Synthese (2017) 194:4477–4497
DOI 10.1007/s11229-016-1145-6

Responsibility for strategic ignorance

Jan Willem Wieland[1]

Received: 1 March 2016 / Accepted: 14 June 2016 / Published online: 29 June 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Strategic ignorance is a widespread phenomenon. In a laboratory setting, many participants avoid learning information about the consequences of their behaviour in order to act egoistically. In real life, many consumers avoid information about their purchases or the working conditions in which they were produced in order to retain their lifestyle. The question is whether agents are blameworthy for such strategically ignorant behaviour. In this paper, I explore quality of will resources, according to which agents are blameworthy, roughly, depending on their moral concern. The account I will propose—the Maximal Account—has three innovative features: (1) it utilizes a suitable concept of maximal moral concern, (2) it offers an accountability version of the account which significantly differs from the more familiar attributability variant, and (3) it maintains that agents without maximal concern are blameworthy (in some sense, to some degree).

Keywords Strategic ignorance · Quality of will · Moral concern · Accountability · Attributability · Degree

✉ Jan Willem Wieland
j.j.w.wieland@vu.nl

1 Department of Philosophy, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

1 Strategic ignorance

Strategic ignorance is a widespread phenomenon. We often avoid information in order to continue dubious practices. Example 1: the experiment by Dana et al. (2007). This 'DWK' experiment has been influential in behavioural economics. It has a dictator game set-up. The first participant, the dictator, has a choice between two options, 'A' and 'B'. If she chooses A, she receives $6; if she chooses B, she receives $5. The second participant, the receiver, has no choice. What she gets will depend on the dictator's choice. There are two different settings. In the first setting, called the 'baseline treatment', the dictator knows what the receiver will get, namely $1 if she opts for A, and $5 if she opts for B:

Action	Dictator	Receiver
A	6	1
B	5	5

In the second setting, the 'hidden information treatment', the dictator does not know whether the receiver will receive $1 or $5. There is .5 chance that the outcomes are as in the baseline treatment, but there is also .5 chance that the outcomes are as follows:

Action	Dictator	Receiver
A	6	5
B	5	1

So the dictator does not know which distribution will apply. She is told that she can come to know this by pressing a button in front of her, but she might also make a decision right away, that is, without knowing what the receiver will get as a consequence of her choice.

The results of the experiment were surprising.[1] In the baseline treatment, 26 % of the dictators chose the egoistic option A, and 74 % the fair option B. In the hidden information treatment, in contrast (with the baseline distribution, where B is the fair choice), 62 % of the dictators chose the egoistic option, and only 38 % the fair one. Moreover, in the latter case, 50 % of these participants did *not* press the button to gain more information, but chose to remain ignorant. As Dana et al. put it: "many subjects behave consistently with a desire to remain ignorant to the consequences of a self-interested choice" (2007, p. 75).

Example 2: consumers. Suppose you want to buy a T-shirt and have two comparable options 'A' and 'B'. Option A is cheaper, though made by workers who are exploited. Option B is a bit more expensive, though made in fair conditions. In the baseline

[1] For more numbers, see Dana et al. (2007), and subsequent studies that have replicated their result. For an overview of these studies, cf. Grossman and Van der Weele (2015); for related experiments, cf. Spiekermann (2016).

treatment, you know this. In the hidden information treatment, you know you have two options (one cheap, the other more expensive), but do not know which is made in fair conditions. Still, you can ask about this, and if you do you'll receive information about the two T-shirts. As in the DWK experiment, my prediction is that most people will buy the fair T-shirt in the baseline treatment, but the unfair one in the hidden information treatment.[2]

Action	Consumer	Product
A	Cheap	Unfair
B	More expensive	Fair

Of course, there are differences between the two cases. The lab abstracts all sorts of factors, including distraction, mistakes about what's going on, and social pressure (in the experiment, the receiver won't know whether the dictator pressed the button). And it matters whether you know what's at stake, whether you're allowed to play several times and revise your choices in the light of what you've learned, whether you're supposed to motivate why you're doing what you're doing, whether you can see what others do in similar circumstances, and so on.

Also, not all consumer cases have a DWK-like structure: sometimes there is no fair alternative or easily accessible information. If we abstract from such different cases, though, the two examples are sufficiently similar. Two parallels stand out. Parallel 1: control. In the game, the dictator has a certain control over the outcome, namely whether the receiver gets $1 or $5. However, she has no control over the distribution of these numbers, nor over what other receivers get in other games. Similarly, consumers have control over whether the product they purchase is made in fair or unfair working conditions, but they have no control over the specific wages, or the working conditions in general.

Parallel 2: strategic ignorance. You don't know whether T-shirt A or B is a fair product made in acceptable working conditions, yet in many cases you can find this out by searching for information about the label online. Moreover, just as many participants in the DWK experiment avoid learning information about whether the receiver gets only $1 (and opt for the $6), many consumers will not seek information about whether their T-shirt is made on the basis of exploitation (and will opt for the cheapest one, or the easiest available, or the most popular).

Generally, S is strategically ignorant of the truth of a proposition p iff, roughly, S could gather information about p, yet does not do this because she does not want to.[3] In the DWK experiment, 'p' is the proposition that the receiver will get only $1 if one chooses to get $6, and in the consumer case it's the proposition that the T-shirt one wants to buy is made on the basis of exploitation. One interesting hypothesis on why people strategically avoid information is that they do so because the information would be inconvenient in some way. As consumers, we generally do not want to consider

[2] Unfortunately, we don't have numbers in this case. For numbers in comparable consumer cases, cf. Ehrich and Irwin (2005), Onwezen and Van der Weele (2016).

[3] Strategic ignorance goes under different labels: motivated, affected, willful, contrived.

whether our consuming behaviour is wrong, first because it's in our interests to stay ignorant: we want to keep buying cheap products (the motive of self-interest). Second, we don't want to consider it because we have gone so long without considering it, are in fact still doing so, and acknowledging this will seriously affect the image we have of ourselves (the motive of self-image).[4]

2 The question

In this paper, I'll address the question: if the dictator decides to stay ignorant, *is she blameworthy for any unfair distribution*?

On the one hand, the cases just discussed appear to be clear cases of blameworthiness. For in the given setting with two clear options and freely available information, it seems rather easy to do the right thing.

On the other hand, many DWK participants and consumers don't do the right thing. This might suggest that they don't care enough about doing the right thing, but might also suggest that doing the right thing is harder than it seems to be. According to several experimental studies, furthermore, people regard strategically ignorant agents (agents who don't press the button, and choose egoistically) as less blameworthy than agents who are knowingly egoistic (agents who choose egoistically after they have pressed the button).[5] This does not suggest that strategic ignorance takes away *all* blameworthiness, but that it might mitigate it. Yet this contradicts the initial reaction that strategically ignorant agents are fully blameworthy. Hence, we're asking a tricky question. Does strategic ignorance excuse, or not?

Before further explaining the question, a brief comment is in order concerning our obligations. In DWK games, I will assume that one has moral obligations to realize fair outcomes and avoid unfair ones. For the purposes of this paper, I will stay neutral about the normative theory from which such obligations can be derived (if they can be derived). Similarly, in the consumer case I'll assume one has a moral obligation to buy fair products rather than unfair ones. This seems plausible, even though it might be tricky to derive such obligations from one's normative theory.[6]

Furthermore, these obligations to realize fair outcomes and buy fair products come with a certain obligation to press buttons. Generally, one should inform oneself because doing so will enable one to see what one has to do (cf. Smith 2014, p. 20). In the present context, this means that agents have to gather information about the outcomes of their choices. In other cases, they may also have to inform themselves about the choices they have, or about what to do, given information about the choices and outcomes.

[4] For more on the motive of self-interest, cf. Moody-Adams (1994). For more on the motive of self-image, cf. Dana et al. (2007), Grossman and Van der Weele (2015). In this paper, I won't assume that S's motives should be transparent to S, i.e. that S can always report on her motives. What I am interested in are facts about S that explain her conduct.

[5] Cf. Krupka and Weber (2013), Conrads and Irlenbusch (2013), Bartling et al. (2014), Grossman and Van der Weele (2015).

[6] For one thing, buying an unfair product does not make a difference to the worker's situation in any straightforward way (cf. Kagan 2011). I'm sympathetic to a suggestion by Lawford-Smith (2015), according to which our obligation to buy fair products derives from our obligation to signal our commitments.

To have these obligations is one thing; to be blameworthy for not complying with them is another. For one might be excused. Questions of blameworthiness should always be posed carefully. The term has several senses, and one can be blameworthy or blameless depending on what kind of blameworthiness one has in mind. In this paper, I'll focus on *moral* blameworthiness rather than legal punishment, and in particular on two different kinds of moral blameworthiness (attributability and accountability, see Sect. 4). When it comes to scope, I'll assume that one might be blameworthy not only for omissions to press the button and the ensuing strategic ignorance, but also for the outcomes of such omissions (such as unfairness). The main question I'll be asking concerns the latter.

Hellman (2009) distinguishes three different views on the issue.[7] In terms of moral blameworthiness for unfair outcomes, these views can be stated as follows. Strategically ignorant agent S is blameworthy for the unfair outcome X she realizes iff

- S's decision to stay ignorant about X was unjustified (justification view);
- S's conduct that led to X was reckless (recklessness view);
- S's motivation to stay ignorant about X was to preserve a defense against liability, or is morally bad in another way (motivation view).

At first sight, these views seem applicable in a fairly straightforward way to the DWK and consumer cases. In these cases, one's decision to remain ignorant was unjustified. Whether such a decision is justified depends on whether the reasons to remain ignorant actually outweigh the reasons to investigate. In these cases, though, the reasons to remain ignorant (self-interest or self-image) do not outweigh the reasons to investigate (fairness). Moreover, to decide to remain ignorant is reckless: you are aware that you take an unjustified risk. In the hidden information treatment, if you don't press the button and choose A, you know there's a .5 chance that you will realize an unfair outcome. Similarly, if you don't ask whether the T-shirt was made in fair working conditions and choose to buy the cheapest one, you know there's a serious chance that the product was made in unfair circumstances.[8]

How about the motivation view? For one thing, it does not seem right to say that the DWK participants and consumers are trying to preserve a defense against liability when they decide to remain ignorant. As Sarch (2014, pp. 1073–1078) convincingly argues, however, agents might still be considered blameworthy if their motivation is sufficiently bad in another way. How can we determine whether this is so?

On this point, the quality of will approach may be helpful. It has proven attractive to many philosophers (at least since Strawson 1962). On this approach, blameworthiness is analysed in terms of an agent's motivations, and especially the extent to which one is motivated by what matters morally. The general idea is that one is blameless when one

[7] Hellman herself endorses the justification view (on which blameworthiness is a function of objective reasons to investigate), but the other views (on which subjective elements matter, such as the agent's awareness or her own reasons) have proponents as well. For the recklessness view, cf. Robbins (1990) and Alexander et al. (2009). For the motivation view, cf. Husak and Callender (1994) and Sarch (2014).

[8] It's controversial whether *all* strategically ignorant agents (such as consumers) are fully aware of what they're doing. If not, their blameworthiness might be analysed in terms of negligence rather than recklessness. You're reckless, the assumption is, if you're aware that there's an unjustified risk of wrongdoing, and you're negligent if you're culpably unaware of this.

acts out of moral concern, and blameworthy when one acts without such a concern. As always, however, everything depends on the details. On the one hand, strategically ignorant agents don't seem to care enough about fairness when they ignore the free information. On the other hand, as Ehrich and Irwin (2005) point out, strategically ignorant consumers do seem to care to a certain extent, given that they are far from indifferent. That their wooden desk comes from endangered rain forests, or that their cell phone is made by child labourers, causes them stress and other negative emotions.

Thus it remains to be seen to what extent and in what sense strategically ignorant agents act without moral concern, and so to what extent they can be considered blameworthy on the motivation view. To determine this will be the core task of this paper. For the most part, I'll set out and defend a version of the motivation view and won't directly criticize the justification and recklessness views. Given the attractiveness of the motivation view, this is an important result on its own. Still, I will return to the dispute between the three views at the end of the paper, where I'll briefly explain why I prefer the motivation view.

This paper builds upon Nomy Arpaly's influential quality of will account as proposed in her (2003) as well as her (2014) with Timothy Schroeder. I will proceed in three steps. First, I introduce and apply a suitable concept of maximal moral concern (in Sect. 3). Second, I identify a novel accountability version of the view which differs from Arpaly's attributability version, and identify and defend certain surprising implications of this distinction (in Sect. 4). Finally, I discuss and reject the idea that we need to relax the notion of maximal concern (in Sect. 5).

3 Moral concern

Do strategically ignorant agents act with moral concern? To determine this, we need to know what it is to act with moral concern. According to Arpaly (2003, chap. 3), S acts with moral concern iff S acts with concern for the wrongmaking or rightmaking features of her action. Huck Finn acts with moral concern when he helps the slave Jim escape because he acts with a concern for Jim's humanity (which, let us suppose, makes the action of letting Jim escape right).[9] Given that we're focusing on blameworthiness for bad outcomes, moral concern here amounts to a concern for those outcomes. For example, participants in the DWK experiment and consumers act with moral concern when they act with a concern for fairness.

What does such a concern involve? First of all, as Arpaly argues, concern for a bad outcome X doesn't involve a belief that X is important and to be prevented. After all, Huck does not believe that Jim's freedom is important (rather, he believes it is *not*), and yet he is concerned about it. Instead, concern for a bad outcome X involves:

(1) noticing X;
(2) detecting and not ignoring available information about opportunities to prevent X;

[9] This is not to be confused with a concern for features that *she thinks* make her conduct wrong or right. After all, Huck Finn does not believe that he should help Jim and that Jim's humanity is something he should act on.

(3) taking these opportunities to prevent X;
(4) not having internal difficulties in (1)–(3) because one finds the truth about X inconvenient;
(5) not making up rationalizations for the claim that X should *not* be prevented;
(6) not being distracted by or concerned about things which matter less, morally or otherwise;[10]
(7) not being distracted by the misleading behaviour and views of others who do not fulfil (4)–(6);
(8) fulfilling (1)–(7) in similar, or even more difficult, circumstances.

This is a carefully designed list. Most clauses derive from Arpaly's work (cf. also Björnsson 2016). My specific additions to the list are (4) and (7), which play an important role in strategic ignorance cases (as I'll discuss soon). This list sounds like a lot, though if you're concerned, these clauses come in a package. For example, if you're concerned about your friends' welfare, you'll notice when things don't go well for them, attend to opportunities to help them, take those opportunities, not find this inconvenient, not make up reasons why you shouldn't do these things, not be distracted by other things you wanted to do with your time, and not be distracted by others who don't help them.

The list is especially designed for strategic ignorance cases and not meant to be exhaustive. For example, just as concern involves an attentiveness to available information about opportunities to prevent X, it will also involve an effort to create or help create such opportunities (where needed and possible). Yet, the latter clause is irrelevant in DWK games and to a large extent also in consumer cases. In those cases, all options are simply given to you.[11] I have also left out the emotional dimension. It's plausible to think that often concern for X involves a certain stress (or other appropriate emotions) if X has not yet been prevented. Still, I don't take this dimension to be essential.

Given that one might satisfy (1)–(8) to a greater or lesser extent, moral concern is a matter of degree. The more these clauses apply to you, the more you care about X. The less they apply to you, the more you're indifferent about X.[12] Clause (8) has a special role in this respect. Call it the 'robustness' of moral concern. If you're not only concerned about a specific bad outcome X (and satisfy (1)–(7) in a single case), but also about similar kinds of outcomes (and perhaps even in more difficult circumstances), then your concern can be said to be greater.[13]

I'll have more to say about these degrees, and how they map onto degrees of blameworthiness, in Sect. 4. For the moment, though, let us assume that concern is a

[10] This is compatible with the counterpart clause that concerned agents do not spend energy on preventing X if they have more important things on their mind (that is, things which are *in fact* more important, not things that the agent *thinks* are more important).

[11] If you're really concerned, though, you might still question the status quo, and ask for information or alternative options even if they don't yet exist.

[12] Modulo certain qualifications. Also, indifference should be kept distinct from ill will. Ill will involves the absence of only some of (1)–(8). For example, if you're ill willed and desire to establish bad outcome X, then you'll still be attentive (though to opportunities to realize X). More about ill will in Sect. 4.

[13] When the circumstances are too difficult, and the difficulty isn't due to a lack of concern on S's part, (8) reads: S is disposed to fulfill (1)–(7).

binary affair: you have it and fulfill (1)–(8), or you don't have it. If you do have it, let us call you a *maximally concerned agent* (or 'moral hero', as I'll call this kind of agent in Sect. 5). The question in the following will be whether such a maximally concerned agent does avoid a given unfair outcome, or whether she is obstructed by some source of difficulty. I'll consider the following possible sources: (a) a lack of alternative outcomes; (b) the sacrifice required by S; (c) the inaccessibility of information about X; (d) S's cognitive incapacities; (e) S's unawareness; (f) S's internal barriers; (g) S's social context.[14] I'll argue that maximally concerned agents will only be obstructed by (a)–(c).

(a) Suppose that you're playing the following game:

Action	Dictator	Receiver
A	6	1
B	5	1

In this game, you face two choices, and both lead to an unfair outcome for the receiver (for the sake of the argument, let's assume that they're equally unfair). Even concerned agents wouldn't realize a fair outcome in such cases.

(b) Suppose that you're playing the following game:

Action	Dictator	Receiver
A	6	1
B	0	5

In this game, you face two choices. If you choose A, you'll get $6 and the receiver will get only $1. If you choose B, you'll get nothing and the receiver will get $5. Would concerned agents avoid the unfair option A? I don't think so. As before, A's outcome is considered unfair, for the dictator gets $5 more than the receiver. Yet, in this case (and in contrast to the baseline treatment), choosing 'B' requires a sacrifice on S's part (namely, she would have to give up $6) which is greater than the unfairness involved in the outcome (namely, a difference of $5). Even if you care a lot about fairness, you're not supposed to make this sacrifice.

The general idea is this: *concern about X does not require a sacrifice on S's part which is greater than the unfairness involved in X to which others would be subjected*. A similar idea has been suggested by Bradford (2016), who motivates it on the basis of what she calls the 'life takeover worry'. Choosing B in the game just mentioned

[14] For some of these factors, cf. Nelkin (2016), Bradford (2016). Again, the list is not intended to be exhaustive, yet it does capture all potential obstacles in strategic ignorance cases. Some factors might seem to overlap (e.g. (c) and (e), or (b) and (g)), though below I'll clearly set them apart.

would take over your life, so to speak. If moral concern does not require any such thing, even morally concerned agents can be obstructed by sacrifice. Generally, I will consider sacrifice a factor which comprises money, time, or something else that the agent has to give up in order to act fairly. So cases of sacrifice include consumer cases where they have to pay a lot for fair products, or where the latter are so exclusive that consumers have to devote a substantial portion of their time to finding them.[15]

(c) Suppose that you're playing the hidden information treatment (where the chance that A is the unfair outcome is 50 %), and that the information about the outcomes isn't easily accessible via a free button. Rather, it costs $6. For reasons just explained, this will obstruct concerned agents. For even if you care a lot about fairness, you're not supposed to pay $6 only in order to get a maximum of $6 in return.

(d) Suppose that you're playing the hidden information treatment, and that you have weak cognitive capacities and so cannot see all the options and consequences in a clear way. In a straightforward sense, some people are smarter than others, and this holds whether or not they care about morality.

Nevertheless, the things you're able to notice do depend in part on your concern. If you're concerned about your football team, then no matter how smart you are you'll be able to recognize things during their matches that uninterested parties will miss (cf. Arpaly 2003, pp. 86–87, 110–111). The same applies to the DWK experiment: if you care about fairness, then you are able to see all the options and consequences in a clear way, and it will be easy for you to see that you should press the button and choose the fair alternative. It follows that, at least in the present context, concerned agents won't be obstructed by weak cognitive capacities.

I'd concede that the abstractness of the set-up makes a difference here (cf. Gendler 2007). It's easier to care about fair products, and to recognize relevant features of one's consuming behaviour, than to care about fairness in DWK games. It's easier to care if the unfairness is represented concretely (involving the exploitation and suffering of specific people), rather than abstractly (involving only numbers). Still, I don't think the DWK games are so abstract that they would be difficult to grasp for someone with moral concern.

(e) Suppose that you're playing the hidden information treatment and that obstacle (c) does not apply: the information is freely accessible. Still, you're avoiding it, and you're unaware or only vaguely aware that you're doing this in order to act egoistically. It might even be that you're in denial (i.e. you tell yourself that you're not really acting egoistically, or that choosing egoistically is permissible), or you're trivializing the issue (i.e. you tell yourself that it is morally unimportant), or you're rationalizing (i.e. you make up nonsense reasons for choosing the egoistic option A in the face of good available reasons against A). These symptoms are actually quite common in the consumer case.[16] Consumers tell themselves, for example, that buying questionable

[15] To apply Bradford's criterion, though, we'd still need to weigh these costs against the costs of the consequences. Importantly, I consider (c) distinct from the factors (f) and (g) below: you don't sacrifice anything in a relevant sense if you have to overcome internal and social difficulties [as Pleasants (2011, p. 148) seems to assume]. They only count as sacrifices in the special cases where overcoming such difficulties costs you money or friends.

[16] Cf. Chatzidakis et al. (2007), Eckhardt et al. (2010), Paharia et al. (2013), Gruber and Schlegelmilch (2014).

products is permissible because it creates work (no matter what work, in what conditions, and for what wages). Or they tell themselves that only others are responsible (the companies, the government). Still, had they cared about fairness, they wouldn't exhibit such symptoms and would be aware of what they're doing (and unawareness wouldn't be an obstacle).

(f) Suppose that you're playing the hidden information treatment where information is accessible via a free button, but that it's psychologically difficult for you to press it. For example, suppose you want to get those $6, but you also care about your self-image. This combination of concerns makes you disinclined to find out whether your choice is fair or not. After all, knowing that you're selfish is bad for your self-image. Or suppose you're a consumer and really want to have the newest smartphone. In such a case, it might be psychologically difficult for you to read about the problems in the supply chain. For one thing, you already bought a whole series of those phones made in problematic circumstances. For another, you want one more of them, and you don't want to make any changes to your lifestyle. Still, if you cared more about fairness, you wouldn't have such internal barriers and it wouldn't be psychologically difficult for you to press the button. So internal barriers wouldn't form any obstacle for concerned agents.

(g) Suppose that you're playing the hidden information treatment, and that obstacles (a)–(c) do not apply. So the information is freely accessible, alternative outcomes are open to you, and you don't have to sacrifice anything. Yet this time there is a social context, which is, moreover, misleading: virtually none of your peers press the button, or act fairly. Furthermore, all of them exemplify one of more of the typical symptoms of strategic ignorance mentioned earlier (denial, trivialization, rationalization). Given this, it's socially difficult for you to press the button and do the right thing. Again, moral concern would make a difference here. If you'd care about fairness, you wouldn't be distracted by your peers.[17] Concerned agents won't be obstructed by a misleading social context.

All in all, only conditions (a)–(c) obstruct maximally concerned agents from doing the right thing.[18] As noted, on the quality of will approach, agents are blameworthy depending on their moral concern. One possible way of making this more precise is as follows: S is blameless for X if she is maximally concerned, i.e. fulfills (1)–(8), with respect to X; and S is blameworthy if she isn't maximally concerned. On this account, call it the 'Maximal Account', agents are only excused if they're obstructed by a lack of alternative outcomes, sacrifice, or inaccessible information. Applied to our two cases, it would follow that both DWK participants and the consumers from Sect. 1 are blameworthy. For they do have information available to them (that they're avoiding), as well as options to realize fair outcomes (that they're not taking), which do not require too much sacrifice on their part. Consumers would be excused only when their situation is sufficiently different from the DWK case, that is, when there

[17] This is not supposed to imply a problematic form of dogmatism. What matters is that you're concerned about X, if X is in fact something to be concerned about, and not that you ignore your peers whenever you strongly (and perhaps mistakenly) believe that X should be prevented.

[18] As argued, (d) might do so as well, though not in the current context.

are no fair alternatives (such as clothes and phones made in fair circumstances), when those alternatives are too costly, or when there's no information about these things.

4 Attributability vs. accountability

On the Maximal Account, then, strategically ignorant agents are blameworthy since they act with a lack of maximal moral concern. You might think the Maximal Account is too demanding. In Sect. 5, I'll argue that this is mistaken: there's no need to relax it. In this section, I'll qualify the Maximal Account and ask *in what sense* and *to what degree* strategically ignorant agents are blameworthy. Standardly, you're blameworthy for X whenever you deserve blame for X and are the proper candidate of reactive attitudes such as resentment, indignation, and condemnation. At least since Watson (1996), however, many philosophers acknowledge at least two kinds of blameworthiness, namely attributability and accountability. Basically, I understand these notions as follows: attributability concerns an agent's moral concern that motivated her conduct, while accountability involves certain expectations regarding an agent's moral concern. I'll argue that this distinction has important ramifications.[19]

More precisely, we'll assume: a bad outcome X is attributable to S, and S is blameworthy in this sense for X, iff, and in the extent to which, *her conduct that led to X speaks badly of her will*.[20] This is a fairly straightforward notion. For example, realizing X when it's obvious to you that X shouldn't be realized speaks worse of your will than doing this when it's unclear to you that X shouldn't be realized. Sometimes, though, X may be attributable to you even when you're not accountable for X. That is, your conduct may speak badly of your will, even though you couldn't be expected to do better. Perhaps you couldn't help your lack of concern. Specifically, I'll understand accountability in the following way: S is accountable for X, and S is blameworthy in this sense for X, iff, and in the extent to which, *S could have been expected to have had more concern regarding X*.[21]

There is a familiar distinction between normative and predictive expectations (cf. Clarke 2014, pp. 181–182). You're normatively expected to comply with your obligations, that is, to press buttons and to realize fair outcomes (as noted in Sect. 2). Expectations in the predictive sense aren't set by what you should do, but by what it is likely that you will do given what we know about you. Perhaps you've never pressed any button, and so we do not predict (or expect, in that sense) that you'll press it this time.

The expectations I have in mind for accountability are normative, but do not have the content just mentioned. When I hold you accountable, I don't particularly think that you should have complied with your obligations (after all, this is normatively

[19] Cf. Mason (2015). Further kinds of blameworthiness (e.g. answerability) fall outside the scope of this paper.

[20] Arpaly uses this notion of blameworthiness in her (2003), and explicitly classifies it as an attributability account in Arpaly and Schroeder (2014, p. 160).

[21] The suggestion that blameworthiness relates to certain expectations can already be found in Strawson (1962). Still, it's not always clear what these expectations are supposed to amount to and below I'll make a specific proposal.

expected of *all* agents). Nor do I think that it was likely you would have done better. The expectations I have in mind aren't immediately defeated by reasons to think that they won't be met, though they can be defeated by reasons that they will *never* be met. If you lack the right sort of capacity for moral concern, expectations that you should have more of it aren't appropriate. I won't need to make any substantial assumptions about these capacities, and will just assume that you have a capacity for moral concern when you have a track record of satisfying clauses of moral concern (1)–(8) (from Sect. 3). Basically the idea is that, when I hold you accountable, you have the right sort of capacity for moral concern and so could have done a bit better than you did. When I expect more of you, furthermore, I think that you should have exercised that capacity and done better.[22]

I take accountability to be an interesting form of blameworthiness. For I don't think we merely resent strategically ignorant agents if we regard them as blameworthy. We also expect more of them. Generally, it's controversial whether attributability is an interesting form of blameworthiness as well. According to FitzPatrick (2016), for example, attributability is no substantial form, since it lacks a historical dimension: people shouldn't be considered blameworthy for their vicious conduct (which does speak badly of them) if they had no chance to be less vicious. Others disagree. According to Talbert (2016), if these agents didn't have any chance to be less vicious, they could still be regarded as blameworthy just in the sense that their conduct manifests disregard for others. I won't take sides here, and will work with both notions. In particular, I'll contribute a novel accountability variant of the quality of will account, which importantly differs from the familiar attributability variant.

4.1 Attributability

As noted in Sect. 2, experimental studies have demonstrated that people regard strategically ignorant agents as less blameworthy than agents who are knowingly egoistic. Importantly, what kind of blameworthiness these studies track is an open question. For example, asking which type is more social (Grossman and Van der Weele 2015) or deserves a stronger punishment (Bartling et al. 2014) is not exactly the same as asking the attributability or accountability question. But let us see whether there is any relevant difference between these types of agents when it comes to attributability (I'll address accountability next).

In my view, there is an interesting distinction to be made among four types of agents. Type (1) cares about fairness (call it 'good will'). This type will press the button, and choose fairly. Type (2) cares about fairness to some extent, but cares more about money and self-image ('impure indifference'). This type will avoid the button and choose egoistically in the hidden information treatment, but choose fairly in the baseline treatment (where all information is given to them, and they don't have the option to remain ignorant). Type (3) doesn't care about fairness or self-image, but only about money ('pure indifference'). This type chooses egoistically in all settings (and

[22] These are *backward-looking* expectations, which are in principle distinct from forward-looking ones (according to which S can be expected to have or develop more concern for X-like outcomes in the future).

won't press any button if that's not needed to maximize her welfare). Types (2) and (3) are similar in that both won't press the button (and count as strategically ignorant), though differ in the robustness of their unfair behaviour.[23] Type (4) does care, yet in such a way that they desire an unfair outcome ('ill will'). This type will press the button in order to ensure that the receiver gets the unfair amount.[24]

On the assumption that behaviour in the DWK experiment is robust enough and representative of different types of agents (with different capacities for moral concern),[25] 38 % has good will (agents who press the button and choose fairly in the hidden information treatment with the baseline distribution), 36 % is impurely indifferent (the difference between fair agents in the baseline (74 %) and agents with good will), 13 % has ill will (agents who press the button and choose unfairly in the hidden information treatment with the baseline distribution), and 13 % is purely indifferent (the difference between unfair agents in the baseline (26 %) and agents with ill will).

Hence, impurely indifferent agents constitute an important group. They're indifferent, but only impurely so, because to some extent they do care about fairness. At least, they appear to care about fairness in the baseline treatment. One might simply say they're all too human. I'd speculate that the majority of consumers fall within this category. We often avoid information about the unfair products we buy, though we would consume fairly if we had been clearly informed about the problems in the supply chain and about fair alternatives (cf. Ehrich and Irwin 2005). Only a few of us care nothing about others, or have ill will.

Given these four types, who is more blameworthy in the attributability sense, that is, for choosing unfairly in the hidden information treatment? The ordering is fairly straightforward (where '<' stands for 'is less blameworthy than')[26]:

$$(1) < (2) < (3) < (4)$$

Pressing the button in order to act unfairly speaks worse of your will than not pressing the button because you only care about your own welfare. The latter speaks worse of your will than not pressing the button because you care about your self-image. And the latter, in turn, speaks worse of your will than pressing the button because you care about fairness. As we can see, this ordering corresponds to the findings of the experimental studies mentioned earlier: at least in the attributability sense, strategically ignorant agents are less blameworthy than those who are knowingly egoistic.

[23] Both types fail to satisfy the clause 'not ignoring available information about opportunities to prevent X' of moral concern (in Sect. 3), though type (2) does satisfy 'taking these opportunities to prevent X' when she's not ignorant.

[24] The behaviour in the DWK experiment is only consistent with these types. It may well be that it's not robust enough to amount to a certain type. Suppose someone presses the button (just out of curiosity), and then decides to act egoistically. Such behaviour is merely consistent with the behaviour of ill willed agents.

[25] And relying on DWK data (and ignoring the behaviour of agents in the hidden information treatment where A, not B, is the fair outcome).

[26] For the purposes of this paper, I'll keep fixed further factors that might have an impact on one's degree of blameworthiness (such as the degree of unfairness involved in X, the degree of the risk that X will be realized, or the degree to which information about X is available).

Markovits (2012, pp. 308–309) has questioned whether purely indifferent agents are really less blameworthy than agents with ill will. She discusses a purely indifferent politician who cares more about his own reputation than about the execution of a man who might well be innocent. I agree with the response by Arpaly and Schroeder (2014, p. 189) that, so long as it concerns attributability, such a politician is less blameworthy than a politician with ill will (who desires to execute innocent people, regardless of whether his reputation is at stake). There's a difference in how their conduct speaks badly of them, and in this respect the purely indifferent politician fares slightly better. Still, as I'll argue next, I'd agree with Markovits that there's something "particularly abhorrent" about the purely indifferent politician.

4.2 Accountability

According to quality of will accounts, lack of moral concern does not excuse, and is a ground for blameworthiness. Yet, as discussed in Sect. 3, lack of moral concern comes with a number of obstacles (i.e. reduced cognitive capacities, reduced awareness, and/or internal and social barriers). To be sure, in strategic ignorance cases, it's still possible for the agents to press the button and go for the fair option. But a lack of concern might make things complicated. If you're not very interested in fairness, then it might take you a great deal of effort to pay more attention, change your lifestyle, and do better. Given this, one might think that lack of moral concern is a real handicap rather than a ground for blameworthiness.[27]

The objection might also be put in tracing terms:

(P1) S is blameworthy for a bad outcome X only if S is blameworthy for her lack of moral concern for X;
(P2) S is blameless for her lack of moral concern for X;
(C) Therefore, S is blameless for X.

In response, I'll concede a version of (P1), but deny that (P2) applies to most of us. (P1) is false when understood in terms of attributability. After all, acting with a lack of concern can speak badly of your will irrespective of whether you could help your indifference. Yet, (P1) may be true when taken in terms of accountability. We might not expect more of you irrespective of whether you could help your indifference. Let us say that you are blameless for the latter if you lack the right sort of capacity for moral concern. For if you lack such a capacity, it won't be appropriate to expect you to be more concerned. Hence the question: do participants in the DWK experiments or consumers have the right sort of capacity, and can they be held accountable for their lack of moral concern?

Of course, you cannot by fiat decide to care about specific things. Still, you do have a certain indirect control over what you care about. To a certain extent, you have the capacity to develop concern for your friends, your job, the philosophical issues you happen to find interesting, your hobbies, the clubs you support, and so on, by devoting time and energy to them. The same applies to consuming: you might well develop

[27] This problem is due to Jojanneke Vanderveen, and is inspired by Wolf (1987).

concern for the circumstances in which your purchases are made. Even granting this, the crucial question is whether we should expect agents to develop such concern.

As explained, when I hold you accountable and expect that you would have done better, you must have the right sort of capacity for moral concern which might have enabled you to have done a bit better than what you did. This reaction seems to fit the large group of impurely indifferent agents. After all, to a certain extent these agents do care about fairness, but they're all too human and are easily distracted by other things (such as money and their self-image). Still, given that they have a track record of choosing fairly in the baseline treatment, they do have the right sort of capacity, and the relevant expectations are well in place.

The same does not apply to the smaller group of purely indifferent and ill willed agents. They seem to lack a capacity for moral concern. Purely indifferent agent exhibit no moral concern and won't be motivated to develop it. Ill willed agents have the wrong kind of concern altogether (namely for what's immoral), and won't be motivated to develop the right kind. Still, in this respect ill willed agents might even fare better than purely indifferent agents, given that the former at least have a track record of satisfying some of the clauses (1)–(8) (for example, they will be attentive to relevant information). At any rate, in both cases the relevant expectations are largely misplaced. We would rather exclude such agents from the moral domain, or offer them moral education.

Given all this, who is more blameworthy in the accountability sense, that is, for choosing unfairly in the hidden information treatment? The accountability ordering is as follows[28]:

$$(1) < (3) \approx (4) < (2)$$

That is, we don't expect *more* of good willed agents (though we might think it would be good if they were to exhibit similar concern in the future). We don't expect much of ill willed agents and purely indifferent agents. Yet, we do expect impurely indifferent agents to do better, and hold them accountable to the greatest degree. As we can see, this ordering differs, and interestingly so, from the attributability ordering.

4.3 Objections

Next, I'll respond to two objections. First objection: my account is preempted by a similar account by Luban (1999). Second objection: if this is so, my account suffers from the same criticism that Luban's account faces from Sarch (2014).

Luban distinguishes three types of strategically ignorant agents: the fox, the unrighteous ostrich, and the half-righteous ostrich. In terms of DWK games, all of them are strategically ignorant that their choice (namely A) is wrong, but they have different counterfactual profiles. When given full knowledge that A is wrong, the fox would choose A anyway, because she wants to do "the crime". In such a situation, the unrighteous ostrich would also choose A, but not because she wants to do the

[28] Where '$<$' stands for 'is less blameworthy than', and '$\approx$' for 'is blameworthy to a similar degree as'. To recall: (1) good will; (2) impure indifference; (3) pure indifference; (4) ill will.

crime. The half-righteous ostrich, by contrast, would not choose A when given full knowledge that A is wrong. Moreover, according to Luban, these differences matter: the fox is more blameworthy than the unrighteous ostrich, and the unrighteous ostrich is more blameworthy than the half-righteous ostrich.

In response to the first objection, I acknowledge that there are some similarities between Luban's classification and my own. The unrighteous ostrich is similar to my impurely indifferent agent, the unrighteous ostrich to my purely indifferent agent, and the fox to my agent with ill will (though the latter is no longer strategically ignorant).

Still, there are at least two important differences as well. First, on my account the counterfactual profiles of these agents are the result of their having certain motives. The impurely indifferent agent is concerned about fairness (to some extent), but she is also concerned about money and her self-image. That is why she decides to remain ignorant, but would have chosen the fair option if she hadn't been ignorant. In contrast, the purely indifferent agent is concerned only about her welfare. That is why she does not inform herself if she can maximize her welfare right away, and also why she would not choose fairly even if she had more information.

Second, and crucially, Luban does not distinguish between attributability and accountability. Given his ordering, he seems to have in mind an attributability form of blameworthiness. The fox is more blameworthy than the others in the sense that her conduct is motivated by the least concern for the interests of others. However, as I just argued, a rather different ordering applies in the case of accountability. When it comes to the latter, the half-righteous ostrich is *more* blameworthy than the others given that she could have been expected to have more moral concern.

For these reasons, my account isn't preempted by Luban's account. Yet it might still suffer from the second objection, namely Sarch's claim that the proposed counterfactual differences between the agents cannot matter to their blameworthiness. Sarch writes: "the mental state one would have had under counterfactual circumstances, but actually lacked, cannot be the basis for how culpable one is for one's actual action. After all, one's counterfactual mental state did not produce the actual action." (2014, p. 1059) Thus, according to Sarch, it's irrelevant that the purely indifferent agent would still have chosen the unfair option A had she not been ignorant, given that she did not actually know that A was unfair and therefore no such knowledge played any role in her choice.

In response, I agree that counterfactual mental states do not produce actual behaviour, and that what matters are actual motives for actual behaviour. But, as I see it, the purely indifferent agent differs from the other agents on more than a merely counterfactual level. Instead, their main difference is that they're concerned about different things. While the purely indifferent agent is merely concerned about her own interests and welfare (and doesn't press the button and chooses egoistically for these reasons), the impurely indifferent agent is also concerned about fairness and her self-image (and doesn't press the button and chooses egoistically for these more complex reasons). These are their actual motives that explain their behaviour.[29] Yet, this does not exclude

[29] At one point, Arpaly and Schroeder suggest: "The amount of good or ill will that an action manifests is not the same as the amount of good or ill will that exists and is being acted on." (2014, p. 188) I agree that there should be a limit on how much good or ill will can be manifested in a particular action, though

that these motives come with different counterfactual profiles (as explained above), and that is where I think the counterfactuals matter. All in all, I take it that Sarch's objection to Luban's account does not apply to my own.

5 Threshold

That moral concern and blameworthiness come in degrees won't surprise quality of will theorists. What's more controversial, I take it, is whether we should accept certain thresholds of moral concern, which are such that if one meets them, one is off the hook. Typically, quality of will theorists will say that agents are supposed to care 'adequately' or 'sufficiently' (cf. Harman 2011; Björnsson 2016), rather than maximally. In Sect. 3, I discussed Bradford's threshold: S meets the threshold (and so is fully blameless for X) if avoiding unfair outcome X requires more sacrifice on S's part than unfairness with respect to others. But one might worry that this isn't enough. For without any further threshold, agents with options to avoid X can only be excused if they're obstructed by sacrifice. All other agents would be blameworthy, if only to a certain degree. The worry I'll address next is that it might be unfair to blame people, if only to a low degree, who aren't moral heroes.

A similar worry applies to the famous experiment by Milgram (1963) on obedience to authority.[30] Here's what Arpaly and Schroeder say about it:

> Participants thought they were delivering harmful, and perhaps even deadly, electrical shocks to other participants in the study. The fact that many participants did what they believed to be so harmful suggests a certain terrible moral indifference on their parts. Perhaps they did not intrinsically desire the welfare of those being shocked very much at all. But a closer look suggests practical irrationality rather than moral indifference. … It might be that, had Milgram's subjects been true moral heroes, they would not have irrationally stayed in the experiment, and so the experiment revealed them to have some very small degree of moral indifference. (2014, p. 168)

In this experiment, the majority of participants delivered the shocks (65 % administered all shocks up to 450 V), and failed to act morally. Had they been moral heroes, and had they cared maximally about the other participants, they wouldn't have obeyed. Yet they did obey. Given that it would take quite some moral concern to do the right thing in the experiment, one might not want to consider them blameworthy. Yet Bradford's threshold won't help here. After all, the experiment abstracts from any sacrifice on the part of the participants (in the relevant sense discussed). They do have to resist certain

Footnote 29 continued
I'd deny that this should be taken to imply that impurely and purely indifferent agents manifest the same amounts of ill will in the hidden information treatment, and would be blameworthy to the same degree in the attributability sense. For, as suggested in Sect. 4.1, not pressing the button because you only care about your own welfare speaks worse of your will than not pressing the button because you care about your self-image.

30 Moody-Adams (1994, p. 299) classifies this experiment as a strategic ignorance case: the participants are strategically ignorant that their conduct is wrong.

social barriers, though it won't cost them any money. In the absence of any further threshold which may render them blameless, they're blameworthy, at least to a certain degree.

To avoid this result, one needs another threshold. If agents meet it, then they're fully excused, even if they're not maximally concerned.[31] For example, if Milgram participants did obey but under protest and stress, we might say they cared enough. Yet, if we want to say this, we need a story. *Where and how to draw the line?* One might suggest that agents meet the threshold if they fail only because their circumstances are too unfavourable (as in the Milgram experiment). However, this is just a restatement of the question: what's too unfavourable? I'll briefly consider two proposals, and argue that it's not clear that we need them.

Proposal 1: S meets the threshold (and so is fully blameless for X) if *more than 60 % (say) of S's peers also realize X*. The idea is not that one's blameworthiness is sensitive to social distraction after all (as discussed in Sect. 3), but that it's sensitive to the actual performance of others. Does a significant group of similar agents act morally? If more than 60 % also fails to act morally, and there's only a small group of moral heroes, then the threshold is low (and you'll meet it). If more people act morally, the threshold is higher (and you might fail to meet it). Such checks can easily be done in the DWK and Milgram experiments. In the baseline treatment, most people act morally (74 %), and if you do not act morally as well, you'll be blameworthy. In the Milgram experiment, as we have just seen, most people fail to act morally, and in that case you'll be blameless as well if you act immorally.

Such numbers are nice indicators of a threshold. Yet they're unreliable. For example, in the hidden information treatment many people fail to choose fairly (62 %), but that doesn't imply that they're thereby blameless.[32]

Proposal 2: S meets the threshold (and so is fully blameless for X) if *changing the environment will be enough to prevent S from realizing X*. This proposal derives from another observation by Arpaly and Schroeder:

> One of Milgram's later findings was that if participants saw another person, who also seemed to be a participant in the experiment, refuse to deliver the harmful shocks then a very large majority of the participants themselves also refused. That their refusal needed such a weak trigger—essentially, just a reminder that it could be done—strongly suggests that many participants were acting contrary to what they on balance desired in seemingly delivering the harmful shocks. (2014, p. 168)

As Arpaly and Schroeder suggest, the participants' lack of concern isn't very robust. For had they been triggered in the right way, they might well have done better. What kind of triggers should we think of? In this case, other agents function as a reminder

[31] As explained, maximal concern does not entail that one devotes one's whole life to preventing X and gives up all one's other pursuits for the good cause. Still, it does mean that if one is under an obligation to prevent X, one does not fail to comply with it due to a lack of moral concern (but rather due to a lack of alternatives or information).

[32] Also, knowing what many people in fact do won't tell us much about what we should expect from any single individual. Cf. Sher (2009, chap. 7) against any statistical reading of the standard of awareness.

that it's possible to act morally. In the DWK experiments, an additional instruction that B is the fair choice might make a difference all by itself. And if it does, the participant meets the threshold, and is blameless according to the current proposal.

The problem with this proposal is that it renders too many agents blameless. Presumably it renders a large portion of the impurely indifferent agents blameless (given that certain triggers will make them press the button after all). Clearly, though, in the cases I've been considering, they're not blameless, neither in the attributability sense, nor in the accountability sense. In fact, as argued in Sect. 4.2, they are the most accountable of all. Perhaps certain changes in their environment could improve their behaviour, but that doesn't mean we shouldn't expect them to do better.[33]

Hence, the idea of a further threshold may seem appealing at first, though in the end it's unclear whether it can be done. The two proposals just discussed do not work. More importantly, it's not clear that we should want further thresholds when it comes to the kinds of blameworthiness under consideration. As to attributability, there doesn't seem to be a need to set a limit on when people's conduct can speak badly of their will. As to accountability, so long as people have the right sort of capacity of moral concern, there doesn't seem to be a need to set a limit on our expectations that they do better. The Milgram participants who delivered the shocks under protest and stress, then, can be expected to have done better than what they did. And the same would apply to strategically ignorant agents with social and psychological barriers.[34]

6 Conclusion

Are strategic agents blameworthy for the unfair outcomes they realize? As noted in Sect. 2, this question can be answered from various angles: in terms of one's justification to remain ignorant, in terms of the recklessness of one's conduct, or in terms of one's motivations. In the foregoing, I have explored the third route. My account—the Maximal Account—builds upon Arpaly's quality of will account, and has three innovative features. First, it utilizes a suitable concept of moral concern, according to which maximally concerned agents aren't obstructed by such things as psychological or social barriers. Second, it accepts a threshold according to which a certain kind of sacrifice will excuse one from blame, though it rejects further thresholds. As a consequence, it will render many agents blameworthy, if only to a certain degree. Third, it makes a strict and significant distinction between attributability and accountability. In the familiar attributability sense, strategically ignorant agents are blameworthy, but less so than agents who are knowingly egoistic. In the accountability sense, an impor-

[33] The impurely indifferent agents we have been concerned with do care about fairness to some extent, but they also care about money and self-image. In other kinds of cases (discussed by Hellman 2009 and Sarch 2014), these agents might have mixed concerns about the interests of others (such as one's family members, or one's patients), in which case it's not always clear that we should regard them as blameworthy.

[34] What if agents are quite concerned and exert a great deal of effort, but just fail to succeed because they live in a society with heavy social barriers? On the Maximal Account, they're blameworthy to some degree (since they do not act with maximal moral concern, and fulfill all clauses (1)–(8)). But, they might still deserve credit for trying to resist their social barriers. Indeed, my story about blameworthiness might well be compatible with a further story about praiseworthiness.

tant group of strategically ignorant agents are blameworthy to the greatest degree (namely, the impurely indifferent ones), given that they seem to be the only type of agent that could, in a relevant sense, be expected to do better.

In closing, I'd like to return briefly to the alternative views of the responsibility of strategically ignorant agents mentioned in Sect. 2. On the justification view, recall, strategically ignorant agents are blameworthy when their decision to remain ignorant is unjustified, i.e. when the reasons to remain ignorant, such as self-interest, do not outweigh the reasons to investigate, such as fairness. On the recklessness view, strategically ignorant agents are blameworthy when their conduct is reckless, i.e. when they are aware that they run an unjustified risk. This is not the place to offer a full comparison between these other views and my motivation view, but let me just mention one main consideration in favor of the latter. On the existing alternative views, if strategically ignorant agents are considered blameworthy, they are blameworthy *full-stop*. My motivation account, by contrast, offers a much more nuanced story. First, it distinguishes between two kinds of blameworthiness: strategically ignorant agents are blameworthy depending on whether we're talking about attributability or accountability. Second, it adds a gradual dimension: strategically ignorant agents are more or less blameworthy depending on whether their indifference is pure or impure (i.e. whether they're concerned only about self-interest, or whether their concern is more complex, as it often is). So far, alternative views have not accounted for these nuances.[35]

Acknowledgements Thanks to: all of my colleagues, several audiences, and the two referees for comments. This paper has been written in the context of the research project 'We should know better. An inquiry at the crossroads of ethics and epistemology' supported by the Netherlands Organization for Scientific Research.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Alexander, L., Ferzan, K., & Morse, S. (2009). *Crime and culpability. A theory of criminal law*. Cambridge: Cambridge University Press.

Arpaly, N. (2003). *Unprincipled virtue. An inquiry into moral agency*. Oxford: Oxford University Press.

Arpaly, N., & Schroeder, T. A. (2014). *In praise of desire*. Oxford: Oxford University Press.

Bartling, B., Engl, F., & Weber, R. A. (2014). Does willful ignorance deflect punishment? *European Economic Review*, *70*, 512–524.

Björnsson, G. (2016). Explaining (away) the epistemic condition on moral responsibility. In *Responsibility—The epistemic condition*. Oxford: Oxford University Press.

Bradford, G. (2016). Hard to know. In *Responsibility—The epistemic condition*. Oxford: Oxford University Press.

Chatzidakis, A., Hibbert, S., & Smith, A. P. (2007). Why people don't take their concerns about fair trade to the supermarket: The role of neutralisation. *Journal of Business Ethics*, *74*, 89–100.

Clarke, R. (2014). *Omissions. Agency, metaphysics, and responsibility*. Oxford: Oxford University Press.

[35] As always, important issues remain and need to be addressed on a further occasion. One interesting issue concerns the practical upshots of the attributability versus accountability distinction, i.e. whether there's an interesting link between them and the various forward-looking functions of legal punishment (such as deterrence, societal protection, or rehabilitation).

Conrads, J., & Irlenbusch, B. (2013). Strategic ignorance in ultimatum bargaining. *Journal of Economic Behavior & Organization*, *92*, 104–115.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*, 67–80.

Eckhardt, G. M., Belk, R., & Devinney, T. M. (2010). Why don't consumers consume ethically? *Journal of Consumer Behaviour*, *9*, 426–436.

Ehrich, K. R., & Irwin, J. R. (2005). Willful ignorance in the request for product attribute information. *Journal of Marketing Research*, *42*, 266–277.

FitzPatrick, W. J. (2016). Unwitting wrongdoing, reasonable expectations, and blameworthiness. In *Responsibility—The epistemic condition*. Oxford: Oxford University Press.

Gendler, T. S. (2007). Philosophical thought experiments, intuitions, and cognitive equilibrium. *Midwest Studies in Philosophy*, *31*, 68–89.

Grossman, Z., & Van der Weele, J. J. (2015). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*.

Gruber, V., & Schlegelmilch, B. B. (2014). How techniques of neutralization legitimize norm- and attitude-inconsistent consumer behavior. *Journal of Business Ethics*, *121*, 29–45.

Harman, E. (2011). Does moral ignorance exculpate? *Ratio*, *24*, 443–468.

Hellman, D. (2009). Willfully blind for good reason. *Criminal Law & Philosophy*, *3*, 301–316.

Husak, D. N., & Callender, C. A. (1994). Willful ignorance, knowledge, and the 'equal culpability' thesis. *Wisconsin Law Review*, *1*, 29–69.

Kagan, S. (2011). Do I make a difference? *Philosophy & Public Affairs*, *39*, 105–141.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*, 495–524.

Lawford-Smith, H. (2015). Unethical consumption and obligations to signal. *Ethics & International Affairs*, *29*, 315–330.

Luban, D. (1999). Contrived ignorance. *Georgetown Law Journal*, *87*, 957–80.

Markovits, J. (2012). Saints, heroes, sages, and villains. *Philosophical Studies*, *158*, 289–311.

Mason, E. (2015). Moral ignorance and blameworthiness. *Philosophical Studies*, *172*, 3037–3057.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal & Social Psychology*, *67*, 371–378.

Moody-Adams, M. M. (1994). Culture, responsibility, and affected ignorance. *Ethics*, *104*, 291–309.

Nelkin, D. K. (2016). Difficulty and degrees of moral praiseworthiness and blameworthiness. *Noûs*, *50*, 356–378.

Onwezen, M. C., & Van der Weele, C. N. (2016). When indifference is ambivalence: Strategic ignorance about meat consumption. *Food Quality & Preference*, *52*, 96–105.

Paharia, N., Vohs, K. D., & Deshpandé, R. (2013). Sweatshop labor is wrong unless the shoes are cute: Cognition can both help and hurt moral motivated reasoning. *Organizational Behavior & Human Decision Processes*, *121*, 81–88.

Pleasants, N. (2011). Moral argument is not enough: The persistence of slavery and the emergence of abolition. *Philosophical Topics*, *38*, 139–160.

Robbins, I. P. (1990). The ostrich instruction: Deliberate ignorance as a criminal mens rea. *Journal of Criminal Law & Criminology*, *81*, 191–234.

Sarch, A. F. (2014). Willful ignorance, culpability and the criminal law. *St. John's Law Review*, *88*, 1023–1101.

Sher, G. (2009). *Who knew? Responsibility without awareness*. Oxford: Oxford University Press.

Smith, H. M. (2014). The subjective moral duty to inform oneself before acting. *Ethics*, *125*, 11–38.

Spiekermann, K. (2016). Four types of moral wriggle room: Uncovering mechanisms of racial discrimination. In M. Fricker & M. Brady (Eds.), *The epistemic life of groups*. Oxford: Oxford University Press.

Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, *48*, 187–211.

Talbert, M. (2016). Akrasia, awareness, and blameworthiness. In *Responsibility—The epistemic condition*. Oxford: Oxford University Press.

Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, *24*, 227–248.

Wolf, S. (1987). Sanity and the metaphysics of responsibility. In F. D. Schoeman (Ed.), *Responsibility, character, and the emotions* (pp. 46–62). Cambridge: Cambridge University Press.