral absolutes.

To illustrate the differences here, let us briefly look at the example of donating to charity, an act which is frequently referred to as paradigmatically moral. Donating can be seen as a maximalist act – other things equal, it is a good thing to donate, and the more donated the better. So, donating 4% of one's income is roughly twice as good as donating 2%.[5] Contrast this with the historical practice of paying a religious tithe (at least at times when this was not strictly enforced), or the contemporary practice of taking the `Giving What We Can´ pledge.[6] In both cases there is an expectation that 10% of one's income will be given away, and there is a strong expectation that the 10% mark will be reached. While we can still talk of a 10% donation being better than a 5% donation, there is also a very significant distinction between giving 10% and giving 9%. And this

---

[5] Donations from individuals in the US hover at around 2 percentage of disposable income. https://www.philanthropy.com/article/The-Stubborn-2-Giving-Rate/154691, retrieved 24.03.2018, 9:36 pm
[6] https://www.givingwhatwecan.org/pledge/, retrieved 19.02.2018, 5:10 pm

latter distinction is likely to be of a deontic sort, between the acceptable and the unacceptable, between success and failure in acting on one's obligation.

With an understanding of the contrast between maximalist and deontic thinking, we can now introduce one final notion, that of 'supererogatory acts'. Whereas duties are thought to set a minimum standard that we are expected to meet, and maximisable values set a target which it is sensible or otherwise good for us to aim at, there is also a commonsense notion that some acts are not just morally acceptable but morally excellent. Great generosity might be a relatively uncontroversial example. These are known as supererogatory acts. To do one's duty is usually not considered to be a supererogatory act. But duties help to set a standard for supererogatory acts – they are acts that go 'well beyond the call of duty'. Conversely, a purely utilitarian, maximalist, account of morality cannot make much sense of the notion of 'supererogation' – achieving the best possible outcome is the whole point, and there is nothing better to do than that. We should distinguish two aspects of the notion of supererogation, both of which need to be recognised as features of commonsense morality. We have already stressed the excellence of some acts, but another feature of them is their *unnecessariness*. People are not typically *blamed* for failing to perform supererogatory acts. The paradigmatic case of a supererogatory act is one which is both excellent and is not required by duty.

Supererogation in particular is a notion which seems to be wholly absent from the spillovers literature. To return to the example of donating 10% of your income, for most of us this will be an act of generosity which we think is admirable but which we have no intention of imitating. This is a greater act than, for example, spending a few dollars more on 'green' product or donating a few dollars to charity. But still, for a person with a low disposable income, even a small amount of money could have been of great use, and giving it up is quite a sacrifice. Hence, not only do

individuals assess opportunities for prosocial or pro-environmental acts differently, they will likely import their own value assumptions into their interpretations of others' decisions. And this simple implicit failure to agree on what sorts of values are in play can, we believe, distort the design and analysis of research studies: leading us to exaggerate the similarities between studies, to conflate important variables, or to misunderstand participants' self-reporting of their experiences and beliefs.

c) Domains of Action

There is a final issue of value pluralism which it is worth making explicit, as its importance is not always recognised. This is the question of how to go about promoting one's values. People who value the environment might refrain from damaging it themselves, they might encourage others to refrain, or they might actively engage in environmental protection. These are already three ways in which an individual can live up to upon her environmental values. But environmentalism might also affect one's diet, career choice, spending (whether consumable or philanthropic), voting and other civic behaviour, travel, even friendship networks or aesthetic preferences. We will call this consideration the 'domains of action'; that is, the plurality of ways in which an individual acts on or expresses her values.

4) MORAL CONCEPTS IN THE CURRENT SPILLOVERS LITERATURE

Some of the above concepts appear in a few existing studies of behavioural spillovers. Cornelissen et al. (2013), for example, link a 'rule-based mind-set' to consistency spillovers (both good and bad). "*Moral rules… do not naturally lend themselves to… trade-offs, because 'a rule is a rule'.*" (Dolan and Galizzi 2015 summarising Cornelissen et. al.'s findings, 2015, p.9). This 'rule-based mind-set' is what we call deontic valuing (with elements of 'aretaic valuing', for details please see Appendix I). Mullen and Monin provide an excellent summary of the deontic distinction

between unacceptable and acceptable when they write that "*perfect duties are black-and-white litmus tests, but they are asymmetrical: Someone violating a perfect duty is immoral, whereas someone respecting a perfect duty does not get much moral credit.*" (Mullen & Monin, 2016). Similarly, Monin and Miller (2016) have recently suggested a distinction between 'moral opportunities' and 'moral tests', the former being something like an opportunity to perform a supererogatory or maximalist act, the latter being a case of one's being under an obligation to perform a duty.[7] And finally, Truelove et al. (2014) talk of different 'decision modes', including 'calculation-based' (maximalist), 'role-and-rule- based' (deontic, with an aretaic element).

In other research papers which focus on the general behavioural spillovers the distinctions in moral values we listed in Section 2 play no part. We should stress that this is not a problem for the work – values are controversial and can be difficult to pinpoint. But caution is needed when the results are subsequently interpreted and compared with other studies. Carrico, Raimi, Truelove, and Eby (Carrico, Raimi, Truelove, & Eby, 2017) for example discuss spillovers using a distinction between 'morally good' and 'morally dubious', as well as using the term 'pro-social behaviour' which is seemingly interchangeable with 'morally good', and the term 'pro-environmental behaviour' which is seemingly a subset of these. This language use overlooks the middle ground of 'moral acceptability' between the good and the dubious, and fails to distinguish moral excellence from more mundane forms of moral good. Further, the assumption that pro-environmental behaviour is always prosocial is questionable.[8] Similarly, in other cases, the focus is on 'environmentally-friendly behaviour', with the link between this and morally correct behaviour being vague but unmistakably present. Lacasse (2017), for example, is interested in the link between

---

[7] That is, a 'perfect duty', to use Immanuel Kant's term, a term which also appears in the previous quote from Mullen and Monin (2016).

[8] See Routley (Routley, 1973) for a classic attempt to rebut this.

'Pro-Environmental Behaviours' and political attitudes, internal motivations, personal norms, and environmental identity, all of which seem intimately linked to subjects' sense of what is 'moral'. The paper provides an interesting examination of what we call 'domains of action', though in all other respects the different 'ways of valuing' we have described here are not discussed.

5) ROOM FOR IMPROVEMENT IN THE CURRENT USAGE

For other research studies, which focus more specifically on the moral dimension in behavioural spillovers, we believe that applying the value scheme suggested by our framework can shed light on research gaps which currently exist. We can illustrate this with some cases of cross-domain spillovers, in which an initial behaviour affects the likelihood of an act in a very different 'domain', or area of moral concern.[9] Our examples here are primarily drawn from those who have followed-up on the work of Sachdeva, Iliev, and Medin (2009). Generally, our concerns are that:

1) Often, it is not clear how distinctions between moral, amoral, and immoral behaviour are drawn, and so whether the researchers participating in the dialogue themselves agree on the extension of these categories. (5.a)

2) Researchers may interpret a subject's 'bad behaviour' as a (temporary, or induced) lack of interest in acting morally, when the behaviour in fact indicates different beliefs about what 'good behaviour' is. In other words, disagreement regarding which 'domain of action' is appropriate – what sort of action must be taken to do the right thing – can be misconstrued as apathy or immorality. (5.b)

---

[9] Environmentalism perhaps being one moral 'domain', or a catch-all term for a number of domains like sustainability, litter reduction, species preservation, etc.

3) Researchers may talk of the severity of wrongdoing, while subjects may not agree with their severity assessments. So what appears to a researcher to be inconsistent behaviour ('moral balancing') may be consistent in the eyes of the subject. (5.c)

To illustrate these concerns we will work through exemplary passages from this literature.

a) Delineating Im/morality

Consider this passage by Sachdeva, Iliev, and Medin, in which we are presented with distinctions between moral, amoral, and immoral, but not given enough guidance regarding where the authors believe that the distinctions are to be found:

> The present research also does not address the question of whether people simply refrain from engaging in moral behavior in cases of moral licensing or are actually more liable to behave immorally. This is an important question to consider because the answer may indicate the strength of the licensing effect in motivating moral behavior. One possibility is that the licensing effect arises because of an accrual of ''moral currency,'' which allows people to more or less passively engage in more secular sorts of activities until that currency has been spent. However, the more insidious possibility is that moral licensing lowers the bar of what is considered to be an amoral activity so that people are more likely to do immoral things that yield various types of secular benefits. These experiments show the first possibility to be true, because people refrained from doing something outwardly good (…), but our experiments do not show if moral licensing allows people to do something that is prima facie bad. A way to test this possibility might be to use cheating as a dependent variable and examine whether morally licensed individuals are more likely than others to cheat on a task that entails some sort of instrumental benefit. (Sachdeva et al., 2009)

This is an important passage, as it helped to trigger a series of studies attempting to observe 'licensed' *immorality*. Following from Sachdeva, Iliev, and Medin's paper, Jordan, Mullen, and Murnighan (2011, p.702) seek to find "*whether moral behavior can license actual immoral behavior (rather than simply reducing the prevalence of moral behavior).*" Jordan et al. (Jordan et al., 2011) are wary of a spectrum model of im/morality, writing that "*moral and immoral behaviors are not opposite ends of a single scale but rather two, distinct dimensions*" (p.703). Note that all of

16

these quoted sections seem to presuppose a 'categories' model of ethics, as in our Fig. 2.1 and 2.2. The examples given of immoral behaviour are acts like cheating and theft, in which the source of the 'immorality' does not seem to be a failure to maximise good outcomes, but rather the breach of an important rule of conduct. So, 'immoral' acts are identified via deontic reasoning. Whereas 'moral' acts included "*donating to charity, donating blood, volunteering*" and "*help[ing] other people*" (Jordan et al. 2011, p.705) which we submit would typically be taken to be voluntary maximalist acts – on a categories model these are merely 'permitted' or 'optional', but they seem to fall somewhere on the 'good' side of a scalar model.

So when interpreting Sachdeva et al. and Jordan et al., we find a sort of 'middle ground' between rule-breaking (labelled immoral) and utilitarian maximalisation (labelled moral), which the writers label 'amoral' or 'nonmoral'.[10] 'Amoral/nonmoral' here could mean anything from behaviour which is unusually selfish but does not actually 'break the rules' to perfectly morally neutral behaviour (like choosing one flavour of icecream rather than another). And this ambiguity is not helpful if we are seeking to generalise about how spillover effects operate. It looks in both of the above cases as if maximalist and deontic thinking is being conflated, and we may be able to get some greater clarity if a distinction is introduced.[11]

---

[10] At another point Jordan et al. differentiate 'absence of moral behavior' from 'explicit immorality' (p.706), which in the context seemingly also identifies their amoral/immoral distinction.

[11] Mazar and Zhong (2010) have also taken up Sachdeva, Iliev, and Medin's suggestion to "*use cheating as a dependent variable and examine whether morally licensed individuals are more likely than others to cheat*" (Sachdeva, Iliev, and Medin 2009, p.527). They found (as did Jordan et al., 2011) that lying and cheating can be among the spillovers of some 'good' behaviours. In this study, Mazar and Zhong take some steps to assess their subjects' own interpretation of the link between ethics and pro-environmental behaviour (in Experiment I) and limit their own discussion of their findings to describing cheating as 'unethical' and 'socially undesirable'.

b) Domains of Action and Domains of Interest

We have seen that as well as dividing behaviour into domains by topic of moral concern (e.g. environmentalism, pro-sociality, personal integrity) we can recognise different interpretations of moral imperatives (maximalist, deontic, etc.). We should also recognise a distinction, between moral concern for actions, or for personal character, or for institutions, or for states of the world. It can be perfectly consistent to think that a certain act is wrong, even that the act is deontically prohibited, but also to believe that we should not interfere with or punish the bad act. A classic example might be the betrayal of a promise to one's spouse – it is the wrong thing to do, but that does not mean that outsiders have any business getting involved and making sure that the promise is kept. Furthermore, as Cristina Bicchieri (2006) has demonstrated, people can frequently disapprove of a social norm while continuing to participate in *and enforce* it – so judgment does not translate directly into action, and certainly not into collective action.

Similarly, we may disapprove of a state of the world or a pattern of behaviour, such as our disapproval of water or energy wastage, while also disapproving of some of the measures taken to fix it. Noblet and McCoy (Noblet & McCoy, 2017) overlook this in their study on energy saving measures. They write:

> [W]e focused on the potential for domain-specific licensing and hypothesized that individuals who participated in prior sustainable behavior in the energy realm would be less likely to support a future sustainable choice in the energy realm because they feel licensed by their prior behavior. (p.16, *emphasis in original*)

In this case, behaviour 1 is (remembering) the household's energy-saving behaviour, while behaviour 2 is to choose whether to support a government policy which would increase energy bills but would also increase renewable energy investment and/or fund energy efficiency programs. But voting for or advocating a certain policy is not part of the same decision mode as acting on related

concerns in private. While these are two decisions within the 'energy' domain of interest, they are not within the same 'action domain'– these are two very different sorts of moral decisions about energy use.

Furthermore, the authors list as a limitation of their study the fact that *"[w]ith our data we are unable to determine if participants followed through on their choice to support the Energy Policy Scenario by actually donating funds to renewable energy or energy efficiency."* (2017, p.17-8). Again the same reasoning applies. Just as being energy-conscious at home is not the same as supporting an efficiency-levy, supporting an efficiency-levy is not the same as donating money to an efficiency-charity.[12] Imagine an anarchist environmentalist (or a less extreme form of 'small-government environmentalist'). We ask them, "Will you buy some energy saving lightbulbs?" They say "yes", because they are an environmentalist. We then ask, "Will you support a government program which increases taxes to pay for more energy efficiency?" They say "no", because they are an anarchist. Their values here are not inconsistent, but the domain of action in which environmentalism is exercised is pivotal.

Noblet and McCoy's paper is concerned with 'domain specific licensing' in the 'energy domain' – but while examining this the study crosses back and forward between the 'domains' of household management, democratic citizenship, and charitable giving. Hence, what the authors observe does not need to be moral licensing effect, but rather shows that people seemingly behave inconsistently across different domains of action, as they are acting on multiple values.

---

[12] For an extended illustration of the difference between approving of a societal policy and acting on it unilaterally, and of approving of high taxes but not engaging in equivalent philanthropy, see Cohen (2000).

c) Subjects' Values versus 'Objective' Values

The final concern, which is tightly linked to the two previous points, is the fact that the value judgments of the researchers are often presented as correct and those of the spillover-prone subjects mistaken or biased. This is only one of many ways in which researchers' values and moral views might shape their research design and analysis, but it is a particularly important one to be aware of. Recall the previous quote in which Sachdeva, Iliev, & Medin (2009) suggested the possibility that "*moral licensing lowers the bar of what is considered to be an amoral activity*" and so "*allows people to do something that is prima facie bad.*" Now, it seems that '*prima facie*' means from the perspective of the authors and their readers – it is clearly not from the perspective of the subjects, given the notion here of 'lowering the bar of consideration'.

Commonly cheating and theft are judged to be particularly bad, probably because there are stronger norms in our societies governing theft or cheating than norms regarding giving to strangers or donating to charity. While refraining from these later actions is seen as 'disappointing', only theft is clearly *impermissible* (a *deontic* notion). It seems this is also the view of many (though not all) of the aforementioned researchers,[13] given their tendency to single out the cheating as particularly 'bad' or 'immoral'. However, it is an open question whether subjects would judge lying about the earned income in an experiment and cheating on a research fund as more immoral than not sharing given money with a stranger, or not donating to charity. Another example of the diverging interpretation of moral actions and its pitfalls are the aforementioned giving examples. While giving away a part of one's earned experimental payoffs, either to a charity or to a fellow-subject, might be perceived by the researcher as a sensible way for the subjects to maximise good outcomes, a

---

[13] See the examples given above in 5.a. Jordan et al. 2011 also use cheating as a paradigm example of an 'immoral' act.

low-income subject (like an undergraduate student) might understand the act as a noble ('supererogatory') gesture which warrants some serious moral praise (or licenses some serious misconduct).

6) DISCUSSION AND CONCLUSION

Overall, the recognition that there are multiple ways in which humans engage in moral valuing suggests precautions for future research on behavioural spillovers.
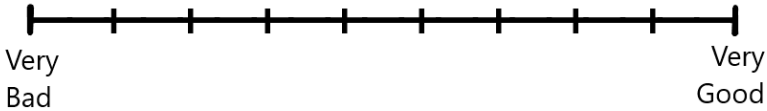
Dolan & Galizzi (2015) have stressed the need for a way to quantify spillover effects, and have highlighted the lack of progress on this so far. We hope that we have demonstrated one important consideration of such a project: that measuring morality cannot (always) be done with a simple spectrum from very bad to very good. Rather, we should distinguish between the harmfulness and impermissibility, between actions and motives, and between acts, recollections, policy preferences, character traits, etc. We are aware that this is a very complex task, particularly since many factors are hard to objectively quantify, or at least to quantify in an intuitive or widely acceptable way. But we believe that also moral decision making is a highly complex process and future work considering this complexity can, in our opinion, help clarifying what leads to problematic behavioural spillovers.

To make a start, it should be helpful, when interpreting the work of others, to try to identify whether the work presupposes a scalar model of ethics (Figure 1) or a categories model (Figure 2.1 and 2.2.), or both, or some other schema. This can help to identify whether apparently similar research projects indeed arrive at comparable conclusions. Furthermore, when doing original research, researchers could ask subjects how they feel about certain sorts of im/moral behaviour with reference to the aforementioned models. For example, when discussing a certain act (such as

cheating in a test or donating a certain sum to charity), subjects could be asked to fill out the following task sheet (note that here we have here added supererogatory acts to the categories model):

Where on the following diagrams do you think that the act is located?

Figure 1:

Very Bad   Very Good

Figure 2:

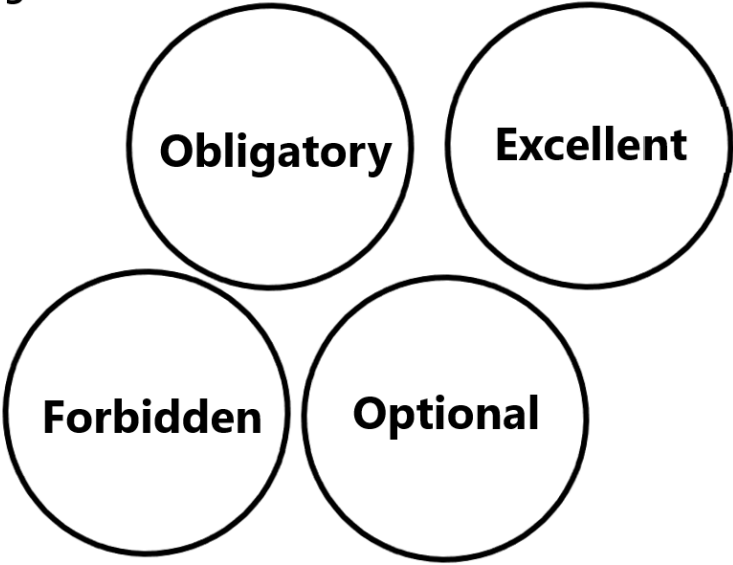**Obligatory**   **Excellent**

Forbidden   **Optional**

Figure 3: Task Sheet for Subjects[14]

---

[14] Note that we have been careful not to label the centre of the scalar model: we suspect that many different judgments will be placed around there, and we do not want to leadingly suggest to subjects that the actions they judge to be 'amoral', 'acceptable', 'pragmatic', or 'cost/benefit-neutral' belong here or on any particular side of the centre.

We hypothesise that some topics will be more fruitfully discussed and classified with reference to one model than to the other. These will include our judgment, seemingly implicit in the work of others in the literature, that cheating and lying are best seen as 'forbidden' than as siting somewhere on a scale of badness, and that while donating to a good cause is widely recognised as 'good' in the scalar model's sense, discussion of its being obligatory, optional, or supererogatory will be more problematic (barring exceptional cases). We will be able to see this if subjects can independently agree on how to classify an act on only one model, and the agreement is on a matter of substance. For example, everyone might agree that a $1,000 theft goes in the 'forbidden' category, but they might not be able to agree on whether it is better or worse than marital infidelity, lying under oath, or illegal polluting; so, they will not be able to agree on where these go on the 'scalar model'. Conversely, while subjects may disagree about whether a donation of a given sum to a good charity is optional, excellent, or even obligatory, they should be able to agree that a larger donation is better than a smaller one (assuming that the task is well-framed by the researchers, and subjects do accept that the charity is indeed a good one). In this later case, we suspect, the scalar model will be more helpful in substantive assessments of just how invested in the 'goodness' of the good act the subject is, and so what degree of balancing or consistency might be occurring. If areas such as these of intersubjective agreement and disagreement can be found, then we will also have identified an important component needed in future work like that of Cornelissen et al. (2013), which seeks to encourage a rule-based (deontic) or consequentialist (maximalist) mindset in subjects: it will be easier to do so if the subjects are considering acts which are widely regarded as fitting into a certain category or are easily placed on a good-to-bad scale.

Another point that we hope to have brought to the reader's attention is this: the assessment of spillover effects can only lead to generalisable conclusions if we researchers are more careful with

the language that we use, so that we can describe exactly what we observe subjects are doing. Interpreting these observations then demands that the possibility of different perspectives must be taken into account. Part of this is explicitly differentiating between the values of researchers and the values of subjects. While it is important to observe that a subject's evaluations are inconsistent, we should be sure that this is real inconsistency and not apparent inconsistency due to researchers' values intruding.

Our theme has been the 'dimensions of difference' within morality. We believe that huge potential for cross-disciplinary work exists here. As Jordan, Mullen, and Murnighan suggest, the results of moral spillover experiments *"have interesting implications for how people cognitively categorize fairly disparate activities (e.g., helping and cheating) within a single moral rubric."* (2011, p.710). And, we hope to have provided with our framework how to categorise moral actions a helpful starting point for future collaboration which seek to understand pro-environmental behaviour and the determinants of potential moral spillovers.

7) BIBLIOGRAPHY

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540–558. https://doi.org/10.1177/0146167215572134

Blanken, I., van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (2014). Three attempts to replicate the moral licensing effect. *Social Psychology*, *45*(3), 232–238. https://doi.org/10.1027/1864-9335/a000189

Carrico, A. R., Raimi, K. T., Truelove, H. B., & Eby, B. (2017). Putting Your Money Where Your Mouth Is: An Experimental Test of Pro-Environmental Spillover From Reducing Meat Consumption to Monetary Donations. *Environment and Behavior*. https://doi.org/10.1177/0013916517713067

Cohen, G. A. (2000). If you're an egalitarian, how come you're so rich. *The Journal of Ethics*, *4*(1–2), 1–26.

Cornelissen, G., Bashshur, M. R., Rode, J., & Le Menestrel, M. (2013). Rules or consequences? The role of ethical mind-sets in moral dynamics. *Psychological Science*, *24*(4), 482–488. https://doi.org/10.1177/0956797612457376

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668.

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, *47*, 1–16. https://doi.org/10.1016/j.joep.2014.12.003

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589–611.

Jessoe, K., Lade, G. E., Loge, F., & Spang, E. (2017). *Spillovers from Behavioral Interventions: Experimental Evidence from Water and Energy Use*. *Economist*.

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, *37*(5), 701–713. https://doi.org/10.1177/0146167211400208

Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research*, *43*(2), 259–266. https://doi.org/10.1509/jmkr.43.2.259

Lacasse, K. (2017). Can't Hurt, Might Help: Examining the Spillover Effects From Purposefully Adopting a New Pro-Environmental Behavior. *Environment and Behavior*, *0013916517*. https://doi.org/10.1177/0013916517748164

MacIntyre, A. (1981). The nature of the virtues. *Hastings Center Report*, *11*(2), 27–34.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644.

Mazar, N., & Zhong, C.-B. (2010). Do green products make us better people? *Psychological Science*, *21*(4), 494–498.

McElwee, B. (2017). Supererogation across normative domains. *Australasian Journal of*

Philosophy, *95*(3), 505–516.

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees

Us to Be Bad. *Social and Personality Psychology Compass*, *4*(5), 344–357.

Miller, D. T., & Effron, D. A. (2010). Psychological license: When it is needed and how it

functions. *Advances in Experimental Social Psychology*, *43*, 115–155.

Monin, B., & Miller, D. T. (2016). Moral Opportunities Versus Moral Tests. In *The Social

Psychology of Morality* (pp. 40–55). Routledge.

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior.

Annual Review of Psychology, *67*, 363–385. https://doi.org/10.1146/annurev-psych-010213-

115120

Nilsson, A., Bergquist, M., & Schultz, W. P. (2017). Spillover effects in environmental behaviors,

across time and context: a review and research agenda. *Environmental Education Research*,

23(4), 573–589. https://doi.org/10.1080/13504622.2016.1250148

Noblet, C. L., & McCoy, S. K. (2017). Does One Good Turn Deserve Another? Evidence of

Domain-Specific Licensing in Energy Behavior. *Environment and Behavior*,

0013916517718022.

Rode, J., Gómez-Baggethun, E., & Krause, T. (2015). Motivation crowding by economic

incentives in conservation policy: A review of the empirical evidence. *Ecological Economics*,

109, 270–282. https://doi.org/10.1016/j.ecolecon.2014.11.019

Routley, R. (1973). Is there a need for a new, an environmental ethic. In *Proceedings of the XVth

World Congress of Philosophy* (Vol. 1, pp. 205–210).

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, *20*(4), 523–528.

Schmidtz, D. (2015). Value in Nature. In I. Olson & H. och Jonas (Eds.), *The Oxford Handbook of Value Theory* (pp. 381–399). New York: Oxford University Press.

Tetlock, P. E., Kristel, O. V, Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870.

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, *29*, 127–138. https://doi.org/10.1016/j.gloenvcha.2014.09.004

Weber, E. U. (2006). Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change*, *77*(1–2), 103–120.

Williams, B. (2006). *Ethics and the Limits of Philosophy*. Routledge.

8) APPENDIX I: VIRTUES AND RELATED CONCEPTS

While debate about the importance of rules and outcomes is a mainstay of contemporary moral philosophy, these notions do not exhaust the range of concepts available to us. There are many others, some of which are associated with 'virtue ethics': the third of the major approaches to moral evaluation typically taught to philosophy undergraduates today. Although it proved difficult to incorporate these ideas into the body of our paper, we want to present them here for readers to consider. We suspect that this sort of valuing is going to be particularly resistant to quantification, for reasons to be explained, and hence we have not offered a model of it akin to the 'scalar' and 'categories' models.

We might feel that we can judge someone's 'moral character', and that such a judgment is *not* reducible to an assessment of their past actions. The term 'aretaic values' is derived from an Ancient Greek term often translated as 'excellence', or more moralistically, as 'virtue'. Aretaic judgments are judgments of character – of traits which are supposed to attach to people over the long term. It is commonplace, though not ubiquitous, to associate virtuous character with sociability, so a virtuous person is sympathetic, respectful, civic-minded, or whatever.[15] It is also common to associate these judgments with deontic or maximalist expectations ('they are a law-abiding person'; 'they are an effective mentor'), though it would be a mistake to see an aretaic judgment as simply description of past or even predicted behaviour. A character trait may exist, presumably, without it is being publicly displayed, and it is not displaying the trait which principally makes someone

---

[15] Some exceptions to this are the 'militant' virtues of someone like Homer's Achilles, and the 'monkish' virtues of a religious ascetic. See MacIntyre (MacIntyre, 1981) for discussion of these and a good general introduction to 'virtue ethics'.

virtuous, but possessing it. The clearest case of this is when someone has a certain character trait (say, generosity) but no opportunity to display it (because they have nothing to offer to others).

Some of the behavioural spillovers literature seems to recognise this distinction between one's perceived character and one's actual actions. For example, Miller and Effron (Miller & Effron, 2010) stress the importance of one's self-identity (one's perceived character) in predicting moral licensing. Some researchers like Sachdeva, Iliev, and Medin (2009) seem to have been able to manipulate the self-image of subjects, such that they can admire someone (or themself) as a 'generous person' (or similar) without actually *doing* anything generous at all. As the subjects are not actually performing a relevant action, but rather just by reflecting on virtues of character, this looks like 'aretaic' valuing (perhaps of a seriously wrong-headed sort, in this particular case).

Furthermore, some would say that being a good person is not so much about knowing the rules and having the motivation to follow them, as about being able to wisely and sympathetically balance one's relationships with others. A generation ago, the alleged neglect of this topic was central to Carol Gilligan's (1982) complaint against Lawrence Kohlberg's account of psychological moral development. Taking relationships of care seriously should have consequences for the use of toy dilemmas like trolley problems and dictator games in research. If it is correct that the concerns of 'care ethics' (a type of virtue ethics) are an important part of our moral practice, then moral dilemmas in a lab may be a poor guide to actual conduct, given that so much of ethical behaviour is embedded in one's social context. The framing of such dilemmas will also be of huge importance. Consider, for example, the difference between introducing a task to a subject as 'a dictator game with a stranger' as opposed to 'a distribution task with a fellow-student'. The framing here may trigger very different feelings about one's role in the task, and one's relationship to other participants.

31

The simple distinctions used in the bulk of this paper have been between right and wrong, good and bad, or permissible and impermissible. But virtues are narrower and more complex conceptions of im/moral phenomena. Some of these concepts may distort the boundaries of our original, simple distinctions. We might say that an act was 'beneficent', 'courageous', 'harsh but fair', 'devious', 'exploitative', or intended to 'save face', to give a few of the many possible examples. Following Williams (2006) it is common to call these 'thick concepts' (as opposed to 'thin' concepts such as 'right' or 'bad'). How important thick concepts are in contemporary moral practice is debatable, but what seems clear is that they are not simple concepts to master – one has to be both a competent language user and a competent social actor to see exactly how the term is properly applied. They may be hard to analyse, and any simple dictionary definition is liable to be a poor guide to their use in practice. While Williams himself rejected the label of 'virtue theorist', it is essential to the understanding of aretaic values to recognise that they are often not *simple* values, and they tend to be somewhat culturally-relative. Even the relatively simple notion of 'generosity' provides an example: a 'generous' person is not simply someone who gives resources away; the label suggests something about the motivations of the person, and the relationship between, or status of, the giver and the recipients. (A relevant example of the use of thick concepts in spillovers research are the lists of virtues in Sachdeva et al.'s 2009 experiments, and Blanken et al.'s 2014 attempts to replicate these with the list translated into Dutch.) The 'thickness' of virtues like generosity is likely to make them resistant to quantification and observation, particularly if we want our conclusions to be widely generalisable. Still, the topic is there for those prepared to engage with it.