

Part I

CONCEPTS

## EXPLOITATION

Although we frequently claim that some act, practice, or transaction is exploitative, the concept of exploitation is typically invoked without much analysis or argument . . . exploitation is often utilized as if its meaning and moral force were self-evident. They are not.<sup>1</sup>

Why do we need an analysis of the concept of exploitation? One reason, a general one, is that although exploitation is a heavily used and widely accepted moral concept, moral philosophers have paid relatively little attention to it. As Wertheimer notes,

despite the frequency and ease with which we make exploitation claims in ordinary moral and political discourse, I think it fair to say that with the major (and I do mean major) exception of the Marxist tradition, exploitation has not been a central concern for contemporary political and moral philosophy.<sup>2</sup>

A second reason for looking at exploitation, one with special relevance to this project, is that asserting that 'markets' are somehow exploitative is perhaps the commonest way of arguing against them. Many commercial practices – including organ sale, prostitution, and paid surrogacy – have been the object of attacks couched in terms of exploitation. So in order to understand the case against such practices, we need to know what exploitation claims mean. As we'll see shortly, though, saying exactly what exploitation is isn't easy. For it's a complex and contested concept and the subject of several competing theories.

### 2.1 Introducing exploitation

To exploit something, in the most general sense, is simply to put it to use, not waste it, take advantage of it.<sup>3</sup>

In the broadest sense, to exploit something, e.g. a natural

resource, means to use it for a purpose. Such exploitation is morally neutral.<sup>4</sup>

The word 'exploitation', like the related expressions 'use' and 'take advantage of', is sometimes used *evaluatively* or *normatively* and sometimes not; it has both *moral* and *non-moral* senses.<sup>5</sup> Hence, as Schwartz reminds us (above), we can talk entirely neutrally of exploiting (or of using) a resource, or an opportunity, or a talent, and when we speak of exploitation in this way we do not thereby express moral disapproval.<sup>6</sup> But, at the same time, as Onora O'Neill suggests:<sup>7</sup>

Few moral criticisms strike deeper than the allegation that somebody has used another.<sup>8</sup>

On the other hand, Wertheimer (drawing on Williams) plausibly suggests that exploitation is a *thick moral concept* – one 'such as *treachery* and *promise* and *brutality* and *courage*, which seem to express a union of fact and value'.<sup>9</sup>

Sometimes, it's hard to tell whether 'exploitation' is being used normatively or not, and at times people deliberately take advantage of this ambiguity. When people do this they rarely lay out their arguments in formal terms. However, we can see how they might trade on this ambiguity by thinking about the following schematic example:

- 1 'Factory farming' involves the commercial exploitation of animals.
- 2 Exploitation is a bad thing.
- 3 Therefore 'factory farming' is a bad thing.

In (1), 'exploitation' is used in its non-moral sense. Hence, (1) appears (and is) obviously true. (If 'exploitation' were used in its moral sense in (1), then (1) might still be true, but it wouldn't be *obvious*, and we'd have to argue for it independently.) In (2), 'exploitation' is used in its moral sense and so (2) is supposed to be an obvious *a priori* truth. (3) The conclusion then *appears* to follow – but doesn't really – from (1) and (2). In order for (3) to follow from (1) and (2) we'd need 'exploitation' to be used in the same sense throughout. But if we use it in its moral sense throughout then (1) loses its obviousness – and, if we use it in its non-moral sense throughout, (2) becomes false, since exploitation in the non-moral sense clearly isn't always bad. So, either way, the argument is flawed. As I've said, people who take advantage of this ambiguity hardly ever present their arguments in such a formal way as this. However, what I've just said does capture the underlying structure of what they're doing (sometimes intentionally, sometimes not).

Except for these equivocal cases, the non-moral use of 'exploitation' is

irrelevant to our present concerns and from now on I'll concentrate just on its moral sense (or senses).<sup>10</sup> We can usually get a good (though not infallible) idea of whether 'exploitation' is being used normatively or not by seeing what it's being applied to. For example, when it's applied to inanimate objects, it's normally being used non-morally, but when it's applied to persons, it implies moral criticism.<sup>11</sup> Thinking along these lines, Robert Goodin even goes as far as to say that 'an act of exploiting a person always constitutes a wrong'.<sup>12</sup> And, as Jonathan Wolff points out, similar things can be said about the word 'use':

There can be no doubt that the idea of exploitation is closely related to some notion of use ... exploiting someone is using someone. Generally, when we say 'x used y', where y is a person, we intend this to be a form of criticism of x's behaviour.<sup>13</sup>

One way of dealing with this is to say not that 'exploitation' has a moral and a non-moral sense, but that it has just one sense (it means *use*) accompanied by different 'connotations' in direct contexts, this difference being explained by the fact that while using persons is wrong, using inanimate objects is not.<sup>14</sup> However, things are not that simple, since:

using other *people* may be OK, as students use teachers to learn. But it is odd to describe such cases as exploitation. Among people, the word implies *merely* using others, taking *wrongful* advantage for one's own purposes.<sup>15</sup>

So the point is not just that using people is wrong, while using other things is OK. For there is (normally) nothing wrong with using people 'as students use teachers', or as patients use their doctors, or as children use their parents. Rather, the concept of exploitation, when applied to persons, is part of an attempt to pick out those uses of persons which are morally wrong, distinguishing them from those which are innocent. Hence, while it's true that 'exploit' means *use* when applied to persons, there is much more to be said. 'Exploitation' means something more specific, *wrongful use*. In fact, it means something even more specific than that since, as Wolff points out, not all 'improper uses' of persons are exploitative:

'exploits' as a moral criticism seems very close to 'improperly uses', although there are some improper uses of people – for example as a punch bag – which would not normally be thought of as forms of exploitation.<sup>16</sup>

We can start our analysis of exploitation, then, by schematically defining

'exploiting a person' as *wrongfully using her in one or more of a limited set of ways*. The main aim of sections 2.2–4 is to say what these ways are and why they constitute wrongdoing, but, before proceeding with this task, we must make three further clarificatory points.

The first is a concern that what I have said so far might be taken to imply that only persons – and not, for example, animals, or nature – can be exploited (in the moral sense).<sup>17</sup> Given certain very widely held and plausible assumptions about persons, it's particularly easy to see how the idea of exploitation applies to them; the concepts 'exploitation' and 'personhood' are, we might go as far as to say, made for each other. That said, there is no reason in principle why exploitation shouldn't be extended (and in its normative sense) to animals, or perhaps even to nature or to works of art. In each case, whether one thinks that the thing in question can be exploited will depend on what one thinks of its moral (or other evaluative) status. For example, someone who regards horses as quasi-persons will think that horses can be exploited just as humans can be. Or someone who thinks that nature is 'sacred' may think that it can be wrongfully used in exploitative ways. Nothing that I have said so far rules these possibilities in or out, since everything depends on what one thinks about the moral status of horses, nature, etc. My own focus, however, will be almost entirely on the exploitation of persons.

The second clarificatory point concerns my suggestion (above) that wrongfulness should be built into the definition of 'exploitation'. This might, understandably, be taken as an endorsement of *absolutism* about exploitation: the view that to exploit is always (and necessarily) wrong. The definition I propose is not, however, meant to be absolutist. Rather, my 'weaker' view is that exploiting persons (in the moral sense) is wrong *other things being equal* or, more specifically, *in the absence of other (sufficiently strong) countervailing moral considerations*. Take, for example, a case in which we have to choose between:

- 1 exploiting an individual (without seriously harming her) and in so doing saving the world from a nuclear catastrophe which would kill billions; and
- 2 not exploiting the same individual, thereby not averting a nuclear catastrophe in which billions die.

What ought we to do? Obviously, we would like to know more about the case before making a final decision. But it nonetheless seems clear that most of us have a strong preliminary intuition in favour of (1). In other words, we are not absolutists about exploitation, even if we think that the exploitation involved in (1) is regrettable and *prima facie* wrong. This intuition is compatible with what I want to say about exploitation, that it is wrong *in the absence of other (sufficiently strong) countervailing moral*

*considerations*. For here we obviously have countervailing moral considerations: the fact that if we don't exploit, billions will die. If those considerations were absent, if there were no nuclear threat, then we probably *would* want to say that we shouldn't exploit. We should bear in mind throughout, then, that the fact that there is a genuine or valid exploitation objection to a practice doesn't necessarily mean that that practice is wrong 'all things considered'. For there may be countervailing moral considerations which are more important or 'weighty' than the fact that there's exploitation. *A fortiori*, that a practice is exploitative isn't always a sufficient reason for banning it, since there can be sound policy reasons for permitting the continuation of exploitation.

The third and final clarificatory point is that exploitation need not be intentional, although whether it's intentional or not may impact on our views about the exploiter's blameworthiness and character. Wertheimer makes the point well:

A may exploit B even though A believes the terms of the transaction are eminently fair – if they are not. And A may exploit B while (and perhaps even because) A is unaware of the effects of the transaction on B. Free riders frequently do not understand that they are riding free. Of course, as with other forms of wrongdoing, we can distinguish between the wrongness of the act and the agent's culpability for that act. We may regard intentional exploitation as worse than negligent exploitation, and there may even be completely non-culpable exploitation.<sup>18</sup>

Thus, even if it were true that *all* capitalists and husbands (for structural reasons) exploit their workers and wives (respectively), it needn't be true that they all *intend* to do so. For some of them (naively perhaps) sincerely believe that they are benefiting and/or treating their workers and wives fairly.

## 2.2 Wrongful use and disparity of value

Within the moral sense of 'exploitation', John Harris usefully distinguishes between 'two differing conceptions'. The first of these is:

the idea of *wrongful* use and may occur when there are no financial or commercial dimensions to the transaction. A classic case here would be where it is claimed that lovers may exploit one another, that is, use one another in some wrongful way. The most familiar of such wrongful ways in this context might be where it is claimed that one partner uses the other or treats the other merely as a 'sex object'.<sup>19</sup>

While the second conception

involves the idea of some disparity in the value of an exchange of goods and services.<sup>20</sup>

Being underpaid for one's work by an unscrupulous boss is a classic example of exploitation in this second sense, as is being made to pay exorbitant prices for scarce goods by profiteers during emergencies, such as in wartime. In each case, there is a disparity of value; you are *underpaid* in one case and have to *overpay* in the other.

Sometimes both kinds of exploitation are present in the same situation. Prostitution is a good example of this. Most street prostitution is believed to be exploitative in the 'disparity of value' sense for reasons which include the fact that the women involved don't receive, or don't keep, enough money. But many people also argue that, even if there were no problems with underpayment, etc., prostitution would still be exploitative in that it involves the wrongful use of women (or others) as 'sex objects'.<sup>21</sup> Goodin, for example, says that:

we would certainly want to say that prostitutes are exploited by their clients and pimps, *however well the prostitutes might be paid*.<sup>22</sup>

The two kinds of exploitation can, however, also function independently of one another. Consider, for example, Harris's example of the lover who says that she is treated as a sex object. Her complaint clearly appeals to wrongful use exploitation and, equally clearly, does *not* appeal to 'disparity of value' exploitation. For she is hardly likely to be impressed if her lover attempts to deal with the sex object complaint by offering to pay her for sex. Conversely, if a philosophy professor complains that she is underpaid for her philosophical labours by an *exploitative* university administration, she is (rightly or wrongly) complaining *only* about 'disparity of value' exploitation. For the chances of her thinking that employing people to teach philosophy and to do philosophical research is a *wrongful use* of them (no matter what the pay) are pretty remote.

The word 'exploitation', then, appears to be used in three different ways. We have first to distinguish between its moral and non-moral occurrences and then, within the moral ones, between 'wrongful use' and 'disparity of value' claims. How should we respond to the fact that ordinary moral thought contains these different conceptions of exploitation?

One response, a highly revisionary one, is to insist that we choose between the conceptions, providing philosophical arguments to show why one is better and adequate on its own and why the other should be abandoned. Sensat, for example, distinguishes between two frameworks

for theories of exploitation. One of these, corresponding to the 'disparity of value' conception, views exploitation as 'basically a misdistribution of the benefits of social cooperation', while the other, corresponding to the wrongful use conception, views it as 'a certain kind of use of what is exploited, namely a use which is contrary to its nature'.<sup>23</sup> He then proceeds to argue against the former, which, he says, is 'fundamentally flawed'; for exploitation is *really* all about wrongful use, and those who think otherwise are mistaken.<sup>24</sup>

A second approach is to produce a 'higher-order' and/or more general theory of exploitation which somehow incorporates both conceptions. An example of this can be found in Jonathan Wolff's paper 'Marx and exploitation', where he offers us the following 'general analysis of exploitation'.<sup>25</sup>

to be an exploiter is to use another's circumstances to obtain their actual compliance with a situation without having sufficient regard to whether that situation violates fairness, flourishing, or suffering norms. To be exploited is to be treated in this way.<sup>26</sup>

Wolff's analysis (and this is not meant as a criticism) is what we might call *enablingly vague*. The precise content of 'fairness, flourishing, or suffering norms' is unspecified, and a variety of different philosophical theories about such norms can be plugged into Wolff's general definition, yielding a more substantive account. As regards wrongful use and disparity of value, it's easy to see how Wolff's definition could incorporate both. For wrongful use is (on most views) likely to violate at least one of our 'fairness, flourishing, or suffering norms', while 'disparity of value' will, at least in the standard case, be unfair and may violate the other norms too.

The third and final response is to regard the two 'exploitations' as entirely distinct moral phenomena, which just happen, by an accident of history, to be attached to the same English word. On this view, the word 'exploitation' is rather like the word 'bank', which can refer to two unrelated things: a financial institution, or the side of a small hill or mound. Another example, closer to home, is the word 'right', which (as well as its non-moral sense, *the opposite of left*) means two different things within ethics. There is an *action being right*, the right or correct thing to do, as in 'it was right for her to save that drowning child'. And then there is *someone having a right*, which here means an entitlement to something, as in 'I have a right to life.'

My aim here is not to provide a complete philosophical theory of exploitation but rather (following Wood)

to clarify the concept ... by exploring what I think people mean when they object to behavior or social arrangements as

exploitative, and identifying the moral convictions which give such objections their force.<sup>27</sup>

In so doing, I hope to improve our understanding of how, and whether, exploitation objections to the commercialisation of the body work. Given this, my general strategy in what follows will be to remain as neutral as possible with respect to the theoretical options outlined above. I'll proceed *as if* there are two distinct types of exploitation, corresponding to Harris's 'two conceptions'. This may or may not, ultimately, be the correct theory of exploitation. It does seem to me, though, that each 'conception' (rightly or wrongly) has an independent role in our moral thought and that each contains, or can be used to articulate, an important (and independent) ethical insight. Since the aims here are more applied than theoretical, these facts are enough to justify the structure of what follows – my offering two semi-independent accounts of the two 'exploitations'. I should add, though, that it seems unlikely that the two 'exploitations' are completely unrelated things which just happen to be referred to by the same word, since, as we shall see, they do share a number of important features.

### 2.3 Benefit

Many people seem to think that exploitation (at least in the 'pejorative' sense) has to be unfair or unjust, and that the injustice consists in a redistribution of harms and benefits, with the benefits flowing from the exploited to the exploiter.<sup>28</sup>

This section aims to explain the connection between benefit and exploitation. In particular, we're going to take a look at the idea that exploiters necessarily derive benefit from their exploitative behaviour. Versions of this position are endorsed by, amongst others, Feinberg and Wertheimer.<sup>29</sup>

Common to *all* exploitation of one person (B) by another (A) . . . is that *A makes a profit or gain* by turning some characteristic of B to his own advantage.<sup>30</sup>

A cannot exploit or take advantage of B unless A obtains some advantage through the transaction with B.<sup>31</sup>

This view, which states a (putative) necessary condition for the occurrence of exploitation, I term the *benefit condition*.

To some people, it seems just obvious that exploiters are getting something, or gaining some kind of advantage, when they exploit. Why else

would they do it? If, however, an argument is needed, perhaps the best is that the benefit condition is an effective way of distinguishing exploitation from abuse (as well as from certain other forms of wrongdoing, including discrimination and oppression). This appears to be one thing that moves Wertheimer to endorse a version of the benefit condition.<sup>32</sup>

It has been alleged, for example, that medical students are frequently abused by oral insults and denigration and that this abuse may leave 'long-lasting emotional scars'. By contrast, it is also sometimes claimed that medical interns are exploited, that they work long hours for low pay. The contrast is just right. There is no reason to think that anyone gains – at least in any normal sense – from abuse, but it is at least plausible to think that the hospitals or patients gain from the exploitation of interns.<sup>33</sup>

Wertheimer's version of the benefit condition is what we might call robustly objective. For him, in order for A to exploit B, A must derive actual benefit:

A does not actually exploit B unless A gains from the interaction, even if A *seeks* to gain from the transaction. We can distinguish between the claim that A *exploits* B and the claim that A *acts exploitatively* towards B.<sup>34</sup>

So, for Wertheimer, people who attempt exploitatively to derive benefit from others, but fail to derive the benefit envisaged, don't count as having exploited at all, although they have 'acted exploitatively'. Someone who rejects this 'objective' benefit condition is Sensat:

Exploitation . . . need not further exploiters' real interests at all. Indeed, if exploitation involves something like Kant's treating persons merely as means, then exploitation may be a type of heteronomy and thus *never* in exploiters' highest-order interests.<sup>35</sup>

In support of this, he offers us the example of

the man who, relying on certain power advantages, gets a woman to engage in public displays of deference and submissiveness toward him in order to boost his status as defined by the standards of a misogynous culture. While the woman's behavior is degrading to her and violates rather than realizes her nature, the man's use of the woman does nothing to further his real interests. Though it does foster the achievement of a goal which he has set for himself.<sup>36</sup>

Sensat's remarks show that we should reject the benefit condition, or rather the 'objective' version of it. The objective version says (implausibly) that exploiters always actually derive benefit. But a more plausible version would be subjective, appealing not to the exchange of benefits, but to the exploiter's aims and intentions: the idea being that exploiters always *aim* to derive benefit from their exploitative behaviour.

Although exploitation need not be intentional (i.e. not all exploiters mean to exploit), there may nonetheless be necessary conditions for the occurrence of exploitation which cite the exploiter's intentions. The subjective benefit condition is one such (putative) necessary condition, saying as it does that in order to exploit, one must intend to derive benefit. Exploiters (or attempted exploiters) can fail to be successful in at least two different ways. First, they can simply not get the exploitee to do, or give them, what they want. At least some of these cases will be cases of merely attempting to exploit, or what Wertheimer calls 'acting exploitatively' (without actually exploiting). Second, they can be mistaken about the value of what they demand and so fail to derive any real benefit from it, even if they get what it was that they originally wanted (under a non-evaluative description). Sensat's case is a failure of the second kind and shows very well why the subjective benefit condition is to be preferred to an objective version. The reason is that it would be more than a little odd if we said of someone, such as Sensat's misogynist, that he wasn't *really* an exploiter because the thing that he was trying to extract from the exploitee wasn't *really* of any value. Sensat, then, has provided us with a good counter-example to the objective benefit condition, and shown that it is rather a subjective benefit condition that we should be considering.<sup>37</sup>

However, even the subjective benefit condition faces problems of its own. The most obvious is that there could, it seems, be cases in which A exploits B not in order to benefit herself, but solely in order to affect (either positively or negatively) a third party, C. A good example of this is biomedical research on human subjects. This can, of course, be exploitative under certain conditions, but it may sometimes be motivated not by self-interest, but by the researcher's desire to do good – for example, finding a cure for AIDS, cancer, or heart disease. Alternatively, consider the following hypothetical:

Anne, Bob, and Carole

Anne employs Bob (for whom she has little regard) under highly exploitative conditions. She does this for one reason only: in order to pay for a lifesaving transplant operation for her sister Carole. Anne derives no material benefits from the arrangement. On the contrary, she (like Bob) works very hard for a minimal wage, saving all the profits for Carole's operation. Furthermore, she doesn't believe that she

will derive any emotional or psychological benefits from prolonging her sister's life, since relations between them are at best 'frosty'. Rather, Anne is doing this entirely 'out of a sense of duty'.

One possible response to Anne's behaviour is to say that she must believe that she will derive some kind of benefit, even if it's just the absence of guilt feelings, otherwise she wouldn't do what she's doing. However, this won't work against hypotheticals (in which the fact that she doesn't believe that she stands to benefit can just be *stipulated*) unless attached to the (implausible) view that people only act in ways which they at some level think will benefit them.

Another response, suggested by Feinberg, is to broaden the notion of benefit ('gain' in his terms) in such a way that Anne's behaviour counts as benefiting her:

A may exploit B for great 'gain' all of which he then gives to charity. Clearly, to accommodate this example we must dilute the sense of 'gain' so that it includes gain either for oneself or for some person or cause that one chooses to benefit. In order to preserve the gain requirement, in short, we must employ an admittedly extended sense of 'gain' including both gain in the strict sense and fulfilment of one's aims, purposes, or desires, including altruistic and conscientious ones.<sup>38</sup>

Wertheimer makes a similar move, introducing the category of *mediated exploitation*:

Although the paradigmatic exploitation claims involve cases in which A seeks to promote his or her *self-interest*, we need a more protean conception of what counts as a benefit to A, one that includes A's purposes, goals, and values. In some cases, for example, A exploits B *on behalf* of C. Call this *mediated exploitation*. Consider the Legal Aid lawyer who advises her client to reject a generous settlement because she wants to litigate a new legal doctrine that, if litigation is successful, will serve the interests of a larger group. The lawyer may have exploited her client on behalf of others, but she has exploited her client nonetheless.<sup>39</sup>

Of course, one *could* choose to use the word 'benefit' in the ways suggested by Feinberg and Wertheimer, thereby rendering a version of the benefit condition consistent with the Anne, Bob, and Carole case (which, in Wertheimer's terms, looks like an instance of mediated exploitation). However, there are good reasons not to.

The first is that saying that Anne benefits from her actions is highly counter-intuitive and appears inconsistent with our ordinary usage of the word 'benefit'. Indeed, what we would probably want to say of Anne's actions (and of many other cases of charitably exploiting) is that they are acts of *self-sacrifice*, rather than *self-benefit*. Now this is not, of course, decisive. For our 'folk' use of the word 'benefit' might involve a mistake or an inconsistency, or be otherwise conceptually sub-optimal. But there should at least be a presumption in favour of ordinary usage, one which can be overturned only by substantive argument.

The second reason, one internal to our interest in producing an account of exploitation, is that if we take benefit to include fulfilment of any of 'one's aims, purposes, or desires', then the benefit condition doesn't really add anything to the idea that exploitation is use. It's obvious that exploitation necessarily involves use, and obvious too that this, in turn, involves the attempted satisfaction of the exploiter's aims, purposes, or desires. Now the benefit condition may or may not be true. But what it's supposed to do is to improve our account of exploitation by *adding an additional constraint*, such that not all A's usings of B, but only the ones where A tries to derive benefit from using B, count as exploitation. Given this, Feinberg's suggested broadening of the idea of benefit totally 'emasculates' the benefit condition, giving it no content over and above the idea of use. So, at least for our purposes, we may as well not bother even to discuss Feinberg's 'weak' benefit condition, for it adds nothing to the central insight that exploitation necessarily involves use.

The third and final reason is that some 'aims, purposes, or desires' are clearly self-destructive, so on Feinberg's extended understanding of 'benefit', I count as having benefited myself even when the desires I've satisfied are wholly (and self-consciously) self-destructive ones. What this shows, it seems, is that the broad understanding of benefit should be rejected on general grounds because it trivialises and renders redundant the idea of benefit, collapsing it into the idea of desire satisfaction.

A better response to Anne, Bob, and Carole is to abandon the benefit condition (in both its objective and subjective forms). For, really, there is in principle no reason why A could not exploit B for reasons which have nothing at all to do with A's welfare. And there could also be cases in which A exploits B in order to *harm*, rather than benefit, herself.<sup>40</sup> Consider, for example, the following:

The Slave Owner, having killed her father during a heated argument, is filled with self-loathing and guilt. She decides (calmly and rationally) that she wants to be mutilated as a penance. She sends The Slave to purchase the necessary equipment and orders her to carry out the mutilation. Although The Slave is not exactly keen on The Owner, she

nonetheless has moral qualms about mutilating her and also finds the whole idea physically repulsive. However, The Owner insists that The Slave must go through with it, threatening to have her and her family tortured if she doesn't do what she wants.

Presumably we *do* want to say that The Slave is exploited here because, in most respects, the situation is just like any other case of forced labour. The Slave is coerced and made to do what The Owner wants – which, here, just happens to be mutilation. However, it's also clear that The Owner's goals are not (or, at least, need not be) self-beneficial. Hence, we have another counter-example to the benefit condition. (A more realistic case like the above might be where someone – again for reasons of 'self-loathing', rather than pleasure – hires a prostitute, under exploitative conditions, to perform harmful 'sadistic' acts.)

What we should say then is not that exploiters necessarily use others to benefit themselves, but that *they necessarily use others to foster the achievement of their own goals*: goals which may or may not be selfish and may or may not coincide with the furtherance of their real interests.<sup>41</sup> In reality, of course, most of the goals that people have are self-interested ones. Hence, exploiters do normally use others to benefit themselves, which is why the benefit condition strikes us as initially plausible. But attempting to benefit oneself is not a necessary part of being an exploiter. This proposal, which we might term the *use condition*, is in substance more or less the same as Feinberg's 'expanded' benefit condition, but (for reasons given earlier) I prefer to drop talk of benefit, instead of expanding the notion of benefit.

## 2.4 Use

There are various things about people which we exploit. We might exploit their strong backs or weak minds. We might exploit their fears, ignorance, superstitions, gullibility, or naiveté. We might exploit their generosity, loyalty or trust. We might exploit their bad luck, their joblessness, homelessness or illness. All those things, however, are merely *attributes* of people and their circumstances. Exploiting a person's attributes is not the same as exploiting a person. Each of us can sometimes exploit some of our *own* attributes – a runner her stamina, a boxer his reach, etc. We would not thereby be exploiting ourselves; we would just be exploiting certain *things about* ourselves.<sup>42</sup>

This section aims to address two questions. The first is simply: what exactly is use? More specifically, what exactly is the kind of use involved in exploitation? The second relates to the distinction drawn by Goodin



above: how does exploiting or using *a person* relate to exploiting or using *a person's attributes*? It won't be possible to answer completely such questions here. However, it will be possible at least to provide a general framework, and to rule certain possibilities out and others in.

One initial linguistic complication and, I suspect, the source of more than a little confusion, is the fact that (as we have already seen) 'use', like 'exploitation', can be used as a moral term. Hence, if B is *exploited* by A, in the moral sense, then we probably also want to say that B is *used* by A, again in the moral sense. Bearing this in mind, the first thing we need to say about use is that 'use' as it occurs in the use condition is *not* a moral term. Rather, for A to use B (in the sense intended) is simply for A to bring it about that that B or, more likely, some part or aspect or behaviour of B, makes a causal or constitutive contribution to the achievement of A's relevant goals (where 'goals' is meant as a general term covering aims, desires, wants, etc). So the use condition (when it's satisfied) is normally satisfied by there being use of one or more of the exploitee's attributes, rather than use of the exploitee *per se*.

As an answer to our questions, though, this is at best schematic. Can anything more be said? One substantive answer is suggested by Wood, who offers a very plausible model of the kind of use which normally occurs when one person exploits another. First, he distinguishes two aspects of the 'standard' exploitative relationship: *advantage-exploitation* and *benefit-exploitation*. We *advantage-exploit* when we 'exploit someone's weakness or vulnerability, which gives us a hold over the person'.<sup>43</sup> (So we might say that to advantage-exploit is to 'take advantage'.) The vulnerabilities which are the objects of advantage-exploitation are often desires and needs. Hence, we can say, for example, that the starving worker's *need* for food and the young 'wannabe' footballer's passionate *desire* to be a soccer star are advantage-exploited by their employers; or that the economically dependent wife's *need* for money and a home is advantage-exploited by her husband; or that the terminally ill patient's desperate *desire* for a 'miracle cure' is advantage-exploited by the biomedical researchers who recruit her into their clinical trial. However, the idea of vulnerability is wider than that of desire/need, and, as Wolff points out:

it is easier to give a list of causes of vulnerability than to explain its nature. Typically you are vulnerable if (other things being equal) you are poorer, more ignorant, less intelligent, less cunning, or less ruthless than another, or have some other bargaining weakness with respect to them.<sup>44</sup>

The relationship between advantage-exploitation and the 'full' exploitation of persons is, it seems to me, as follows. One can be advantage-

exploited without being exploited (in the moral sense): i.e. advantage-exploitation is not sufficient for exploitation. But one can't be exploited, or at least one can't be the object of disparity of value exploitation, without being advantage-exploited: i.e. advantage-exploitation is necessary for disparity of value exploitation.

The reason why advantage-exploitation isn't sufficient for exploitation is simply that there are entirely innocent, or even beneficent, cases in which one or more of a person's vulnerabilities are advantage-exploited. For example, a friend might advantage-exploit my gullibility in order to give me a surprise party or an unexpected gift which she knows I will enjoy – and enjoy all the more for having been 'taken in'.<sup>45</sup> It would be bizarre to describe such cases as exploitation, although, interestingly, similar cases could be exploitation (or attempted exploitation) if my friend's aims were different: for example, if she advantage-exploited my gullibility by playing a practical joke on me in order to impress third parties.

The reason why advantage-exploitation is a necessary part of disparity of value exploitation is that this form of exploitation normally involves getting the exploitee to *agree to a 'bad deal'* by advantage-exploiting her personal or situational weaknesses. Indeed, Wood is right when he says that without advantage-exploitation, there wouldn't be the required 'control or manipulation of the object of exploitation'. For people (at least if they are fully informed, rational, etc.) are hardly likely to agree to 'bad deals' unless they are advantage-exploited (or otherwise pushed or tricked) into doing so.

Turning now to benefit-exploitation, we benefit-exploit when we take advantage of 'some attribute of the person from which we derive benefit or use to achieve our end'.<sup>46</sup> So to benefit-exploit is to derive benefit or otherwise make use of. Hence, we can say of the earlier examples that the employers and husbands benefit-exploit the capacity of their workers and wives (respectively) to provide productive labour; or that the biomedical researchers benefit-exploit some useful aspect of their research subjects' bodies.

Like advantage-exploitation, benefit-exploitation is not *sufficient* for the 'full' exploitation of a person. If it were, then exploitation would occur whenever anyone was useful to anyone else (for example, in employment or friendship) and accusations of exploitation would hardly be serious moral criticisms – if indeed they were criticisms at all – given that exploitation would be endemic. It does, however, seem that benefit-exploitation is *necessary* for the 'full' exploitation of a person, since this is entailed by (what I at the end of the previous section termed) the use condition. For the use condition says that (successful) exploiters necessarily use exploitees as means of achieving their goals. And exploitees can contribute to the achievement of their exploiters' goals in this way only if

they have one or more usable attributes which are benefit-exploited. Hence, in order for a person to be exploited, one or more of her attributes must be benefit-exploited. (Note that this is compatible with my earlier rejection of the benefit condition, since Wood's idea of benefit-exploitation appeals not only to benefit, but to 'benefit or use'.)

Wood, then, is right to describe the relationship between the two aspects of exploitation as follows:

As their names are meant to imply, advantage-exploitation and benefit-exploitation constitute a complementary pair, and advantage-exploitation is the foundation of benefit-exploitation. Without benefit-exploitation there would be no *use* of the object of exploitation, while without advantage-exploitation there would be no control or manipulation of the object of exploitation.<sup>47</sup>

So in order to be exploitable (at least in the disparity of value sense) one must be both useful and vulnerable (which presumably everyone is to differing extents). Someone who is entirely useless can't be exploited because she can't contribute to the achievement of the exploiter's goals, while someone who is totally invulnerable can't be exploited because she can't be made to agree to an exploitative deal by having her weaknesses advantage-exploited. This is a helpful insight, since it provides a structure for thinking about why certain people are more exploitable than others. The most exploitable people are those who are *both* very useful *and* very vulnerable. A young woman who was both sexually attractive and poor would be an obvious example. Her sexual attractiveness makes her 'useful' – both directly, as a means of satisfying men's sexual desires, and indirectly, as a way of making money for third parties – while her poverty makes her vulnerable. So her level of exploitability is high. In contrast, people who are equally useful but *not* vulnerable, such as workers with desired occupational skills that are in short supply, are not very exploitable. On the contrary, they tend to do well in market economies, since they can enter into non-exploitative contracts to sell their useful attributes, or the products of those attributes. Indeed, workers like these may well be in a position to exploit employers who need their services. What about people who are vulnerable but not useful? Such people are not exploitable, but only because they are in the unfortunate position of not having anything to exploit. Even worse, given their vulnerability, if they were to acquire any useful attributes, they could only do so at the cost of becoming the potential victims of exploitation.

At several points, I've qualified claims about exploitation by saying that they apply only to disparity of value exploitation. With this in mind, I'd like to end this section by discussing briefly how the preceding

discussion of use (in particular, of advantage-exploitation and benefit-exploitation) relates to wrongful use exploitation.

As I've already suggested, benefit-exploitation is required for both types of exploitation. So the first thing to note is that in order for there to be wrongful use exploitation, there must be benefit-exploitation. Where disparity of value and wrongful use differ is in their relationship to advantage-exploitation. I suggested earlier that in order for there to be disparity of value exploitation, the exploitee must agree (or acquiesce) to a bad deal and that the reason why she agrees to it is that she is advantage-exploited: i.e. a weakness or vulnerability is used to make her agree or acquiesce. However, these considerations don't apply to wrongful use exploitation, because this form of exploitation doesn't require the exploitee to agree to a 'bad deal': *both* because she needn't agree at all *and* because, even where there is agreement, the 'deal' agreed to need not be bad (or, at least, need not be bad considered apart from the intrinsic badness of wrongful use itself). This point will be discussed in more detail in Chapter 5. For the present, it will suffice to think again about wrongful use objections to prostitution and, more generally, to treating other persons as sex objects. In the case of prostitution, the whole point of the wrongful use objection is that it's meant to apply *even if* the prostitute consents and *even if* she gets a 'good deal' (financially, etc.). This shows that, where there is agreement, the 'deal' agreed to need not be bad. And, to think more generally about 'sex object objections', it should be obvious that it's possible for A to treat B as a sex object (and thus arguably subject B to wrongful use exploitation) without B having any say in the matter, without B's agreeing to anything at all. Pornography, rape, sexual abuse, and voyeurism could all be examples of this.

The two different types of exploitation, then, differ not only in their relationship with *advantage-exploitation*, but also in their relationship with *consent*. The question of how exactly exploitation relates to consent is one of the most interesting issues within exploitation theory and will be explored further in subsequent chapters.

## 2.5 Summary and conclusions

This chapter has introduced and provided the first moves in an analysis of the concept of exploitation. Such an analysis is required because philosophers have paid relatively little attention to defining exploitation, because exploitation is a heavily used and widely accepted idea, and because exploitation arguments have special relevance to the principal concerns of this book. Indeed, I'd go as far as to say that exploitation is perhaps *the* single most important moral concept in the body commodification debate.

The main findings of this chapter's investigations are as follows:

- 1 The word 'exploitation', like the associated expressions 'use' and 'take advantage of', has both a moral and a non-moral sense. Our interest is almost solely in the first of these.
- 2 Behind the moral sense of 'exploitation' lie two distinct worries: one about *disparity of value*, the other about *wrongful use*. These can usefully be thought of as two different kinds of exploitation.
- 3 Although not totally unrelated, these two sorts of exploitation are importantly different in a number of respects: for example, the ways in which they relate to consent and to taking advantage of people's vulnerabilities (what Wood calls *advantage-exploitation*). When assessing exploitation objections, therefore, we must be sure to disambiguate 'exploitation' and to establish which of the two types is being referred to.
- 4 Exploiters normally derive benefit from exploitation, but don't necessarily do so, or even aim to do so, since there can be charitable or self-destructive exploitation.
- 5 Exploiters always *use* exploitees: i.e. they bring it about that the exploitee, or some part or aspect or behaviour of the exploitee, makes a contribution to the achievement of the exploiter's goals.

## OBJECTIFICATION, EXPLOITATION, AND COMMODIFICATION

[I]t is impossible to be a person and a thing, the proprietor and the property.<sup>1</sup>

This chapter aims to give an account of three closely related concepts, each of which plays an important part in debates about the commercialisation of the human body. The three concepts are objectification, (what in the previous chapter I termed) 'wrongful use exploitation', and commodification.

### 3.1 Objectification

If anything a prostitute treats herself like a chair for someone to sit on. Her mind goes blank. She just lies there. You become an object. A lot of clients say, 'Respond. It doesn't seem normal just lying there.'<sup>2</sup>

Let's start with objectification, since this is the broadest of the three concepts. Indeed, I'll argue later that both commodification and wrongful use exploitation are particular types of objectification. Not surprisingly, to objectify is to treat as a mere object. However, treating *objects* as objects doesn't count as objectification. Rather, to objectify is to treat as a (mere) object something which isn't really (or merely) an object. Nussbaum makes the point as follows:

in all cases of objectification what is at issue is . . . treating one thing as another: One is treating *as an object* what is really not an object, what is, in fact, a human being.<sup>3</sup>

Treating things as objects is not objectification, since . . . objectification is making into a thing, treating *as a thing*, something that is not really a thing.<sup>4</sup>

In the first of the above quotations, Nussbaum restricts objectification to human beings and, because this book is mainly about the commercialisation of the *human* body, I'll do the same. There is, however, no reason to rule out *a priori* the objectification of extraterrestrials, animals, nature, or works of art – or, indeed, of anything which isn't a mere thing but can be treated as such.

So what exactly is it to treat something as a mere object? Nussbaum plausibly claims that it is impossible simply to read off a definite answer to this question, or indeed a precise definition of 'objectification', from ordinary language. She argues instead that objectification is a 'loose cluster-concept' and proceeds usefully to identify seven ways of treating something as an object, suggesting that 'we sometimes treat any one of these as sufficient' for objectification, 'though more often a plurality of features is present when the term is applied'.<sup>5</sup> The seven ways are described below (in terms of the objectification of persons).

- 1 *Instrumentality* – the person is treated as merely a tool for, or a means of, achieving the objectifier's goals.
- 2 *Denial of autonomy* – the person's autonomy is not recognised or not respected.
- 3 *Inertness* – the person is treated as lacking in agency ('and perhaps also in activity').
- 4 *Fungibility* – the person is treated as interchangeable with other similar persons (or even with relevantly similar non-persons).<sup>6</sup>
- 5 *Violability* – the person's right to bodily integrity is not recognised or not respected.
- 6 *Ownership* – the person is treated as property.
- 7 *Denial of subjectivity* – the experiences of the person are not recognised or not regarded as significant.<sup>7</sup>

Nussbaum spends a considerable amount of time constructing a complex account of the interrelations between these seven aspects or kinds of objectification and applies this to a variety of literary examples, most of which contain pornography or depict sexual objectification. There's not space to do the same here and, in any case, my own position is that this part of her analysis is basically correct. What I'll do instead is to proceed directly to the question of why we do (and perhaps should) regard her seven forms of objectification as morally bad ways of treating people.

### 3.2 Kantian underpinnings

[W]hen we use the term pejoratively, as in 'objectification of persons,' we mean, roughly, 'what Kant would not want us to do.' The person is a subject, a moral agent, autonomous and self-

governing. An object is a non-person, not treated as a self-governing moral agent.<sup>8</sup>

Margaret Radin (above) suggests that objectification (in the pejorative sense) is very much a Kantian moral concept, and it seems to me (and to Nussbaum) that she's right about this.<sup>9</sup> Underpinning our ethical worries about the objectification of persons are two principles that are broadly Kantian, though one certainly doesn't need to be a fully-fledged Kantian in order to subscribe to them.

The first of these – which is very obviously connected to the first sort of objectification, *instrumentality* – is Kant's well-known doctrine that people should be treated as ends-in-themselves rather than as (mere) means:

Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only.<sup>10</sup>

The second principle – which is most clearly connected to the fourth kind of objectification, *fungibility* – relates to Kant's less popularly known distinction between *price* and *dignity*:

In the kingdom of ends everything has either a *price* or a *dignity*. Whatever has a price can be replaced by something else as its equivalent; on the other hand, whatever is above all price, and therefore admits of no equivalent, has a *dignity*.<sup>11</sup>

According to this second principle, persons have *dignity*, in the sense explicated by Kant, and should be treated and viewed as such. Treating them as fungible (interchangeable) fails properly to respect their dignity and, in Kantian terms, is to regard them as having mere 'price'. This specifically Kantian way in which the term 'dignity' is used should be carefully distinguished from other 'looser' uses of the term, because rather vague ideas of 'human dignity' are often wheeled out during ethical debates about (for example) biotechnology, without there being any clear sense of what exactly dignity is, or why exactly the technology under consideration is an affront to it. Hence, Burley and Harris are quite right to talk of ethical objections often being couched in terms of '*mysterious appeals* to human rights and human dignity'. (They are referring to objections to cloning, but their point applies much more generally.)<sup>12</sup>

Next, I want to show how the wrongness of each of Nussbaum's seven types of objectification is entailed by one or both of these Kantian principles. In so doing, I'll be explaining why the view that objectification is wrong is commonly held, since adherence to the Kantian principles is itself widespread. Furthermore, on the assumption that these Kantian

principles not only are but also ought to be believed, I'll be going a long way towards justifying people's hostility to objectification.

The first type of objectifying (*instrumentality*) is easy to deal with since this is exactly what is proscribed by the first Kantian principle: treat people always as ends and never only as means. We'll be returning to look at instrumentalisation in more detail later on in this chapter.

The status of the second type of objectifying (*denial of autonomy*) is less clear because the meaning of 'autonomy' is unspecified. Also, there is an important distinction between *failing to recognise* that a person has the psychological property of being autonomous and *failing to respect* that autonomy (the latter probably being best understood in terms of unjustifiably restricting the person's freedom). Our difficulty then is that there is more than one way of understanding 'autonomy'. However, it can plausibly be claimed that, on at least most of these understandings, recognising and respecting people's autonomy is a necessary part of treating them as ends-in-themselves. Hence, denial of autonomy would be forbidden by the first Kantian principle.

If *inertness* is understood as the absence of agency, then it also falls foul of the first Kantian principle, since seeing people as ends-in-themselves requires seeing them as agents – people who do, or at least could, act (with action being broadly understood so as not to exclude, say, people with severe physical disabilities).

*Fungibility* is also easy to deal with since this is exactly what is proscribed by the second Kantian principle. According to this principle, we should treat people as having 'dignity' rather than mere 'price'. What this means is that we should not treat persons as if they were *replaceable* or *interchangeable*. We should instead treat them as *irreplaceable* and *unique* – and perhaps even as *incommensurable* with respect to one another (which means that the value of each person is neither more than, nor less than, nor the same as the value of each other person). Nussbaum says that

The connection between fungibility and instrumentality is loose and causal, rather than conceptual.<sup>13</sup>

About this, I would suggest, she is not quite right. For there does seem to be at least one *a priori* connection. The link is that anyone who adopts an entirely instrumental attitude towards X, in so doing, regards X as fungible. This is because to have a fully instrumental attitude to X is to value X only for its relevant causal powers – only for how effective it is as a means of achieving one's own ends. So if I value X in this way, I am indifferent as to whether I have X, or have a replacement, Y – provided that Y has the very same relevant causal powers as X.<sup>14</sup> Indeed, not to be indifferent as to whether I have X or have Y in such a situation would suggest strongly that I had not *fully* 'instrumentalised' X. A fairly trivial

example should illustrate the point. Some people have fully instrumental attitudes to their cars. Others are emotionally attached to their cars, perhaps because they are sentimental about experiences that they've had in them. People of the first type also regard their cars as fungible. Because they only care about the car's ability to get them 'from A to B', they don't mind its being replaced with any other equally effective vehicle. People in the second category, however, don't regard their cars as fungible – and this is precisely because they don't regard them as mere instruments. Rather, they are emotionally attached to them and would be upset if their car were replaced by another, even if it were the same make, model, colour, etc. They (rightly or wrongly) regard their vehicles as unique and irreplaceable.

For completeness, I should add that although regarding something as purely instrumental necessitates regarding it as fungible, it *may* be possible to view something as fungible without also viewing it as merely instrumental. Aesthetic value is a leading example here. Someone may not care which of two equally beautiful paintings she owns and yet still regard the aesthetic value of each as intrinsic rather than instrumental. Cases like this are tricky because it's possible to redescribe them in ways which make it sound as if the viewer is 'instrumentalising' the paintings – for example, we can say that she uses the paintings as *means* of experiencing beauty (or pleasure). But they do at the very least show that viewing-as-fungible doesn't *obviously* entail viewing-as-merely-instrumental. Perhaps a better (though rather theoretical) example of viewing something as fungible without also viewing it as merely instrumental is classical utilitarianism.<sup>15</sup> This theory sees human happiness as the one and only *intrinsic* (non-instrumental) good. However, it also sees happiness as a *fungible* good. For classical utilitarians don't (from a moral point of view) care *who* is happy, or which particular 'bits of happiness' exist, so long as overall happiness is maximised. People often attack utilitarianism on precisely this point, arguing that it wrongly obliges us to 'trade off' (in a thoroughly impersonal, 'agent-neutral' way) the welfare of one person against others, or even against the welfare of possible people.

The fifth item on Nussbaum's list is *violability*: failing to recognise and/or respect the person's right to bodily integrity. The right to bodily integrity is a limited right not to have one's body non-consensually damaged, invaded, touched, viewed, etc. It is typically invoked to explain such things as the wrongness of non-consensual medical treatment and rape. A commonplace view in ethics is that this right to bodily integrity follows from the principle of respect for autonomy, although saying why this is so is far from simple. In brief, one argument would be that an important part of being self-determining (autonomous) is having substantial control over one's own body – and, in order to have such control, one's body

must be free from non-consensual interference by others. If this line of reasoning is correct (as appears plausible), then the wrongness of treating people as violable follows from the first Kantian principle. For the right to bodily integrity follows from the principle of respect for autonomy, which in turn follows from the idea that we should treat persons as 'ends'.

*Ownership*, the penultimate category, is not entirely straightforward since 'owning a person' can mean a variety of things. More specifically, to own someone is to have certain rights over her, and so there will be different sorts of ownership corresponding to different sets of rights. An extreme example of ownership would be a system of slavery under which slave owners were legally permitted to kill and otherwise physically violate their slaves. Under such a system, the slave owners would have what we might call *fully-fledged* or *unrestricted* property rights in their slaves – meaning that they're allowed to do whatever they like with and to them, including destroying them. Clearly, ownership of this kind is wrong. Not only is it cruel and harmful, but it also very obviously falls foul of Kantian principles because the slave's autonomy is not respected and it is highly likely that such slaves will be regarded as fungible. As Anderson puts it:

In Kantian theory, the problem with slavery is that it treats beings worthy of *respect* as if they were worthy merely of *use*.<sup>16</sup>

However, not all ownership is like this, as Nussbaum reminds us:

Slaves are not necessarily regarded as violable; there may even be laws against the rape and/or bodily abuse of slaves.<sup>17</sup>

So someone could have a much more limited set of property rights over another person. For example, there could be a *relatively* humane system of slavery under which, although the slave owners had complete control over their slaves' labour, they were nonetheless obliged not physically to hurt them, and to provide them with health care, adequate food and shelter, and rest days and holidays. We can construct a series of progressively more 'humane' examples like this and, if we keep going, we'll eventually reach a point at which the 'owned' person starts to look more like a well-treated employee than a slave – though at that point we may start to wonder if the person is really owned. This raises the following (closely connected) questions. First, what kind of rights over B (a person) must A have in order to count as *owning* B? Second, does A's owning B *necessarily* involve A's wrongfully objectifying B? I won't attempt to answer these here, since such difficult questions are the concern of the entire book. I'll simply note for the present that, although almost

everyone is agreed that owning persons is wrong – since it amounts to or is like slavery – it's rather hard to say how we should distinguish between those clusters of rights which constitute owning a *part or aspect of a person (or of her life)* (for example, employers' contractual rights) and those which constitute *owning a ('whole') person* (for example, slavery). This issue will arise in various different forms throughout the book.

The final item on Nussbaum's list, *denial of subjectivity*, again follows pretty directly from the first Kantian principle. For, it can be argued, part of treating people as ends-in-themselves is recognising the existence and importance of their mental states and especially of the fact that they are subjects.

So to sum up, the view that each of Nussbaum's seven types of objectification is (*prima facie*) wrong can be justified by reference to two Kantian principles. The first says that we must always treat people as ends-in-themselves and respect their autonomy. The second says that we must recognise and respect people's dignity and not treat them as fungible.

### 3.3 Wrongful use exploitation

The fundamental synonym for the verb 'exploit' is 'use'.<sup>18</sup>

When market norms are applied to the ways we treat and understand women's reproductive labor, women are reduced from subjects of respect and consideration to objects of use.<sup>19</sup>

With an analysis of objectification and its Kantian underpinnings in place, we're now in a position to say what wrongful use exploitation is. Wrongful use exploitation is identical with Nussbaum's first kind of objectification, instrumentality. A exploits B (in this sense) if A treats B merely as a tool for, or a means of, achieving A's goals.

What's the status of this claim, and what are the arguments for it? As regards its status, it is meant to be a rational reconstruction of what people in fact mean when they use the word 'exploitation' in this way – i.e. what really concerns them morally is that one person is *using another person as a (mere) means*. As regards arguments for it, these take two forms. First, if we analyse academic and non-academic ethical discourse we'll find that this is one of the ways in which 'exploitation' is actually used. The same goes for the analysis of case studies involving alleged exploitation. Many of these centre around the question of whether the putative exploitee has been used as a mere means or not. Second – and here we move into more normative territory, as opposed to simply attempting to capture the facts of ordinary moral language – there are some theoretical advantages (coherence, explanation, justification, neatness, etc.) to defining wrongful use exploitation in this way. One is that

(provided we accept Kant's 'treat people always as ends' principle) it's easy to see what's wrong with wrongful use exploitation. For, by definition, this involves failing to treat people as ends and so always falls foul of the first Kantian principle. A second is that, on the definition offered, it's possible to fit wrongful use exploitation neatly into a family of moral concepts including commodification and objectification. Or, at least, that's what I aim to show in this chapter.

A question that we touched on during Chapter 2 and to which we can now return is: how exactly are the two exploitations ('disparity of value' and 'wrongful use') related to one another? Are they merely unrelated things which just happen to be referred to by the same word or do they have important features in common? Even though the analysis of exploitation isn't finished, it should already be clear, in general terms, what the main similarities and differences are. Their foremost shared feature is that, in all cases of exploitation, the exploitee (and/or one or more of her attributes) is used by the exploiter. This point is expressed in what I earlier called the use condition: the view that A exploits B only if A uses B in order to foster the achievement of A's goals. Where the two exploitations differ is in what's wrong (or thought to be wrong) with the use in question. In wrongful use exploitation, the exploitee is wronged by being *used solely as a means*, whereas in 'disparity of value' exploitation, the exploitee is wronged because she is *unfairly used* – with the unfairness typically consisting of being under-rewarded and of having her vulnerabilities taken advantage of (what Wood calls 'advantage-exploitation').

So although both exploitations involve using persons, they are really very different moral concepts. One is grounded in the Kantian 'treat people always as ends' principle and has nothing whatever to do with the unfair distribution of goods, or with taking advantage of the vulnerable; the other is all about the latter and nothing to do with the former. In practice, of course, both exploitations are often present together. This is no accident. If A subjects B to wrongful use exploitation, the fact that A also subjects B to 'disparity of value' exploitation shouldn't surprise us. For if A doesn't respect B as an 'end', A is not likely to be too concerned about treating B fairly or about not taking advantage of B.

This substantial difference between the two exploitations, combined with the facts that (a) they 'share' the same name and (b) they often occur together, helps to create a good deal of moral confusion. For example, it's sometimes hard for people to tell whether it's the *unfairness* of a particular using or *the plain fact that it's a using* (or both) that they find objectionable. For this reason and for stylistic reasons, I shall in subsequent chapters change my own terminology as follows. I'll reserve the word 'exploitation' for what I've so far been calling 'disparity of value' exploitation and I'll use the word 'instrumentalisation' for what I've so far been calling wrongful use exploitation. Does this mean that I think that

wrongful use exploitation isn't *really* exploitation? It's not clear to me that asking this question is terribly helpful. I have said what, in my view, wrongful use exploitation is. I have said what it has in common with 'disparity of value' exploitation and what it does not. And I have said that people *call* wrongful use exploitation 'exploitation'. Beyond that, it's not clear to me that asking whether or not it's *really* exploitation has much point. Why, then, have I chosen to allocate the 'exploitation' label to 'disparity of value' rather than to 'wrongful use'? My reasons for doing so are linguistic rather than philosophical: it seems to me that using the word 'exploitation' to refer to 'disparity of value' exploitation is simply *more common* than using it to refer to wrongful use exploitation. Furthermore, it is harder to find a suitable alternative word for 'disparity of value' exploitation than it is to find one for wrongful use exploitation.

The next two sections explore instrumentalisation (wrongful use exploitation) further by (in 3.4) revisiting the idea of treating persons as ends-in-themselves and (in 3.5) looking at its relationship with treating people as means.

### 3.4 Treating people as ends-in-themselves

That people should treat others as a means to their own ends, however desirable the consequences, must always be liable to moral objection. Such treatment of one person by another becomes positively exploitative when financial interests are involved.<sup>20</sup>

As we've seen, Kant's idea that we must always treat persons as ends-in-themselves and never as 'means only' plays a crucial role in explaining both what objectification is and why objectifying persons is *prima facie* wrong.

One feature of this principle which is sometimes overlooked (and which seems, incidentally, to be overlooked in the passage from the Warnock Report quoted above) is that it doesn't in fact prohibit treating people as means. Rather, it asserts that we've a positive obligation always to treat them as ends. So treating people as means may be permissible (according to the principle) so long as we *also* treat them as ends. That this is the meaning of the Kantian principle should be unsurprising, because if it ruled out treating people as means altogether, it would be wildly implausible. John Harris, commenting on the passage from the Warnock Report quoted earlier, makes the point as follows:

it by no means follows, as Warnock claims, that where people treat others as means to their own ends there is automatically exploitation. We all do this perfectly innocuously much of the

time. In medical contexts, anyone who receives a blood transfusion has used the blood donor as a means to their own ends.<sup>21</sup>

So, to reiterate, the Kantian principle should be understood as a requirement always to treat people as ends – even if, at the same time, we treat them as means. One thing that follows from this is that there may be ways of breaching the principle which don't involve instrumentalising persons. In other words, there may be cases of wrongfully failing to treat people as ends which aren't also cases where those wronged are treated as means. That this is so can be shown by thinking back to Nussbaum's seven types of objectification. Some of these breach the Kantian principle not because people are treated as means, but just because they are (for other reasons) not treated as ends-in-themselves. *Denial of subjectivity* and *violability* are perhaps the clearest examples of this. Now, of course often it's the case that when people have their 'subjectivity denied' and/or their rights to bodily integrity ignored, it's because someone else is using them as a mere instrument. But this isn't the only reason why these kinds of objectification can occur.

At this point, it's helpful to draw a (probably not very sharp) distinction, within objectification, between regarding a person as a *useful object* and regarding a person as a *useless object*. Most discussions of objectification focus on people who are wrongfully *used* and hence on those who are, in one way or another, viewed as useful. However, there is another way of being objectified: to be viewed as an object in which the objectifier has no interest and to which she assigns not even instrumental value. Totally ignoring people who sleep rough on the streets is a possible example of this phenomenon. Sometimes, people on their way to work, or on the way home after a night out, regard these 'street people' as like objects, not in the sense of being objects of use, but on the contrary as being a bit like bags of rubbish which are inconvenient and/or aesthetically displeasing. The personhood of 'street people' is often ignored and, even when they actively approach passers-by, they are more likely to be blanked than argued with. Indeed, one common complaint made by 'street people' is that they are treated as if they don't exist, as if their voices can't be heard, as if they're invisible. Their subjectivity is denied or ignored. They are treated as mere irritations rather than as people with intrinsic value and are at risk of being treated as (in Nussbaum's terms) 'violable'. This seems clearly to be objectification (treating as a mere object), though not a kind that involves use.

We've seen, then that treating a person instrumentally, as a means, is not *necessary* for breaching Kant's 'treat people as ends' principle. This is because it's possible to fail to treat persons as ends-in-themselves without also treating them as means (the leading examples here being those where the objectifier treats people as 'useless' objects). We've also seen

that treating someone instrumentally is not *sufficient* for breaching the Kantian principle, because it is possible to treat people *both* instrumentally *and* as ends-in-themselves. Given this, the obvious question to ask is: what exactly is the relationship between breaching Kant's principle and treating persons as means? This is the subject of the next section.

### 3.5 What's wrong with using people as means?

To answer this, let's first return to Nussbaum's article on objectification and, in particular, to what she has to say about Kant's views on sex and marriage:

Central to Kant's analysis . . . is the idea that sexual desire is a very powerful force that conduces to the thinglike treatment of persons . . . above all, the treatment of persons not as ends in themselves, but as means or tools for the satisfaction of one's own desires. That kind of instrumentalizing of persons was very closely linked, in his view, to both a denial of autonomy – one wishes to dictate how the other person will behave, so as to secure one's own satisfaction – and also to denial of subjectivity – one stops asking how the other person is thinking or feeling, bent on securing one's own satisfaction.<sup>22</sup>

Kant may or may not be right about the nature of sexual desire (and this isn't, sadly, something that there's space to explore further here). But even if we don't accept his views on sex, we can still learn something from the above passage about the way in which viewing persons instrumentally might lead us to deny their autonomy. For Kant's proposal seems in general terms a good idea, and – what's more – a remarkably 'common-sense' one. His thought is that when we become focussed on a person's instrumental value, on her usefulness to us, we have a tendency to disregard the fact that she is an autonomous being who deserves respect. In other words, thinking of a person in terms of use encourages us to overlook the fact that she's not *only* for our use, but is also an end-in-herself.

Some ways of focussing on a person's instrumental value seem more dangerous in this regard than others, with the risk varying depending both on the type and on the extent of any use involved. Some people would, for example, say that sexual use of other human beings is especially morally hazardous – meaning that there's a considerable risk of the sexually used person's autonomy being ignored by the sexual user (who may be overwhelmed by desires, feelings, etc. and, in some sense, forget that the sexually used body 'belongs' to another person and/or that the other person deserves respect and consideration). By contrast, most of us



would regard using someone as a teacher in order to learn as generally unproblematic. Because of the nature of teaching and of the teacher-student relationship, the use involved is not normally of a kind which encourages students to overlook the teacher's autonomy or personhood. Lying behind these views are, I suspect, a conception of sex which views the sexually used body as a passive and impersonal object, and a model of teaching which regards it as active, autonomous, and expressive of the teacher's personality. From this we might generalise as follows (without necessarily endorsing these particular views of sex or of teaching). When people (or people's bodies) are used as passive and impersonal objects, the risk of objectification is greater than when they're used in ways which are somehow expressive of their personalities.

Thoughts like these raise questions about the status of both (a) the general claim that when we think of people instrumentally we tend to disregard the fact that they're also 'ends', and (b) more specific claims about use, such as that using people for sexual gratification is more likely to make us overlook their personhood than using people as teachers. Are such claims just empirical psychological generalisations? Or are they 'deeper' conceptual truths? The word 'treat' – in the expression 'treat as an object' – is a useful shorthand for both the *attitudes* and the *behaviour* of the objectifier. And in order to answer this 'status question', we need to appeal to this kind of distinction, since we may need to give different answers depending on whether we're looking at how we *think* of people or how we *act towards* them.

Let's start with attitudes. It's clear that there's no reason in principle why valuing a person instrumentally must conflict with also valuing the same person as an 'end'. That this is so can be seen from our everyday lives. We all know individuals to whom we attach both instrumental and intrinsic value (such as those who are both treasured friends and proficient colleagues). Furthermore, not only *can* we attach both kinds of value, but often we *ought* to do so because the people in question really are *both* ends-in-themselves *and* undeniably useful. It would be bizarre to claim, surely, that in order to recognise and respect such people as persons, we need somehow to deny or ignore their usefulness.

In fact, this is just a particular application of a more general truth: that there's no reason in principle why one can't value something (or someone) instrumentally while at the same time recognising its other kinds of (intrinsic) value. We can use aesthetic value as an example here. Let's say that I own a beautiful figurine which I keep on my desk to look at and use as a paperweight. What's to stop me from *both* regarding it as useful *and* regarding it as an object of great beauty? Obviously, nothing. Indeed, assigning both kinds of value would be the rational thing to do, on the assumption that it really is both beautiful and useful, and doing so needn't involve any sort of conflict or contradiction. Or consider a similar

case – architecture. Many buildings can be, should be, and are designed to be appreciated both for their aesthetic merits and for their utility. In such cases, what's clear is that recognising something's instrumental value doesn't *need* to displace our appreciation and regard for other kinds of value. This is as true of persons as it is for works of art, or for buildings.

### 3.6 The displacement thesis

As we've seen, there's no conceptual or logical conflict between thinking of something instrumentally and valuing it in other ways too. However, there is instead what I'll term the *displacement thesis*. This says that, even though different modes of valuation don't logically contradict one another, they nonetheless conflict *psychologically*: that, as a matter of fact, instrumental valuation tends to *displace* other important modes of valuation in the human mind. Consider, as an example, Elizabeth Anderson's assertion that

The commodification of sexual 'services' destroys the kind of reciprocity required to realise human sexuality as a shared good. Each party values the other only instrumentally, not intrinsically.<sup>23</sup>

According to the displacement thesis, this claim should be understood as follows. For *psychological* reasons (which may be construed broadly) each party is very likely to neglect the other's personhood, intrinsic value, etc. This is because one party is interested primarily in using the other person as a *means of achieving sexual satisfaction*, while the other party is interested mainly in using the other person as a *means of obtaining money*. Both parties, though, are subject to broadly the same psychological phenomenon: the 'personhood perspective' on the other is *displaced* by a 'use perspective'.

Since my main aims are philosophical rather than psychological, I can't address in any detail the question of whether the displacement thesis is a good piece of psychology. However, before we move on, two further things can be said about it.

First, the fact that the displacement thesis is an empirical generalisation doesn't make it any less ethically significant (although it does, of course, raise the question of how good the empirical evidence for it is). Indeed, the displacement thesis is an important part of the answer to the question I posed earlier: what exactly is the relationship between treating people as means and the Kantian requirement that we must always treat people as 'ends'? There are two things to say in response to this question. One is to make a point about *action*, which we'll be looking at in the next section.

The other is to appeal to the displacement thesis and to say that what's wrong with thinking of people instrumentally (at least in certain ways or in certain circumstances) is that it encourages us to disregard their moral status as ends-in-themselves.

The second thing to note about the displacement thesis is that there are actually many different possible displacement *theses*. More specifically, some versions of the thesis are more or less *global*, whereas others are *local* – focussing specifically on *particular kinds* of use-valuation and/or on *particular kinds* of context or relationship. Clearly, a totally global displacement thesis – i.e. the view that instrumentally valuing *x* *always* increases the risk of failing to value *x* in other ways – would be implausible. This is because there are some cases of regarding-as-useful (such as the case of the teacher) in which valuing the person instrumentally doesn't appear to have any negative impact whatsoever on our appreciation of the person as an end-in-herself. Indeed, there are cases in which thinking that someone is good at something (and hence useful in that respect) *enhances* one's appreciation of 'the whole person'. What seem much more plausible are various local displacement theses. Sex is a leading example here and, as we've already seen, many people (including Kant) appear to believe in a local displacement thesis concerning sexual use – the idea being (roughly) that regarding people as sex objects encourages us to view them as *mere* sex objects. Another interesting example – one which seems to me to be more convincing than the sex case (not least because sexual relations can take so many different forms) – is cannibalism. Most people would find it hard simultaneously to regard someone as a person and as a tasty piece of meat. Thus, given this psychological incompatibility, one might argue persuasively that seeing people as succulent roast dinners is indeed likely to encourage us to ignore their moral status as persons.

One fascinating question (again, one which can't be addressed here) is: *why* are some kinds of instrumental valuation better than others at co-existing with respect for personhood? In fact there's really more than one question here. There's a question about causation. What are the *origins* of various attitudes to use? Are their causes mainly biological or are they 'social constructs'? And to what extent do these attitudes vary according to culture, ethnicity, gender, sex, etc.? There's also a structural question: is there anything about the *nature* of certain kinds of use that makes them more likely to displace respect for personhood?

### 3.7 Using persons, contexts, and relationships

[I]nstrumentalization does not seem problematic in all contexts. If I am lying around with my lover on the bed, and use his stomach as a pillow there seems to be nothing at all baneful about this, provided that I do so with his consent (or, if he is asleep, with a

reasonable belief that he would not mind), and without causing him pain, provided, as well, that I do so in the context of a relationship in which he is generally treated as more than a pillow. This suggests that what is problematic is not instrumentalization *per se*, but treating someone *primarily* or *merely* as an instrument. The overall context of the relationship thus becomes fundamental.<sup>24</sup>

Let's turn now to the other sense of 'treat', *acting towards*. The position here is a little more complex, chiefly because it can be argued that some ways of using things are incompatible with recognising and respecting their intrinsic value. Let's go back to the case where I own a beautiful figurine which I also use as a paperweight. We can change the example slightly and say that not only is it beautiful, but it can also be used as a high-density projectile, although if I ever used it in this way it would be destroyed. First let's ask: is valuing it as a thing of beauty incompatible with recognising its value as a projectile? The answer is *no*. A rational person could (and maybe should) simultaneously acknowledge the existence of both kinds of value. But now let's ask a different question, one about action rather than attitude: is *actually using it* as a projectile incompatible with recognising and respecting its aesthetic value? Arguably, yes. For by using it as a projectile, I destroy it. And destroying a thing of great beauty is usually symptomatic of a failure to recognise and respect its aesthetic value. So we appear to have a case which shows that some ways of using things are incompatible with recognising and respecting their intrinsic value (although, of course, in extreme circumstances, even someone who respects its aesthetic value might destroy it, for example, to save a hundred human lives).

It's not hard to see how this might apply, and in a very direct way, to persons. Consider, for example, the case of Manuel Wackenheim. Wackenheim is a (so-called) dwarf who (until a ban was imposed by the local mayor) made a living from being 'tossed' by customers in bars and nightclubs. This 'tossing' formed part of a *dwarf-throwing competition* – a sport 'in which the aim of the competitors is to fling a dwarf over the furthest distance possible'.<sup>25</sup> (Apparently, the French record for dwarf throwing is 3.3 metres.)<sup>26</sup> Wackenheim appeared keen to pursue his chosen career and didn't welcome the ban on dwarf throwing, saying 'this spectacle is my life; I want to be allowed to do what I want'.

Wackenheim, rather like the figurine/paperweight, can be *regarded* both as intrinsically valuable (*qua* person) and as instrumentally valuable (*qua* projectile). That is, there's no reason in principle why a friend of his couldn't both respect him as an 'end' and recognise the fact that his body is formed in a way which makes it instrumentally valuable to dwarf throwers. But could Wackenheim's friend *actually use him* as a projectile

while at the same time respecting the fact that he's an 'end'? The passage quoted at the start of this section (from Nussbaum) gives the key to answering this: it is the overall context of the relationship, along with other structural features of the situation, which determines whether he is appropriately respected. Surely this must be right, for it would be extraordinarily hard to argue plausibly that it's *impossible*, in *all* contexts, to use Wackenheim as a projectile while at the same time respecting his personhood. For what if he enjoys being thrown, gets paid for it, and freely and knowingly consents to it? If for these reasons – i.e. because I want to ensure that he derives pleasure and money, and am certain that it's what he really wants, etc. – I throw him, then there seems no basis whatsoever for saying that I'm failing to respect his personhood. For in such a case, I'm (let's assume) deliberately giving him what he wants (and wants in a free, informed, and otherwise autonomous way) and deliberately benefiting him. How can this be a failure to recognise his status as an 'end'?

It can't be. Although I should straight away add two qualifications. First, this doesn't rule out the existence of other separate moral objections to dwarf throwing. Second, there certainly are contexts in which dwarf throwing would constitute wrongful instrumentalisation. These may include cases in which the dwarves are substantially harmed and cases in which the dwarves don't consent or in which their consent is invalid. Indeed, this is my main point. What does all the ethical work here is context: in particular, issues relating to the existence and quality of the dwarf's consent – and, for the more paternalistically inclined, those relating to harm and welfare.

Lying behind this way of thinking is one of the following principles (which differ just in that the second, more cautious, principle includes a 'substantial harm' constraint).

If A seeks and obtains from B valid consent to do x to B, that is sufficient to guarantee that B's status as an end-in-herself is respected by A (other things being equal).<sup>27</sup>

If A seeks and obtains from B valid consent to do x to B, and x is not substantially harmful to B, that is sufficient to guarantee that B's status as an end-in-herself is respected by A (other things being equal).

These principles are driven by the view that: (a) respect for autonomy is either identical with, or is the most important part of, Kant's 'treat people always as ends' doctrine, and (b) the relationship between autonomy and consent is of the first importance. On this view A, just by seeking and obtaining valid consent, has respected B's autonomy and status as an 'end' – though I should add that it must also be the case that *A wouldn't have gone ahead and done x to B if B hadn't consented*. In other words, what

matters isn't so much the *existence* of consent as the fact that *A requires* it. It is by *requiring* consent (i.e. by not being willing to act without B's consent) that A respects B's autonomy.

We'll be looking at valid consent again later. So, for now, I'll just restrict myself to making two observations. First, we should remember that the word 'valid' is doing a lot of work here. Consent *per se* is relatively insignificant, since an invalid consent, in ethical terms, is often no better than no consent. It is *valid* consent which matters most morally. Indeed, what 'valid' *means* here is that the consent in question is morally significant, that it goes at least some way towards justifying A's actions. Second, the view that A treats B as an 'end' so long as A requires B's consent does *not* entail the view that there's nothing to which B couldn't validly consent – that is, it's entirely compatible with the view that in practice there are certain kinds of use to which no one could validly consent. One may think (for example) that selling parts of one's body or being used as a projectile can't be validly consented to, because only someone whose consent was invalidated by coercion, manipulation, insanity, desperation, or whatever would agree to be used in such awful ways.

The picture that I'm offering is one in which there are no (non-normatively categorised) types of use that by their very natures preclude respect for personhood. Everything depends on the context in which the use takes place, on the relationship between the user and the used, and (most importantly of all) on the quality of the used person's consent. Having said that, there may well be kinds of use which nearly always involve failing to treat the used person as an 'end', and, for applied ethics and policy formation purposes, this is often what matters. In other words, for these purposes, a reliable statistical connection (for example, 'in 99 per cent of cases, B's personhood isn't respected') is as important as a necessary and universal one. The sorts of use which are likely to fall into this category are ones which are substantially harmful and/or unpleasant to the user. There are two possible reasons for this. First, we might think that if A respects B as an 'end', A won't harm B and so won't subject B to harmful use. A difficulty with this view, however, is what to say about cases in which B autonomously *wants to be harmed* by A. In such cases do we disrespect B's autonomy more by harming B or by adopting a paternalistic stance towards B, depriving her – 'for her own good' – of what she wants? A second (and better) reason for thinking that harmful usings are liable to involve not treating the person as an 'end' is that, more often than not, when people knowingly consent to be harmed, their consent is defective. For most people simply don't want to be harmed – and so why on earth would they consent to be harmed unless they were subjected to some kind of consent-invalidating pressure or manipulation? This seems to provide a very good reason for at least being suspicious of and scrutinising carefully cases of harmful use, since there's a fair chance that any

consent offered in such cases is invalid. We should, however, note that it is possible in principle to consent validly to be harmed, both because it's possible for people autonomously to desire self-harm and because there may be external reasons, should as wanting to help a third party, for subjecting oneself to harmful use. Living organ donors would be a good example of the latter. They typically consent to be harmed in order to benefit the recipient.

### 3.8 Commodification

[F]or some, any financial inducement to increase the supply of organs constitutes a 'commodification' of something that is essential to saving lives, and this is intrinsically objectionable. This position is typically an emotionally charged assertion and not a reasoned argument.<sup>28</sup>

In the next two sections, I aim to say what commodification is, to say how it relates to objectification and to instrumentalisation, and to assess Manga's claim (above) that appeals to commodification are typically no more than emotionally charged assertions.

Like 'exploitation', the word 'commodification' has both a non-moral and a moral sense. In the first of these, 'commodification' refers to a social practice and/or legal system under which the relevant class of things is (or is allowed to be) bought and sold:

Commodification is a *social practice* for treating things as commodities, ie as properties that can be bought, sold, or rented.<sup>29</sup>

To commodify something is to exchange it for money, to buy or sell it.<sup>30</sup>

Radin calls this the *narrow* construal of commodification:

Narrowly construed, commodification *describes actual buying and selling (or legally permitted buying and selling) of something.*<sup>31</sup>

Used in this way, 'commodification' is purely descriptive: i.e. to say of a practice that it's commodification (in this sense) isn't to criticise it. This idea can be contrasted with the second sense, in which 'commodification' is to be understood as a negative moral term. 'Commodification', in this sense, is supposed to denote a specific kind of wrong.

Our interest is mainly in the second sense of 'commodification'. This is because what we're looking for are possible objections to commercialising the body, and the idea of commodification can't (of course) provide an objection if 'commodification' is *just another word* for commercialisation.

As an illustration, consider the following criticisms of organ sale (along with various other commercial practices) from Bob Brecher:

The point about buying a pint of blood, or a kidney [or] renting someone's body for an hour or two ... is that all these are ... based on making a commodity of human beings.

... the possibility of people's buying a kidney represents the further commoditisation [commodification] of human beings, [and] to that extent the practice resembles prostitution, certain forms of surrogacy, and ... page three of *The Sun*<sup>32</sup> in symbolising, partly constituting, and encouraging a moral climate within which the commoditisation of human beings proceeds apace.<sup>33</sup>

Brecher's point here is not merely that permitting commercial surrogacy, organ sale, and prostitution would constitute an extension of the *social practice* of commodification. That claim taken alone would be too obvious to be worth a mention and certainly wouldn't constitute an *argument* for prohibition, but would be more like a description of what's to be prohibited. Instead, his point (which is far from trivial) seems to be that we ought not to permit these practices because doing so will encourage people *wrongfully to treat other people 'as commodities'*.<sup>34</sup>

To sum up, there's an important distinction between a purely descriptive sense of 'commodification', in which commodification just equals actual buying and selling, and a normative sense. It is the second (normative) sense which concerns us here.

### 3.9 Commodification, fungibility, and dignity

If poor people ... find themselves seeking to enter into 'desperate exchanges', such as selling their kidneys or their sexual services, some people will be troubled by their willingness to commodify attributes of self that our culture ... does not conceive of as fungible.<sup>35</sup>

I have some sympathy with Manga's 'emotionally charged assertion' point (quoted at the beginning of section 3.8) insofar as words like 'commodification' do tend to get overused and sloppily used. For this reason, it's impossible to provide a coherent account of the moral concept of commodification which isn't in some way revisionary. Nonetheless, I hope that what I'm about to say will at least capture the spirit of and count as no more than a refinement of what people mean when they object to things by appealing to the idea of commodification.

Before we turn to the substantive analysis, two preliminary points need to be made. First, for reasons that will become clear shortly,

commodification (in the normative sense) is a type of objectification. So, just as it's not possible to *objectify* something which is really an object, it's not possible to *commodify* something which is really a commodity. For example, if I were to regard baked beans or sugar as commodities, I wouldn't thereby be commodifying them, since they *are* commodities. So it's not the possession of the commodifying attitude *per se* which is wrong but the inappropriate application of it to entities which aren't (proper) commodities (notably persons). We should, however, keep in mind that some people think that nothing is really a commodity and/or that nothing should be treated as a commodity – and that, for them, anything could be commodified.

The second preliminary point is that, although I have suggested on occasion that our interest should be principally in commodificatory attitudes, this is actually slightly too narrow. We might instead usefully employ the word 'treat' here (which, as we saw earlier, covers both attitudes and actions) and say that what we're interested in is the idea of *treating persons as if they were commodities*. Nonetheless, it's still best to carry out the analysis of commodification mainly in terms of attitudes because whereas treating *x* as a commodity always involves the possession of commodificatory attitudes, it may or may not involve acting on or expressing those attitudes. In other words, having the attitude is a necessary part of commodifying; acting on the attitude is not.

To turn now to the analysis, what we have so far is that to commodify is to treat as a commodity something which is not a (proper) commodity. In order to make this less schematic, and to distinguish commodification from objectification-in-general, we need to look a little more closely at what exactly a commodity is. The word 'commodity' can be used in a broad sense to refer to all those things that are traded. However, it also has a narrower sense, and it is this which can help us to understand what commodification is and what's wrong with it. In this narrower sense, *commodities are fungibles*: things such as coffee beans, gold, and oil. All these things come in slightly different forms and at different levels of quality. But they are nonetheless fungibles in that the buyer normally just buys a certain amount of a certain good at a certain price per unit mass/volume without caring which particular (token) sack of coffee beans, or lump of gold, or tanker of oil she ends up with. Commodities, then, are fungibles, and to regard something as a commodity is to regard it as fungible. Hence, to *commodify* (in the normative sense) is to treat as fungible something which isn't fungible and/or oughtn't to be viewed as such. Commodification then seems to be identical with Nussbaum's fourth kind of objectification (fungibility). And when applied to persons it breaches the second Kantian principle, since treating persons as fungible fails properly to respect their dignity; it is to regard them as having mere 'price'.

The overall picture which is emerging, then, is that we have a comparatively general moral concept, objectification, of which there are roughly seven varieties. (There is, I suspect, a lot of arbitrariness about how we carve up the territory here – but Nussbaum's list of seven seems to me to be as good as any of the alternatives.) That each of these seven ways of objectifying is *prima facie* wrong (when applied to persons) can be shown by appealing to two Kantian principles: (a) don't treat people solely as means, and (b) don't treat people as fungible. As well as objectification, there are two more specific concepts: instrumentalisation (wrongful use exploitation) and commodification. These are forms of objectification. Each is closely related to one of the Kantian principles. To exploit someone in the 'wrongful use' sense is to breach the first Kantian principle by treating her solely as a means. To commodify someone is to breach the second Kantian principle by regarding her as fungible. Although distinct, instrumentalisation and commodification are closely linked. Perhaps the strongest link is that (for reasons offered in section 3.2) anyone who adopts an entirely instrumental attitude towards *X*, in so doing, regards *X* as fungible. Hence, it would seem that wherever there is instrumentalisation there is also commodification.

The final claim in the previous paragraph – that wherever there's instrumentalisation there's also commodification – will appear false to many people. This is because it's clearly possible to instrumentalise someone without any money changing hands and without trading of any sort taking place. For example, people are sometimes used (instrumentalised) as 'sex objects' within personal relationships in entirely non-commercial ways. So, one might ask: given that there are cases of non-commercial instrumentalisation, how can instrumentalisation entail commodification, since surely treating someone as a commodity must involve commercial transactions of some sort? What I've suggested so far does, I admit, sound strange because of the close association between commodification and money. This association is so close that many writers on the subject (including Radin) use terms like 'commodify' and 'monetise' more or less interchangeably.<sup>36</sup> Elsewhere, I have dealt with this by arguing that something can be treated as a commodity without any sort of commercialisation being involved.<sup>37</sup> One argument for this is that people can treat gifts, things that they find or steal, or persons with whom they have no financial relationship as commodities (i.e. as fungible).

However, I am not quite so confident about this answer as I once was. The reason is that the core moral concept here isn't, perhaps, commodification, but what we might term *fungibilisation*: wrongfully treating as fungible that which is not fungible and/or oughtn't to be treated as fungible. What's the relationship then between commodification and fungibilisation? About this, we have a choice to make – though one which is more

terminological than philosophical. One option is simply to *identify* commodification with fungibilisation. This is what I've been doing so far in this section and what I've argued for elsewhere.<sup>38</sup> The other is to reserve the word 'commodification' for those cases of fungibilisation which involve commerce, money, etc. – in which case commodification becomes a *particular kind of* fungibilisation (maybe the most prevalent one). It now seems to me that there's a lot to be said for this second option. It has two main advantages. First, it stops the account of commodification from clashing with ordinary language and with people's intuitions about cases: for example, on this view, people who are non-commercially instrumentalised as 'sex objects' *aren't* also commodified (though they are fungibilised). Second, it enables more conceptual-linguistic precision, because we now have not only the general term, 'fungibilisation', but also a more specific term ('commodification') which usefully picks out those cases of fungibilisation which are caused by or otherwise intimately related to commerce, money, the market, etc.

So the account of commodification (of persons) that I now support is as follows. Fungibilisation is a type of objectification which involves wrongfully treating persons as fungible. And commodification (in the normative sense) is a type of fungibilisation: more specifically, to commodify is to fungibilise through, or because of, commerce. Although we express ourselves in very different ways, this – it seems to me – makes my view of commodification a little like Radin's. According to her, there are

four indicia of commodification in conceptualisation . . . (i) objectification, (ii) fungibility, (iii) commensurability, and (iv) money equivalence.<sup>39</sup>

By *commensurability*, she means

that the value of things can be arrayed as a function of one continuous variable, or can be linearly ranked.<sup>40</sup>

And by *money equivalence* she means

that the continuous variable in terms of which things can be ranked is dollar value.<sup>41</sup>

The ideas of *commensurability* and *money equivalence* are useful when it comes to thinking about commodification's role within fungibilisation-in-general and about how commercialisation generates fungibilisation. The explanation is as follows. When a thing is commercialised, it is (by definition) assigned a monetary value. This forces, or at least encourages, us to view it as commensurable both in relation to money itself and in relation

to other things with the same monetary value. What commensurability means is that if I'm a purely 'rational economic agent' (and interacting with a perfectly fluid and otherwise 'ideal' marketplace) I'll be indifferent as to whether I have x (worth \$1,000), y (worth \$1,000), or \$1,000. It's not hard to see, then, how we get from the commercialisation of persons to their fungibilisation. For commercialising persons encourages us to regard them as commensurable (i.e. not to care whether we 'have' A, B, or \$n), and this is in turn liable to cause or constitute fungibilisation (and hence commodification).

### 3.10 Uniqueness

[P]art of me is only me and mine and part of you is only you and yours. No one has had my or your exact history.<sup>42</sup>

Up to now, the discussion of treating people as fungible has proceeded at a fairly high level of abstraction. So I'd like now to raise a question about how these ideas might be applied, one which leads me to be a bit sceptical about the moral force of claims about the uniqueness of persons.

A good place to start is this passage from Sandra Marshall's paper on prostitution:

it is not difficult to see how fungibility can be an aspect of a market relation. One well-trained waiter is just as good as another, one paying customer is just as good as another. There may be nothing wrong with this: to treat people as interchangeable within the market activity need not be in any way damaging. Indeed, it is perfectly possible to imagine that with developments in robotics we could do without waiters altogether without loss.<sup>43</sup>

Marshall's remarks remind us that, in many contexts, regarding people as fungible is morally unproblematic. For I'm not obliged to care which particular hairdresser, or taxi driver, or 'checkout operative' happens to service me – and while personal engagement with people in these roles may well be desirable for all sorts of reasons, it's surely not morally required. Radin tells us that 'the idea of fungibility . . . undermines the notion of individual uniqueness'.<sup>44</sup> About this she is right. But, as we've just seen, the problem with applying this idea to persons is that, in many respects, people don't appear to be unique. At least as role occupants, one customer/hairdresser/waiter is much like another. Or, even if we don't want to go that far, it's clear that before applying the idea of uniqueness to persons we need to know much more about the kind of uniqueness that's at stake. That is, the uniqueness thesis stands in need of clarification, so that we can say what it is that we're supposed to be wrongfully

disregarding when we treat people as fungible. All of which takes us to the question that I want to pose in this section. How exactly is the claim that each individual human being or person is unique to be understood? And ought we to believe it?

There are three main interpretations of this uniqueness thesis. Under the first, it's an empirical claim about diversity. People (on this view) *needn't* be, but as a matter of fact *are*, unique – meaning just that they, as a matter of fact, differ from one another considerably (for example, supporters of this view sometimes claim that genetics is on their side, because each human being has its own 'unique' set of genes). There are at least two problems with this version of the uniqueness thesis. First, as Thomas Kuhn reminds us, 'it is a truism that anything is similar to, and also different from, anything else'. Or to put it another way, it's very hard non-arbitrarily to specify what should count as difference/diversity and what should count as sameness/similarity. For this reason, unqualified statements such as 'everyone's basically the same' and 'everyone's a unique individual – we're all different' are hopelessly vague and lacking in content. The second problem is that this version of the thesis makes people's moral status dependent on accidental and seemingly irrelevant differences between them and others. This seems (at best) odd. To see why, consider hypothetical identical twins who are exactly similar in all respects. These twins aren't (*ex hypothesi*) unique: i.e. they don't differ from one another significantly. Yet it would be bizarre to claim that they lack the same moral status as other people, or that it's OK to treat them – but not other people – as fungible, or that the moral status of one twin would be higher if the other twin died, or if the other twin had never been born. For these reasons, we must reject the first interpretation of the uniqueness thesis.

According to the second interpretation, the uniqueness thesis is supposed to be a necessary truth about persons, something along the following lines: *no two persons share all the same properties*. The problem with this interpretation is that it appears to be either false or what philosophers call *trivially true* (i.e. true, but lacking in content or importance). There's clearly a sense in which two people could have all the same (non-relational) properties. For example, I could (in theory) be replicated, or split into two persons. And surely it would then be at least possible for me and the replicant to have all the same properties, both mentally and physically. At this point, defenders of this version of the uniqueness thesis will have to resort to such things as *spatio-temporal properties*. For example, they might say that the replicant is 3.7 metres closer to Canada than I am and therefore different from me in this respect. Furthermore, they might say that we're *necessarily* different, because our spatio-temporal properties must *always* differ (since two different objects can't occupy the same space at the same time).<sup>45</sup> Hence, it can be argued, even

the replicant and I are unique because we're never in exactly the same place at the same time; we never share all the same spatial properties. However, this renders the uniqueness thesis *trivially true*. Or at least it's far from clear why this is a morally significant kind of uniqueness, because fruit flies and stones are as unique in this sense as persons are (for example, no two stones occupy the same space).

The third interpretation of the uniqueness thesis says that what's unique about persons is consciousness or 'the self'. It is extraordinarily hard to know what to make of such claims – not least because, as I suggested earlier, it's very hard non-arbitrarily to specify what counts as being different or unique. Is my consciousness or 'self' unique? Well (leaving aside sceptical worries about knowledge of other minds), I can think of no reason to believe that my lived experiences are radically qualitatively different from anyone else's. This is not to deny that people have different experienced lives. As far as I can tell, they do. But this isn't going to be enough to ground a (non-trivial) *uniqueness* claim, because just as some people have very different experienced lives from mine, others may well have rather similar experienced lives to mine.

Perhaps, though, this misunderstands the point. Maybe the claim isn't so much that my 'self' is *qualitatively different* from others' 'selves' but rather that I am unique and irreplaceable in the sense that *there will only ever be one Stephen Wilkinson* (though there are, of course, other people called Stephen Wilkinson). There's certainly a sense in which this is true. Unfortunately for the uniqueness thesis, though, this has more to do with the way proper names work than with the existence of some more profound philosophical truth about uniqueness. In other words, the reason why 'there will only ever be one Stephen Wilkinson' is true is just that 'Stephen Wilkinson' is being used as a proper name. Proper names are what Saul Kripke famously calls *rigid designators*.<sup>46</sup> This means that they pick out the very same individual in all possible worlds and are to be contrasted with *definite descriptions* such as 'North Staffordshire's best-dressed philosopher', which, though it may (or may not) designate Stephen Wilkinson in the actual world, can designate other individuals in other possible worlds. So uniqueness of this sort seems to be just a function of naming which makes the uniqueness thesis look ethically insignificant. For I may as well give my PC a name (say, John) and claim that 'there will only ever be one John' and that John is irreplaceable and unique. This will be true so long as John is understood as a proper name and in spite of the fact that, in practical terms, John is eminently replaceable – just as (I suspect) Stephen Wilkinson is, though with a little more difficulty given the present state of our biotechnologies.

What can we conclude from all this? There appear to be two fundamental problems with using uniqueness in ethical arguments. The first is that it is hard to make sense of the claim that persons are unique.

Depending on how it's interpreted it seems to be false, trivially true, or morally insignificant. And (of course) if persons aren't really unique in any interesting sense, then it's not clear why we're obliged to regard them as unique. Indeed, doing so seems irrational. The second problem is that there are clearly lots of contexts in which treating people as fungible is ethically unproblematic. The most clear-cut cases are perhaps ones where someone is filling a particular social or economic role, for example, driver, teacher, waiter. Here, at least in uncomplicated cases, it's hard to see why we shouldn't regard individual role occupants as replaceable and non-unique, at least *qua* role occupant. These considerations lead me to believe that we should regard appeals to uniqueness with a fair degree of scepticism.

### 3.11 Bodies and persons

So far I've talked almost exclusively about the objectification of *persons* and have said very little about the objectification of *bodies*. This is perhaps surprising given the concerns of the book, but there are good reasons for it. One is that, as we'll see in a moment, the distinction between 'innocently' treating a body as an object and wrongfully *objectifying* it is itself based on the idea of objectifying persons. A second is that (as we've seen throughout this chapter) the moral concepts of objectification, instrumentalisation, and commodification are, we might say, almost *made for* the concept of personhood. In particular, these concepts are linked directly to personhood via the two Kantian principles: (a) don't treat persons solely as means, and (b) don't treat persons as fungible. Furthermore (because of this conceptual closeness), it's fairly easy to show why the objectification of persons might be thought to be wrong. Things are not so straightforward, however, when it comes to bodies. And they are even less straightforward when it comes to bodily parts, products, and services. For, even if the objectification of persons is wrong, it doesn't follow from this (or, at least, it doesn't follow obviously or directly) that there's something wrong with treating body parts as (mere) objects.

Some readers may find it strange that I'm distinguishing between persons and bodies in this way, and it may appear to them as if I'm advocating Cartesian Dualism and suggesting that persons are disembodied souls.<sup>47</sup> I don't mean to suggest anything so metaphysically extravagant. Nor do I need to, because fortunately the relationship between mind and body, or persons and bodies, is not an issue in which we need become embroiled. For it can be shown that, regardless of what one thinks about the metaphysics of mind, most (if not all) of our ethical concerns about treating bodies as objects are really concerns about the treatment of persons.

To see why, let's start by reminding ourselves of something which is

obvious and yet sometimes overlooked by writers on the commercialisation of the body. Bodies and body parts are physical objects. Hence, any ethical concerns that we have about the objectification of bodies can't be about whether bodies are treated as objects – since they are objects. Rather, our concerns (insofar as they're rational) must be about whether bodies are treated as mere objects. But what does 'mere' mean here? Or, to put it another way, what might bodies be over and above physical objects? What is treating a body as a *mere* object to be contrasted with?

Some people think that bodies are nothing over and above physical objects. For these people, it's impossible to *objectify* bodies, since (mere) objects can't be objectified. If, however, we reject this 'objectification-sceptical' view, then there seems to be only one remotely plausible answer to the question just posed. *Bodies are more than mere objects insofar as they are intimately related to persons*. This is why, in the main, people attach much less moral importance to dead bodies than to living ones. It is because dead bodies aren't connected to the person in the way that live bodies are. That's not to say, of course, that people don't attach *any* importance to the bodies of the dead. They often do, though when they do it's normally because they view the dead body as somehow significantly related to the person who used to exist – or, for those who believe in 'life after death', still exists in a non-bodily form.

To *objectify* the body, then, is to treat it as a *mere* object. This means treating it as if it weren't *intimately related to a person*. This raises two further questions. What is the nature of the intimate relationship between body and person? And what counts as treating a body as if it's not intimately related to a person? The first question is essentially about the relationship between body, mind, and personhood, and it is at this point that we can opt out of the metaphysics of mind by simply not answering it. This is possible because (a) it's the second question which is *ethically* significant, not the first, and (b) provided that we're working within a framework which accepts a few basic assumptions, such as that there are such things as persons which are somehow related to bodies, then the answer that we give to the second question needn't depend on the answer that we give to the first. Thus, Cartesian Dualists, functionalists, and physicalists (in other words, people with a wide range of different views about the relationship between body, mind, and personhood) might all agree that if A removes B's vital organs to feed to A's dog, then A is treating B's body as if isn't intimately related to a person (either that or A is deliberately harming B, in which case our moral objection to A's behaviour is likely to be something other than objectification). Ethical views like this don't depend on one's particular views in the philosophy of mind.

What about the second question? What counts as treating a body as if it's not intimately related to a person? A general answer is that treating a



body as if it's not *a person's* body is one common way of objectifying a person; when we treat a person's body as if it's not the body of that person, we thereby objectify not only the body, but the person too. Thus, what Nussbaum terms *denial of autonomy*, *inertness*, and *denial of subjectivity* will all often involve treating a person's body as if it's not the body of a person (as if it's a mere thing). In practical terms, consent and context will again be vitally important in every case. For example, if I act on another person's body without requiring and obtaining her consent this may amount, in a sense, to treating her body as if it's not 'hers' (as if it's a mere body). Conversely, if I require and obtain valid consent before acting on her body then (arguably) I'm not treating her body as a mere body, *no matter what I do to it*. What is clear, then, is that the idea of body-objectification is very much parasitic on the idea of person-objectification.

### 3.12 Summary and conclusions

This chapter doesn't have a great deal to say about the body. Its relation to the body commodification debate is that it aims to make sense in general terms of a certain sort of objection to commercialising the body. Objections of this kind focus principally on the idea that we shouldn't treat people solely as objects, or as means (instruments), or as commodities – and on the corresponding moral concepts of objectification, instrumentalisation (wrongful use exploitation), and commodification. All such arguments have a strongly Kantian flavour and are often based on Kant's view that we should treat people always as ends-in-themselves and never as fungible. Objectification (of which instrumentalisation and commodification are types) is supposed to be wrong in that it involves breaching these Kantian principles.

A lot of objectification arguments against commercialising the body do little more than draw attention to the fact that a person, or her body, is used as a means. We've seen, however, in Chapter 3 that such objections, if they're to work, need to say a lot more, since the relationship between using someone as a means and failing to respect her as an end is not straightforward. In particular, it's possible to use someone as a means while at the same time respecting her status as an end; using someone as a means doesn't entail a failure to respect personhood. Rather, everything depends on the overall context of the relationship and on the quality of the used person's consent.

That said, we've also seen that there may well be important links, albeit indirect ones, between treating people as means and failing to treat them as ends. One of these is (what I earlier called) the displacement thesis. This says that, even though different ways of valuing people don't logically contradict one another, they nonetheless clash psychologically in

that instrumental valuation tends to displace other important modes of valuation in the human mind.

Turning now to commodification and fungibilisation, we've seen that the main problem with deploying these concepts is that it's hard to make sense of the claim that persons (and *a fortiori* their bodies) are unique. For, depending on how it's interpreted, it is false, or trivially true, or ethically insignificant. We're left wondering therefore whether commodification and fungibilisation are really independent wrongs or whether, instead, regarding persons as commodities is perturbing just because it's symptomatic of instrumentalisation, of regarding them solely as means.

## HARM

Exploitation is an evil that is not typically free-floating. More often than not, perhaps, it is harmful to the interests of the exploitee . . . But a little noticed feature of exploitation is that it *can* occur in morally unsavory forms without harming the exploitee's interests.<sup>1</sup>

Accusing commercial practices of being exploitative and accusing them of being harmful are common moves in the body commodification debate. This chapter aims to improve our understanding of such assertions by exploring the connection between exploitation and harm and by looking at the nature of harm. The question of whether the commercial utilisation of the body is harmful is vitally important, especially given the widespread acceptance of Mill's *harm principle*. According to this,

The only purpose for which power can rightfully be exercised over any member of a civilised community against his will is to prevent harm to others.<sup>2</sup>

In other words, only harmful practices should be prohibited. So while showing the commercialisation of the body to be *harmful* would provide a powerful argument against it, showing it to be *harmless* would (if we accept the harm principle) give us a powerful argument for not prohibiting it (even if it's morally wrong for other reasons). The same goes for exploitation in general: i.e. supporters of Mill's harm principle should only want to ban harmful exploitation, not harmless exploitation (if there is such a thing).

The chapter starts by looking at the harm condition, the view that exploitation is necessarily harmful to the exploitee (the exploited person). It then attempts to say what harm is. The conclusion is that, properly understood, the concept of harm is messy and moralised and that, for this reason, moral principles such as *nonmaleficence* ('do no harm') and Mill's

harm principle aren't as useful (or, at least, aren't as straightforward) as many people think that they are.

## 4.1 Mutually advantageous exploitation?

Not surprisingly, philosophers disagree about whether the harm condition is true and about whether or not there can be such a thing as mutually advantageous exploitation. Buchanan, for example, supports the harm condition, saying that

To exploit a person involves the harmful, merely instrumental utilization of him or his capacities, for one's own advantage or for the sake of one's own ends.<sup>3</sup>

Against, this Wertheimer offers the following perceptive remark, which constitutes an argument against the harm condition:<sup>4</sup>

exploitation would be of much less theoretical interest on a 'no harm, no exploitation' rule. We do not need to be moral rocket scientists to know that it is wrong for A to gain from an action that unjustifiably harms or coerces B . . . By contrast, it is more difficult to explain when and why it might be wrong for A to gain from an action that benefits B and to which B voluntarily consents.<sup>5</sup>

The intuition Wertheimer appeals to here is that exploitation is a somewhat mysterious, or at least complex, moral concept, whereas harm appears not to be. Later on, we'll see that the concept of harm is not quite as straightforward as it at first seems. Nonetheless, we do normally feel that we know what harm is, and that we know that (in the absence of some special justification) harming others is wrong. Therefore, if exploitation were necessarily harmful, or a 'harm concept', we would expect it to seem more straightforward than it does – not least because the wrongness of exploitation could then be grounded in the idea of *non-maleficence* (roughly, the principle that inflicting harm on persons is wrong, other things being equal). Furthermore, although some exploitation claims are closely connected to associated harms, their content clearly goes beyond the fact that harm has been caused.

A more tangible way of making the same point is to look at some examples of exploitation in which the exploitee appears to benefit. Consider, for example, the following hypothetical:

Andrew and Brian

Brian is unemployed and has no food. If he doesn't get food soon,

he will die. There are no welfare agencies, charities, or benevolent individuals around to help him, and turning to crime is unlikely to succeed, because Brian lives in a police state. His only way to survive, therefore, is to get a job. Andrew, a local factory owner, knows what position Brian is in and knows that there are plenty more like him in the local economy. With this in mind, Andrew (for purely selfish reasons) offers Brian a job in his factory. However, he requires Brian to do 50 per cent more work than existing staff and to accept 70 per cent less pay than the going rate. Brian is glad of the job, which pays him enough to survive, and enthusiastically accepts Andrew's offer. However, Brian later complains to his co-workers and friends that he is being exploited.

Andrew and Brian looks like a classic case of exploitation. Our intuitive concern is that Brian is overworked and underpaid (and also that his labour is, in some sense, forced). However, he appears not to be harmed, but rather to benefit overall, since his only alternative to taking the job was death by starvation. Indeed, as Wertheimer suggests, there's a good chance that Brian stands to gain much more than Andrew and that that's why Brian is keen to accept the job:

I suspect that the *exploitee* often gets much more utility from a transaction than the *exploiter*. It is precisely because the *exploitee* stands to gain so much from the transaction (relative to the *exploiter*) that his bargaining position is comparatively weak.<sup>6</sup>

We appear, then, to have a counter-example to the harm condition (the view that *exploitees* are necessarily harmed by exploitation) since Brian seems both to benefit and to be exploited.

How might defenders of the harm condition respond to the existence of cases like this? One option is to insist that Brian is not really exploited. Someone might, for example, argue that 'market forces' always generate just pay levels and, hence, that he isn't really underpaid. Or someone might argue that fully consensual arrangements can't be exploitative and that, since Brian consents, he's not exploited (even if he is underpaid). The first of these arguments – in particular, the assertion that 'market forces' alone always generate just pay levels – is implausible. For we can simply change our case, bit by bit, gradually making Brian's pay and conditions worse and Andrew's behaviour more repulsive. At some point, most reasonable people will concede that Brian is underpaid. And those who won't make this concession needn't be considered here, because their claim that 'the market' always generates fair pay is either a mere stipulation, or driven by a theory which is likely to be 'exploitation sceptical': i.e. one in which the moral concept of exploitation does not

have a place. The consent argument is perhaps more promising. The difficulty in Andrew and Brian, though, is that the status of Brian's consent is unclear. Of course, he consents in *some* sense. But is his consent valid? Is it of a high enough quality to render the arrangement non-exploitative? We'll be returning to questions about consent in the next chapter.

#### 4.2 Harm to interests

Another way of trying to reconcile the harm condition with cases like Andrew and Brian is to appeal to a more complicated understanding of harm. Wolff, for example, would concede that Brian is better off with the job than without it, but nevertheless wants to retain something like the harm condition:

while exploitation involves making someone worse off, it cannot be defined in terms of making someone worse off than they would have been without the exploitative arrangement. But worse off than what? The best alternative is the view that the person is made worse off than they ought to be.<sup>7</sup>

In order to make sense of such claims, we need to look more carefully at the concept of harm. Feinberg's work is a good place to start.<sup>8</sup> He usefully distinguishes three different senses of 'harm'. First, 'harm' can be used in an extended sense to apply to inanimate objects. In this sense, it means pretty much the same as 'damage' or 'destroy':

By smashing windows, vandals are said to harm people's property; neglect can harm one's garden; frost does harm to crops. Quite clearly this is harm in only a transferred sense; we don't feel aggrieved on behalf of the windows or the tomatoes, nor are they the objects of our sympathies.<sup>9</sup>

The word 'harm', used in this sense, is both non-moral and unconnected (or, at least, not directly connected) to the welfare of persons, or other sentient beings. For these reasons, we can disregard this first sense.

The second sense of 'harm' is *harm-to-interests*. Feinberg defines this as 'the thwarting, setting back, or defeating of an interest'.<sup>10</sup> Interests can be set back by a wide variety of things, including accidents, 'acts of God', illness, malicious attacks by persons, and others' (or one's own) negligence. Hence, to say that someone has been subjected to harm-to-interests isn't necessarily to make a moral judgement, because people can be harmed in this sense (by a falling tree or a virus, for example) without being in any way wronged. Feinberg does, however, talk of being harmed 'in the legal sense'. This kind of harm (which falls within the more

general category, harm-to-interests) occurs when one's interests are 'invaded' by another person:

One person harms another in . . . [this] sense then by invading, and thereby thwarting or setting back, his interest. The test . . . of whether . . . an invasion has in fact set back an interest is whether the interest is in a worse condition than it would have been in had the invasion not occurred at all.<sup>11</sup>

For all kinds of harm-to-interests the criterion is the same. As Feinberg suggests above, X has harmed me *if and only if one or more of my interests are in a worse condition than they would have been if it weren't for X.*

### 4.3 Relevant baselines

Harm, then, is a *comparative* concept. Judgements about harm compare someone's *relative* levels of welfare in two actual or possible situations. Where these two comparators are actual, we are looking at a welfare differential *over time*. For example, we might say that someone has been harmed by an event, meaning that she is worse off *after* the event (and because of it) than she was *before* it. However, more often than not, statements about harm involve comparing the actual world with a *merely possible* world: the idea being that someone is harmed by X (an event, say) if *she would have been better off if X hadn't happened*. Often, disputes about whether a thing is harmful, or about how harmful it is, centre on the question of which alternative possible world should be used as a comparator. Wertheimer offers us a good example of this. What if a psychotherapist sexually exploits a patient, but that patient (all things considered) is better off within this exploitative relationship than she would have been if she'd never had a psychotherapist at all? Has the patient been harmed? Wertheimer writes:

it is not clear against what baseline we should evaluate the claim that the patient has been harmed by her psychotherapist. If we adopt a *pre-treatment* baseline, we say the patient is harmed only if she leaves the therapeutic encounter worse off than when she entered therapy. If we adopt a *normative* baseline, we say that the patient is harmed if she leaves the therapeutic encounter less well off than she could otherwise have reasonably expected.<sup>12</sup>

Clearly the use of pre-treatment baselines (to generalise, *pre-interaction* baselines) is unsatisfactory in such cases. There are two main reasons for this. First, thinking of benefits and harms in this way encourages too much *aggregation*, leading to counter-intuitive results. For example: what

if S, a technically excellent surgeon, uses her skills to take P's (a patient's) welfare up from level 1 to level 9 but then, through sexual (or other) abuse, takes P's welfare back down to level 3? Should we say that S has harmed P? Use of the pre-treatment baseline suggests *no* because P used to be at welfare level 1 and now she is at level 3, which is an improvement, not a reduction. But surely this is the wrong description, and what we should say instead is that S has benefited P in one way, but harmed P in another. In other words, we should *disaggregate* the benefits and harms and insist on seeing the surgery and the abuse as two independently evaluable actions. For although S has benefited P by the use of surgical skills, S has *also* harmed P by abusing her. A second reason for rejecting the pre-treatment baseline is that there is a range of counter-examples based around such things as P's improving naturally, or P's improving because of third-party intervention, which render the pre-interaction baseline understanding of harm implausible. For example: P is unwell and at welfare level 2, but will if left untreated get better naturally, rising back to level 10 within a year. P goes to D (a doctor) seeking treatment. D intervenes and the effect of D's intervention is that P's rate of improvement is made slower, so that after one year P is only at level 5. Has D harmed P? Obviously *yes!* But again, use of the pre-treatment baseline gives the wrong answer (in this case, *no*) on the grounds that P ends up better off (level 5) than she was when she met D (level 2).

The use of pre-interaction baselines does seem in some cases to yield intuitive results. I would suggest, however, that all of these cases are ones where the pre-interaction baseline happens to coincide with one or both of the other two sorts of baseline: *closest possible world* baselines and *normative* baselines.

When we use a closest possible world (henceforth, CPW) baseline to assess whether harm has been caused, we are comparing B's *actual* level of welfare with the level of welfare that B *would have had if A hadn't acted, or hadn't occupied the role that she did, or hadn't existed at all*.<sup>13</sup> Whether the pre-interaction baseline coincides with the CPW baseline often depends on what would have happened if (for example) A hadn't been born. In many cases what would have happened is that B's level of welfare would have stayed at its pre-interaction level, in which case the two baselines (CPW and pre-interaction) coincide. For example, if a stranger assaults me there's a fair chance that, *had she not existed at all*, my level of welfare would have remained more or less 'flat': i.e. at roughly its pre-interaction level. Pre-interaction and CPW baselines are much less likely to coincide where A interacts with B mainly because A occupies a particular social role, such as being a member of a profession. For in these cases, the relevant possible worlds in which A does not exist will probably be ones in which A's role is not empty, but occupied by another role occupant, A\*. Hence, pre-interaction baselines and the CPW baselines often come apart,

since (if we use the latter) we assess whether or not B has been harmed by comparing her actual level of welfare with the level of welfare that she would have had if she'd interacted with A\* instead of with A. And since there's not usually any reason to suppose that A\*'s treatment will have no net effects on B's welfare, there's usually no reason to suppose that the pre-interaction baseline will be the same as the CPW baseline.

In order to see better how this idea works, let's turn back to Wertheimer's psychotherapist case. In this case, use of the pre-interaction baseline indicates that the patient has not been harmed, because she is better off after 'treatment-plus-sexual-exploitation' than she was beforehand. She is also, we can assume, better off relative to *some* possible worlds, because she is better off than she would have been if she'd never had a psychotherapist at all. So we *could* say that she's not harmed because treatment-plus-sexual-exploitation is better for her than *neither-treatment-nor-sexual-exploitation*. However, to use comparators like this is to invoke the wrong alternative possible worlds. For surely, it is (at least in the standard case) plausible to say that if the patient hadn't had psychotherapist A (for example if A had never been born) then she would instead have had psychotherapist A\*. Hence, in assessing harm here, we should be comparing A's performance with that of A\*, not with a world in which A's role is unoccupied. And, on the assumption that A\* is competent and is not a sexual exploiter, this means comparing treatment-plus-sexual-exploitation (the actual world) with treatment-without-sexual-exploitation (the possible world in which A\* is the therapist). This will, presumably, yield the conclusion that the patient really is harmed, since A is worse for her than A\*.

CPW baselines, then, are generally preferable to pre-interaction ones for the purposes of assessing harm claims. However, there is a problem with CPW baselines, one which – at the very least – shows that we can't use them uncritically, or in every case. Consider, for example, the following:

Ian is treated at Stokehampton Hospital by Dr Van Zyl. Dr Van Zyl is seriously incompetent and his incompetence causes Ian to become moderately disabled. Subsequently, Dr Van Zyl is struck off, is forced to leave his job and never practises again. When Dr Van Zyl was interviewed for his post at Stokehampton he narrowly beat another candidate, Dr Trawlerman, who is (unbeknown to the authorities) not only less competent even than Dr Van Zyl, but also a serial killer who routinely murders his patients. So (let's say) the following conditionals are true. First, if it weren't for the presence of Dr Van Zyl, Dr Trawlerman would have treated Ian. Second, if Dr Trawlerman had

treated Ian, Ian would have been murdered. (And, for completeness, let's specify that Ian has a 'life worth living', despite his disability.)

Cases like this reveal a problem with using CPW baselines in certain cases. For if we make a harm assessment here by comparing the actual world with the closest possible world, we must conclude that Ian is *not* harmed by Dr Van Zyl. On the contrary, he is benefited, because if it weren't for the existence of Dr Van Zyl, Ian would have been killed by Dr Trawlerman. This analysis is wildly counter-intuitive. For what we want to say is that Dr Van Zyl clearly harms Ian, regardless of the fact that he is, in a sense, 'shielding' Ian from Dr Trawlerman. Furthermore, this case reveals a structural feature which makes the use of CPW baselines problematic (at least in certain types of case). The problem is that using a CPW baseline in cases like Van Zyl and Trawlerman makes the question of whether or not Ian is harmed rest on a seemingly irrelevant and overly distant factor: namely, who would have got Dr Van Zyl's job if he hadn't got it. This is problematic in at least two ways. First, it just seems not to be a relevant consideration; in order to know whether or not Dr Van Zyl has harmed a particular patient we do not need to know who came second at the interviews. Second, allowing the alternative candidate to count as the comparator, for the purposes of harm assessment, can either generate impossibly high standards or let incompetents off the hook far too easily. For say that Trawlerman is the comparator and is both a spectacularly incompetent physician and a serial killer. If this is the case, then it will be almost impossible for Van Zyl to harm *any* of his patients, since it's almost impossible for him to be even worse than Trawlerman. Conversely, if the situation is reversed and Trawlerman is a sort of 'wonder-doctor', the best in the world, then it will be almost impossible for Van Zyl to avoid harming his patients – since, even if he does his best, he will be worse than (and hence harm them relative to) Trawlerman.

Faced with these problems, the attractions of using a *normative* baseline seem compelling.<sup>14</sup> Put in its most stark form, the normative construal of harm says that *A harms B if and only if A causes B to be worse off than she ought to be*, with the normative baseline being the *level of welfare which B ought to have had*. Both 'cause' and 'ought' hide a multitude of sins here. Many of these will have to be ignored, although some of the ones relating to causation will be picked up on later: for example, the question of whether omissions can cause harm. Normative baselines often coincide with one or both of the other types of baseline. Take a case in which A's harming of B violates one of B's negative rights: for example, where A wrongfully assaults B. Here the normative baseline coincides (in the standard case) with the pre-interaction baseline. This is because the normative baseline is B's level of welfare in a world in which A acts in accordance with his relevant duties. And since this is a world in which A

refrains from interacting (so as to affect B adversely) the normative and pre-interaction welfare levels turn out to be the same. Or take a case in which A's harming of B breaches one of B's positive rights: for example, B is entitled to competent treatment from A, but A fails to be competent. Here, the normative baseline will often coincide with the CPW baseline, because it's often true that if B hadn't been treated by A, she would have been treated by a competent person instead: i.e. by someone who would have given B the level of welfare that she ought to have had.

Ought we to conclude, then, that normative baselines are always to be preferred and that the apparent appropriateness of other kinds of baseline is entirely parasitic on the fact that they coincide with the normative baseline in all those cases for which they look plausible? Unfortunately, things are not quite that straightforward. For there are at least two kinds of case in which we can't or shouldn't use normative baselines.

One is where harm is justly inflicted on B by A. In such cases (by definition) A does not make B worse off than she ought to be. One could, of course, deny that there are any cases in which harm is justly inflicted, by denying either that the standard counter-examples involve real harm or that they involve real justice. But such a view seems counter-intuitive and unmotivated. The other type of problem case (a very common one) is where the harm in question is not caused by a moral agent. As was noted earlier, interests can be set back by a wide variety of things, including accidents, natural disasters, and illness. In these cases, using a normative baseline to assess the harm claim makes no sense. For how can, say, a virus make someone worse off than she ought to be (leaving aside biological weapons and such things)? Here, we have to fall back on the use of CPW or sometimes even pre-interaction baselines.

So my conclusion is that things are somewhat messy. Usually, the claim that one person has harmed another should be assessed by reference to a normative baseline. However, sometimes, such claims (for example, about 'just harm') must be assessed instead by reference either to pre-interaction or to CPW baselines. The same goes for claims that a non-person has harmed a person. These must be assessed by reference to pre-interaction, or to CPW baselines, because there is quite simply no normative baseline available in such cases (since there are no relevant duties).

#### 4.4 Omissive harm

The idea that if we are able to change things, to elect not to do so is also to determine what will happen in the world, is very old indeed . . . In such cases, many men have found it natural not only to blame those who could have prevented the harm, but did

not do so, but also to think of such men as having brought the harm about, as being its cause.<sup>15</sup>

As we have seen, assessing harm claims is less straightforward than it at first appears, because of the difficulties associated with selecting an appropriate welfare baseline as a comparator. Things are even more difficult when it comes to the claim that someone is harmed by an omission. Consider the following example:

#### Lawrence and Miss Ellis

Lawrence's school music teacher, Miss Ellis, is moderately incompetent. What she does, she does just about satisfactorily, but she completely fails to teach him anything about music history before 1900, or to develop his aural skills. When Lawrence takes his exams, he becomes aware of Miss Ellis's inadequacies. For he fails his history exam and his aural tests. He had been planning to apply to music college, but is now too demoralised to retake his school exams. He never gets to music college and his parents are dismayed about the fact that his promising career as a musician has been destroyed by what they regard as Miss Ellis's gross negligence.

What should we say about harm in this case? On the assumption that Lawrence would have been better off if he'd been a successful musician, and that he had enough 'potential' to become one, the intuitive answer is that he has indeed been harmed by Miss Ellis's inadequacies. In order to explain why, we need to think about relevant comparators. One comparator, the pre-interaction one, is *not having a music teacher*. If we use this one then Miss Ellis hasn't harmed him at all, as she's certainly better than nothing. Better comparators, though, would be possible worlds in which Miss Ellis is *non-negligent*, or *competent*, or *does what she ought*. And this is because we want to use a normative baseline and say that Miss Ellis has harmed Lawrence by *failing to do her duty*.

Miss Ellis harms Lawrence *without* exploiting him, whereas Andrew (in our earlier case) is an exploiter. In other important respects, though, the case of Lawrence and Miss Ellis is similar to that of Andrew and Brian. In particular, we can say of both cases that the question of whether harm has been inflicted depends crucially on whether we take as our relevant comparator (a) a possible world in which the putative harmer does nothing, or is simply non-existent, or (b) a possible world in which the putative harmer does more, or performs better, than she does in the actual world. In each case, if we take the first kind of comparator to be the relevant one, then the putative harmer turns out not really to have caused harm at all. For Brian would be *even worse off* if it weren't for Andrew's job offer and

Lawrence would be *even worse off* if he didn't have a music teacher at all. However, if we use the second kind of comparator, we get a different result. We've already seen how this works in the case of Lawrence and Miss Ellis: Lawrence is harmed in that he is worse off than he would have been if she had taught him competently (which she should have done). How this might apply in Andrew and Brian is more contentious. However, on the assumption that Brian is underpaid (an assumption we've been making thus far) we might say that Brian is harmed in the following sense: he is worse off than he would have been if Andrew had paid him a *fair wage* and demanded *fair hours*.

We can, then, distinguish, *within* harm-to-interests, between *active* and *omissive* harm. To be actively harmed is to be harmed by someone's actions and is what we were discussing in the previous section. To be omissively harmed, on the other hand, is to be harmed by someone's omissions.<sup>16</sup> An omission, here, is not a mere non-doing. Otherwise, I would count as being omissively harmed every time someone didn't benefit me. Rather, 'omission' is a moral(ised) term referring to someone's failure to comply with a duty to act, or to perform at or above a certain level.<sup>17</sup> Professional negligence, underpayment (for example, by employers), and failing to rescue can all be examples of omissive harm. For, in each case, the 'victim' might reasonably say that she has been harmed by others' failure to do what they ought.

The extent to which one deploys the concept of omissive harm will depend on how many, and what, positive duties one thinks there are. This is especially clear in the 'good (or bad) Samaritan' cases, frequently discussed by moral philosophers: cases in which a passer-by decides to rescue ('good Samaritan') or not ('bad Samaritan') a stranger who is drowning or who is injured at the side of the road. One of the main questions about such cases is: does the bad Samaritan, by not saving the stranger's life, *harm* the stranger, or does she merely *not benefit* him? If we apply what I have said so far about omissions and omissive harm, we get the following (admittedly schematic) answer. The bad Samaritan *omits* to save and therefore *omissively harms* the stranger only if she has a duty to save her. Otherwise, the bad Samaritan's not acting is a *mere non-acting*, rather than an *omission*, and so she merely fails to benefit the stranger. Hence, one could (and, on my view, should) think one of two things. Either one could believe that there is a duty to rescue, an omission and omissive harm. Or alternatively

One might be tempted to argue that bad samaritans are not omit-  
ters at all, but only nondoers . . . On this account, an omission is  
the nondoing of an act one had a duty to do, and duties come only  
from special relations to the endangered party. The bad Samari-  
tan, having no such relationship to the endangered party, had no

duty to assist him. Therefore, although he did not assist him, it is false that he *omitted* to do so.<sup>18</sup>

I can't hope to settle the issue of whether there is or is not a general duty to rescue here. My point, rather, is that those of us who see the moral landscape as densely populated by positive duties will and should – on the view presented – be much more inclined to ascribe omissive harm than those 'minimalists' who think that positive duties only arise within a very narrow range of special relationships.

Turning our attention back to exploitation, we can now see how Wolff's suggestion that 'exploitation involves making someone worse off than they ought to be', combined with the idea of omissive harm-to-interests, might be used to rescue a version of the harm condition.<sup>19</sup> The argument goes as follows. It was initially suggested that, necessarily, exploitees are harmed by exploitation (the harm condition). A counter-example to this, Andrew and Brian, was then provided in which Brian (the exploitee) appeared to be better off exploited than not. However, we are now in a position to argue that, even if Brian has not been *actively* harmed, he has nonetheless been *omissively* harmed, because he is worse off than he would have been if Andrew had paid him a fair wage. (Alternatively, by describing the situation in an appropriate way, it might be possible to argue that he has been actively harmed relative to a comparator baseline in which he is – for example – not made to work excessive hours.)

Importantly, this manoeuvre is not just a way of neutralising one particular counter-example. For it can also be a general strategy for establishing the existence of a necessary connection between exploitation and harm, by generalising the following argument:

- 1 Andrew only counts as an exploiter because there is *disparity of value*.
- 2 What disparity of value means (or, at least, entails) in this case is that Andrew breaches his duties towards Brian (for example, by not paying him enough or making him work too much). If there were no breaches of duty, there would be no disparity of value.
- 3 Therefore, given that, because of Andrew's actions or omissions, Brian is *worse off than he ought to be*, Andrew must have harmed Brian.

So the idea is, in brief, that exploitation entails disparity of value, which entails a breach of duty, which entails harm (relative to a normative baseline) and that, therefore, the occurrence of harm is a necessary condition for the existence of exploitation.

Even if the validity of this argument is established, the version of the harm condition that it saves only applies to disparity of value

exploitation. So if a link between harm and the other kind of exploitation, instrumentalisation, is to be established, a quite separate argument will be required. (One such argument, concerning 'moral harm', is considered in the following section.) Furthermore, by broadening the notion of harm to include omissive harms and harms judged against normative baselines, this (modified) harm condition states something less controversial and remarkable than did the original version. In other words, the modified harm condition is relatively 'weak', because it still allows that there are cases of 'mutually beneficial exploitation', if by this we mean cases in which both the exploiter and the exploitee are better off than they would have been if there had been no exploitation. In these cases, the exploitee is not harmed relative to pre-interaction or closest possible world baselines and is only harmed (actively or omissively) relative to a normative baseline.

#### 4.5 Moral harm

[I]t may be argued that the[re is] a deeper – Kantian – way in which what I have called mutually advantageous exploitation is actually harmful to B [the exploitee], namely, that A treats B merely as a means to be utilized to his own advantage rather than an end in herself. *And so treating a person is a harm to her.*<sup>20</sup>

In Chapter 3, I introduced the case of Manuel Wackenheim, a (so-called) dwarf who earned a living by being 'tossed' by customers in bars and nightclubs. This 'tossing' formed part of a *dwarf-throwing competition* – a sport 'in which the aim of the competitors is to fling a dwarf over the furthest distance possible'.<sup>21</sup>

One view of dwarf throwing is that it is not wrong provided that it is consensual and harmless, in the harm-to-interests sense. Another is that it is a case of harmless wrongdoing; for example, one might concede that dwarf throwing is harmless but nonetheless invoke the idea of instrumentalisation, along with related notions such as degradation and (in)dignity. A third view is that dwarf throwing is harmful in the third sense of 'harm' noted by Feinberg:

To say that A has harmed B in this sense is to say much the same thing as that A has wronged B, or treated him unjustly. One person *wrongs* another when his indefensible (unjustifiable and inexcusable) conduct violates the other's rights.<sup>22</sup>

So someone might say that Manuel Wackenheim has been harmed, in this third sense, meaning that his rights (which will need, for this to work, to be *inalienable*, given that he seems to want to waive them) have been

violated. Such (supposed) rights could include his right to dignity and his right to respect as a person.

At this point, though, we should ask: even if his rights have been violated, what is the point of calling any violations *harm*? If the rights in question exist and have been violated, then, of course, Wackenheim has been *wronged*. But using 'harm' to refer to wrongs which are not harmful in the standard (harm-to-interests) sense seems counter-intuitive and not to add anything to our understanding of these cases. Furthermore, and more importantly, using 'harm' in this expanded way is potentially very confusing, obfuscating as it does the distinction between wronging people by harming them (in the standard harm-to-interests sense) and wronging them in other ways. It would be much better, I suggest, to leave open the possibility that these cases are cases of harmless wrongdoing, reserving the word 'harm' for harm-to-interests, and rejecting talk of purely moral harm.<sup>23</sup>

#### 4.6 Summary and conclusions

Practices involving the commercialisation of the body are frequently accused of being exploitative and of being harmful and, more often than not, these exploitation and harm claims accompany one another. With this in mind, Chapter 4 aimed to improve our understanding both of the nature of harm and of the relationship between harm and exploitation.

The discussion of the relationship between harm and exploitation centred around the harm condition, which states that exploitation necessarily involves harm to the exploitee. Evaluating the harm condition meant engaging in a deeper analysis of the concept of harm. The analysis started by drawing on Feinberg's tripartite division of harm into (a) the extended sense (under which *harm* means simply *damage*), (b) harm-to-interests, and (c) moral harm. The extended sense was deemed largely irrelevant, while the idea of purely moral harm was rejected on the grounds that using 'harm' in this way is confusing, as it obscures an important distinction between wronging people by harming them and wronging them in other ways. The focus for the remainder of the analysis therefore was harm-to-interests, which Feinberg defines as 'the thwarting, setting back, or defeating of an interest'.<sup>24</sup> People are harmed, in the harm-to-interests sense, if they are made worse off. However, lurking behind this seemingly simple idea are two thorny questions. What counts as *being* worse off? And what counts as *making someone* worse off?

The first of these questions could be restated as: worse off *than what*? For harm-to-interests is a comparative concept and, in order to know whether or not someone has been harmed, we need to know what to compare her actual level of welfare to; we need a welfare baseline to use as a comparator. Three types of baseline – pre-interaction, closest possible



world, and normative – were discussed in some detail. It was concluded that things are ‘messy’. More specifically, while *usually* the claim that one person has harmed another should be assessed by reference to a normative baseline, *sometimes* such claims should be assessed instead by reference either to pre-interaction or to closest possible world baselines.

The second question – what counts as *making someone worse off?* – is principally about whether we should allow both actions and omissions to count as harmful, or whether we should say instead that only actions can be harmful. Our conclusion was that there *is* such a thing as omissive harm, being harmed by someone’s omissions. Omissions, however, are not mere non-doings, otherwise I’d be omissively harmed every time you decided not to benefit me. An omission is, rather, a failure to comply with a relevant (‘positive’) duty to do something, or a duty to do something at or above a certain level (for example, of competence).

The idea of omissive harm-to-interests can be used to rescue a version of the harm condition. According to this version, the following necessary connection between (disparity of value) exploitation and harm exists. In order for there to be a disparity of value, the exploiter must (a) *actively harm* the exploitee *and/or* (b) fail to give the exploitee something which she ought to give her (for example, by underpaying her). Obviously (a) involves harm, but (b) also involves *omissive harm* as defined above. Hence, one way or the other, exploitation always involves harm.

There is, then, a version of the harm condition which appears plausible. However, the plausible version seems ‘weaker’ and a good deal less interesting than what we started out with. For it relies on the acceptance of a fairly broad notion of harm – in particular, one which uses normative baselines and allows omissions to count as harmful. As was argued earlier, there are good reasons for having such an understanding of harm, and so these points are not objections to the ‘weakened’ harm condition. However, we should bear in mind that, even with this ‘weakened’ harm condition in place, there is still an important distinction to be made (within exploitation) between *exploitation which makes the exploitee worse off than she would have been if she hadn’t been exploited* and *exploitation which makes the exploitee better off than she would have been if she hadn’t been exploited*. It is tempting to call the former *harmful* exploitation and the latter *mutually advantageous* exploitation (assuming that the exploiter also gains). But this may be misleading, given that cases in the latter category will also turn out to be harmful if normative baselines are used.

I’d suggest, however, that we can still legitimately use the expression *mutually advantageous exploitation* to refer to cases of the second kind, even though such cases are harmful to the exploitee when measured against a normative baseline. It’s important for us explicitly to recognise this category for two reasons. First, the fact that an instance of exploitation is mutually advantageous (in this sense) is often morally relevant, especially

in the context of arguments about legal prohibition. For what we might sensibly say about some such cases is that the exploitation should be allowed because the exploitees are *better off with it than without it*. (Another, forceful, way of making the same point is to say that to ban it would harm the exploitees, relative to a pre-interaction or CPW baseline.) Second, the fact that some instances of exploitation are mutually advantageous plays an important role in the minds of prospective exploitees, for they will often have strong prudential reasons for subjecting themselves to exploitative arrangements of this kind. And this may itself have moral significance, for it is at least worth taking seriously the possibility that exploitation which is rationally consented to and desired by the exploitee is morally preferable to exploitation which is forced and unwanted. These matters, including the issue of consent, will be looked at in the next chapter.

So to sum up, the relationship between exploitation and harm is as follows. It is right to classify *all* (disparity of value) exploiters as harmers. This follows from the broad understanding of harm which I have argued for in this chapter – an understanding which includes omissive harm and harm measured against normative baselines. I take this to be not merely a trivial or linguistic claim but – especially given the importance of harm in our moral-conceptual landscape – an interesting and substantive moral truth, grounded in proper understandings of exploitation and of harm. For example, it has practical implications, such as that exploiters cannot (or should not) defend themselves by claiming that what they are doing is entirely harmless. We should keep in mind, though, that, while all exploitation is harmful, there is an important distinction to be drawn between what I earlier called ‘mutually advantageous exploitation’ and other forms of exploitation, and that mutually advantageous exploitation (for reasons explained above) will often be less morally objectionable than other forms.

## 5 CONSENT

When exploiting friends or lovers, we are not *forcing* them to do anything whatsoever. They do what they do happily, willingly, voluntarily.<sup>1</sup>

The idea of consent, or valid consent, has both a primary and a secondary role within the body commodification debate. The primary role is that it's thought by many to explain directly the wrongness of certain commercial practices. For example, it has been argued both that surrogate mothers and that people who sell parts of their bodies (for example, kidneys) don't, or can't, validly consent. For this reason (so the argument goes) such practices are wrong, since interactions that are as intimate, or as dangerous, as these ought to be properly consensual. Consent's secondary role is the contribution that it makes to the justification of exploitation claims. For, as we've already seen, most (maybe even all) instances of exploitation appear to involve a consent that is somehow defective or inadequate.

This chapter addresses the following questions. First, what is valid consent? Second, what factors can invalidate a consent? And third, what is the relationship between consent and exploitation? Specifically, is genuinely consensual exploitation a possibility, and (conversely), is defective (or absent) consent a necessary condition for a transaction's being exploitative?

### 5.1 Absent consent and defective consent

I've already suggested that there's an important link between defective consent and exploitation. However, one puzzle is that we don't normally want to describe those wrongs which involve the *complete absence* of consent as exploitation.<sup>2</sup> Consent to sexual intercourse is an interesting case in point. It would be extremely odd to say (at least of relatively straightforward cases) that what's bad about rape is that it's exploitative; to do so

would both understate and misidentify the wrong involved. By contrast, though, the claim that prostitution is exploitative is commonplace. What's more, the rationale for accepting this exploitation claim is that (many) prostitutes are, in effect, *forced* into prostitution by poverty, social exclusion, etc. So the puzzle is that, although (arguably) both rape and prostitution are wrong because they involve forced sex, we want to describe only the latter as exploitation. Why could this be?

What we have here is an awkward intuition about how the word 'exploitation' should be applied. We have a choice about how to handle this intuition. Either we can get our account of exploitation to accommodate it (or, even better, show that it already accommodates it). Or we can try to show that the intuition is somehow flawed or misleading.

Perhaps the best answer of the second kind appeals to Wertheimer's idea of moral occlusion.<sup>3</sup> Rape, we might say, really is exploitative, but the exploitation involved is *occluded* by more serious and tangible wrongs, such as bodily violation and mental and physical harm. In other words, the reason why it would be peculiar to say that what's bad about rape is that it's exploitative is not that rape isn't exploitative, but rather that its being exploitative is a relatively minor evil compared to the others involved. So it's the decision to *foreground* exploitation and the suggestion that exploitation is the main evil which are strange, rather than the claim that it is exploitative *per se*. An analogous case here is murder. Often, this will involve lesser occluded evils, such as violation of bodily integrity and physical injury. These occluded evils exist. But it would strike us as odd if they were foregrounded, or if someone argued that what's wrong with murder is that it involves violation of bodily integrity, since the main evil is obviously death.

In general, the idea of moral occlusion is an interesting and useful one. However, it does seem to me that the answer it delivers here is unsatisfying. There are two reasons for this. First, other things being equal, we should prefer accounts which save the appearances, those which trust our intuitions, to those which view our intuitions as defective or misleading. Of course, other things are not always equal, and there are often good theoretical (or other) reasons for holding that things are not as they seem. However, here (as we'll see in a minute) there does seem to be a plausible appearance-saving account available. The second reason is that it's not clear that appealing to occlusion really does explain why we view rape and prostitution differently. In the case of rape, the claim is that the exploitation involved is occluded by the other more serious evils. But it is tempting to ask why the same isn't true of prostitution. For many people believe that the most serious evils involved in prostitution are similar to those involved in rape – for example, bodily violation and mental and physical harm – and that its being exploitative is a *relatively* minor matter. Nonetheless, we don't regard the claim that prostitution is exploitative as

odd and, perhaps more importantly, don't seem to have a problem with believing (even pre-analytically) *both* that it is exploitative *and* that exploitation is not the main evil involved.

Let's turn now to the other kind of answer, the one which takes our intuition at face value and attempts to explain why rape is not normally exploitative, while prostitution often is. Allen Wood suggests (below) that there is a sense in which typical victims of exploitation (such as the starving person and the indebted gambler) are likely to be all *too* willing to be exploited, all *too* willing to give their consent, owing to the absence of less bad alternatives. Hence, according to Wood, 'exploitation is often voluntary' and consensual:

Someone who is propertyless and starving has a lot to gain by striking a deal with an employer who is willing to offer bare subsistence in exchange for long, hard labour under dangerous conditions – and a lot to lose (namely, life itself) if no such exploitative bargain is in the offing . . . A gambler who owes a large amount of money to ruthless and violent characters will be in desperate need of the loan shark who offers the needed funds at a usurious rate of interest; such a person will be more than willing under these conditions to consent to virtually any terms of payment.<sup>4</sup>

These do seem to be classic cases of exploitation and are structurally identical to the case of the prostitute who may be keen to take on sex work given her poverty and lack of preferable alternatives. I would suggest, then, that Wood is at least partially right, and that the following is true of exploitation: *it takes place only where there is (at least) 'minimal consent' from the exploitee*. What I mean by this is that, although the exploitee need not (and normally does not) provide *valid* consent (a concept that we'll be looking at in more detail shortly), she does consciously *choose* the exploitative arrangement rather than some other (even worse) alternative.

To draw on Wood again, a plausible explanation of why this is so can be found in the model of exploitative use suggested in section 2.4. According to that model, exploitation necessarily involves (what Wood calls) *advantage-exploitation*, making use of 'someone's weakness or vulnerability' – such as her desires and needs – in order to gain 'a hold over the person'.<sup>5</sup> This is played out in practice by A using B's vulnerability to get B 'minimally' to consent. Entirely non-consensual uses of persons, by contrast, fail to involve advantage-exploitation, for they bypass the will, or conscious choice, altogether. For example, to rape someone does not (or need not, except in the crudely physical sense) depend on using the victim's weaknesses or vulnerabilities to gain 'a

hold over' her. The reason for this is simply that no 'hold' is required, since there is no need to make the victim choose sex and no need for her minimal consent. Rather, rape is a physical violation. Similar things can be said of burglary (which isn't normally exploitative), and this can be usefully contrasted with the actions of, say, confidence tricksters (who normally do exploit). Burglary doesn't involve advantage-exploitation. On the contrary, its victims don't even need to be present. So burglary isn't normally exploitation. Confidence tricksters, on the other hand, do rely on getting 'minimal' consent from their victims. The victim needs actually to *hand over* the cash or to *sign* the contract. Hence, the victim does need to be advantage-exploited. That is, a psychological weakness – such as fear, greed, impatience, or ignorance – is taken advantage of.

For completeness, I should add that things look very different when we turn to instrumentalisation ('wrongful use' exploitation). This is mainly because, as we saw in section 2.4, instrumentalisation needn't involve advantage-exploitation. Take, for example, the kind of objectification claims that people make about prostitution. The leading idea behind these claims is that *it is wrong to treat another person as a mere sex object* and that treating people in this way constitutes instrumentalisation. Treating another person as a mere sex object, though, doesn't require even minimal consent and can entirely 'bypass the will'. Rape, for example, does (or can) involve the victim being treated as a mere sex object and so would seem to be a case of instrumentalisation. But rape doesn't (or needn't) involve 'minimal' consent and so doesn't (or needn't) involve advantage-exploitation.

## 5.2 Valid consent

We've already seen that, as well as the distinction between consensual and non-consensual transactions, there is a further distinction – *within* the consensual – between *valid* and *invalid* consent. Whereas the distinction between consensual and non-consensual is essentially psychological or social, the valid/invalid distinction is moral. Invalid consents are perfectly *real*: for example, if I hand over my money in response to a 'Your money or your life!' threat, then I really have agreed to give you the money. But such consents don't have the same moral significance as valid consents. In particular, valid consents (for example, to sexual intercourse, surgery, or the taking of property) can morally justify acts in ways that invalid consents can't. Indeed, as I suggested in section 3.7, what it means to say that a consent is valid is precisely that it provides a justification for certain action(s) (sometimes on its own, sometimes in conjunction with other facts). For example, to say that I have *validly* consented to having my tooth removed is to say that my consent is of a sufficiently high

quality (in conjunction with other facts) to *justify* my dentist's (tooth-removing) actions.

Cases like this aren't quite the whole story, though, because one other morally significant thing that valid consents sometimes do is to create additional duties for the consentor. What I have in mind here are things like contracts and promises which involve someone's agreeing (consenting) to take on particular moral and/or legal duties, normally in exchange for something. In such cases, the consent offered must be valid in order for the additional moral duties to be generated. For example, if a promiser only makes her promise because she is lied to by the promisee, most of us would say that she has no moral duty to keep it. This is because, although we wouldn't usually put it in quite these terms, her consent (her agreeing to take on the obligation) is invalid.

So valid consents are morally significant in one or both of two ways. *Either* they justify (partly or wholly) the interventions of third parties, such as dentists and doctors; *or* they generate additional obligations for the consentor if, for example, she agrees to enter into a contract. The aim of the next few pages is to say what valid consent is. In areas like medical ethics valid consent is normally called *informed consent*. The expression *valid consent* is preferable, however, because, as we'll see, adequate information is only one part of validity, and consent can be invalidated by factors other than lack of information.

The concept of consent is widely used both in everyday contexts and by moral and political philosophers. One of the places where it has been most heavily employed and analysed in recent times is within healthcare ethics. Consider, for example, these two definitions of 'consent', both provided by bioethicists (Gillon, and Faden and Beauchamp respectively):

consent means a voluntary, uncoerced decision, made by a sufficiently competent or autonomous person on the basis of adequate information and deliberation, to accept rather than reject some proposed course of action.<sup>6</sup>

an informed consent is an autonomous action by . . . a patient that authorises a professional . . . to initiate a medical plan for the patient . . . an informed consent is given if a patient or subject with (1) substantial understanding and (2) in substantial absence of control by others (3) intentionally (4) authorises a professional.<sup>7</sup>

Plausibly, both definitions suggest that in order for a consent to be valid, three main elements must be present (or present in sufficient quantities): information, competence, and voluntariness.

Let's now take a look at these three main elements, starting with information. Medical ethicists have spent a great deal of time on attempts

to construct a substantive account of how much information a patient or research subject needs in order for her consent to be valid. I'll not try anything that ambitious here, not least because my concerns are rather more general than theirs. It is, however, worth saying something about why information matters and, schematically, about what sort of information matters and how much there should be. The reason why information matters is simply that in order for A to consent to X, A needs to know what X is and what the likely consequences of X are. Exactly how much A needs to know about X is not something that I'll investigate here. However, it seems fair to say that, at least in the standard case, A needs to know about X's main benefits and drawbacks and about any other factors which one would reasonably expect to affect A's decision about whether or not to accept X.

Two further points about information need to be made before we proceed. Both are fairly obvious but nonetheless important, not least because they are sometimes overlooked in the body commodification debate. The first is that *full* or *complete* knowledge is not required for validity. The main reason for this is that even if we allow that such a complete knowledge state is a theoretical possibility, it is certainly not a state in which humans very often (if ever) find themselves in practice. Hence, if we decided to make full knowledge a requirement, that would commit us to the (absurd) view that there are virtually no cases of valid consent. The second point (which is in many respects similar) is that *experiential* knowledge is not generally required for validity. Experiential knowledge of X is knowing what it is like to experience X, which normally means having experienced X and being able to remember experiencing X. Experiential knowledge – for example, of the pains or pleasures associated with certain things – can be very important and may well help to validate consents and, more generally, help people to make good decisions. However, it can't be a general requirement because otherwise it would be, at best, difficult for people to consent to things for the first time. Again, sexual consent provides a pretty clear case in point. If experiential knowledge were required then it would be almost impossible for virgins to consent validly to their first sexual encounters. And depending on how finely we individuate items of experiential knowledge it may, similarly, be impossible even for non-virgins validly to consent to their first sexual experience with A, with B, with C, of X, and so on. We'll be returning to the idea of experiential knowledge and consent during the discussion of surrogacy, because some people have objected to surrogacy on the grounds that the surrogate's consent is rendered invalid by her lack of experiential knowledge.

To turn now to *competence*, this term refers to the consentor's mental capacities. Is she capable of making reasonable decisions? Can she understand relevant information and issues and rationally assess the pros and

cons of what is proposed? Examples of consents being invalid because of lack of competence include consents given by young children, consents given by people suffering from mental disabilities or illnesses, and consents given by people whose minds are under the influence of alcohol or other mind-affecting drugs. As I've suggested, people in all three categories can validly consent to *some* things, depending on the degree of mental incapacity. For example, we'd view most contracts between drunks and taxi drivers or bartenders as valid, and many drunken consents to sexual intercourse as valid (except, perhaps, in very extreme cases of intoxication, or where the intoxication is itself involuntary).<sup>8</sup> Given their mental limitations, though, people in the categories just mentioned do, in general, run a substantial risk of providing invalid consent, especially for complex and/or important decisions. After all, we wouldn't, for example, take seriously an extremely drunken consent to organ donation from a prospective live donor.

As I suggested earlier, the third element of valid consent is *voluntariness*. This is not to be confused with what philosophers call *free will*. That is, you don't need to be completely free from all causal determinants in order to act voluntarily. Rather, to act voluntarily is just to be free from certain specific types of influence, notably coercion. Because coercion is such a difficult, important, and highly relevant concept, it merits a chapter of its own (Chapter 6).<sup>9</sup>

### 5.3 Consensual exploitation?

The idea of exploitation usually implies that the exploiter is able to apply some coercive pressure that those whom she exploits are unable, or ill equipped, to resist.<sup>10</sup>

So far in this chapter we've seen that, in order for exploitation to occur, the exploitee must provide at least 'minimal' consent. This is one of the main differences between exploitation and theft; theft needn't be even minimally consensual. In this section, I'd like to ask whether it's possible for an exploitee's consent to be valid or whether instead what I'll term the consent condition – the view that A exploits B only if B's consent to the arrangement is invalid – is true. I'll contend that it is, that exploitation always involves 'defective' consent from the exploitee.

Why would anyone believe the consent condition? Perhaps the best reason for doing so is suggested by this passage from Schwartz:

If I am entitled to a coat because I made it, you may still receive it justly, not because you are entitled to it, but because I give it to you. That isn't theft, nor do you exploit me in receiving the coat. The reason is that the transfer is free and unforced . . . The

problem with surplus transfer in class societies is that in general it is arguably not free and unforced. In slave and feudal societies that is clear enough. Producers hand over the surplus *or else*. Marx argues that this is true, despite appearances, in capitalism.<sup>11</sup>

For Schwartz, what differentiates the exploitative economic transactions that are endemic within capitalist societies from (say) genuine gifts is that the latter are 'free and unforced' – or, we might say, validly consensual – whereas the former are not. It seems to me not only that Schwartz is right about this, but also that we can generalise this insight and say that the presence or absence of valid consent is what distinguishes exploitation from gifts, and being exploited from being generous.

As we've seen, people who are exploited suffer some (negative) 'disparity in the value of an exchange of goods and services'.<sup>12</sup> They are (for example) under-rewarded for something which they provide for the person or institution which exploits them. However, not all cases of under-reward are cases of exploitation. If someone very poor is forced to work in a factory for \$1 per day in order to avoid starvation for her family, then she is probably being both under-rewarded and exploited by her employer. But if, on the other hand, a wealthy professional person agrees, entirely voluntarily, to spend a day working for charity for a 'token' \$1 per day, then she is (probably) being under-rewarded but *not* exploited. The key difference between these two cases is the nature of the consent that each person gives. In the former case, the validity of the consent is questionable, but in the latter case, the consent to underpayment is unproblematic. This is the thinking underpinning the *consent condition*, according to which A exploits B only if B's consent to the arrangement is invalid. In short, what the consent condition does is to allow us to distinguish cases of exploitation from non-exploitative cases in which an individual waives her rights to a fair share of the burdens/benefits. The consent condition is neutral with respect to particular conceptions of valid consent and furthermore is compatible with the view that different standards of consent are applicable in different types of case. Hence, the consent condition will only yield substantive results when allied with a particular account of valid consent.

Both Wertheimer and Wood argue against my view that exploitation necessarily involves defective consent. Consider again Wood's starving person and indebted gambler (cited above).<sup>13</sup> As we've seen, Wood says that these victims of exploitation are likely to be all *too* willing to be exploited, because of the absence of less bad alternatives, and hence 'exploitation is often voluntary' and we oughtn't to make defective consent part of our definition of exploitation.<sup>14</sup> This does seem to present a problem for the consent condition, since both of Wood's cases look like fully voluntary cases of exploitation. We should, however, bear in mind

that the proposed criterion is not *mere voluntariness*, but *valid consent*, which is a more demanding and more 'moralised' standard.

One way of dealing with Wood's cases, then, is to argue that neither the starving person nor the gambler validly consents, because of their lack of acceptable alternatives. Although this seems initially intuitive, I'd be reluctant to put too much weight on the idea that lack of acceptable alternatives is in and of itself sufficient to invalidate a consent. There are two main reasons for this. The first, a very general worry, is that it will be difficult, if not impossible, to specify non-arbitrarily what counts as a sufficient number of sufficiently good alternatives. Second, it would seem strange to say (as this proposal implies) that if someone were faced with an entirely free choice between X, which is extremely good, and Y, which is extremely bad, that that person could not validly consent to X because of the lack of acceptable alternatives. Consider, for example, this case provided by Wertheimer:

If A (a physician) should say to B (a patient), 'You can choose to have this leg amputated or you will die', we don't say that B's decision to have his leg amputated is coerced because death is an unacceptable alternative. Rather, we seek B's informed consent to the procedure.<sup>15</sup>

It would be quite bizarre to rule out the possibility of valid consent here just because of the lack of acceptable alternatives.

Much more promising is the idea that it is *coercion* which invalidates consent in Wood's cases. This looks straightforwardly plausible in the case of the gambler who is being threatened by 'ruthless and violent characters'. He is, we might say, coerced into taking the loan from the loan shark since the ruthless and violent characters will harm him if he doesn't pay them and (let's assume) getting money from the loan shark is his only means of doing so. So *de facto* they are coercing him into taking out the loan. Wood's other case, the starving person, is one which we might naturally talk about in terms of the person's being forced by poverty or by starvation. But is this really coercion, or is it merely a lack of options, or some other sort of non-coercive forcing? This isn't an easy question and, in order to deal with it, we'll need to tackle directly the question of what exactly coercion is. This is one of the main aims in the next chapter.

#### 5.4 Summary and conclusions

Valid consent has both a primary and a secondary role in the body commodification debate. The primary role is that it's thought by many to explain directly the wrongness of certain commercial practices, while the

secondary role is its contribution to the justification of exploitation claims. It has this secondary role in virtue of its relationship with exploitation. I've suggested that this relationship's key elements are (i) that A exploits B only if B minimally consents, and (ii) that A exploits B only if B's consent is somehow defective. Element (ii) is what I earlier termed the consent condition.

Valid consents are morally significant in one or both of two ways. Either they justify (partly or wholly) the interventions of third parties, such as healthcare professionals, or they produce obligations for the consentor, if, for example, she agrees to enter into a contract. In order for a consent to be valid, three main elements must be present in sufficient quantities: information, competence, and voluntariness. Occasionally (and notably in the debate about commercial surrogacy) consent objections to body commodification are based on concerns about (the absence of) competence or information. However, more often than not, they're based on concerns about voluntariness, especially coercion. For this reason, and because coercion is conceptually difficult, Chapter 6 will investigate coercion in more detail.

## COERCION

Both coercion and deception infringe upon the voluntary character of the agent's actions. In both cases a person will feel used, will see herself as the instrument of another's will.<sup>1</sup>

Whenever coercion takes place one will is subordinated to another. The coerced is no longer a completely independent agent. If my will is overborne by yours I serve your ends and not mine. I am motivated by your interests and not mine. I do what you want, not what I want.<sup>2</sup>

This chapter aims to provide an analysis of the concept of coercion. Accusations of coercion often underpin attempts to show that particular commercial practices involve the wrongful use of people, or the violation of their autonomy, or that any consent given is invalid. More specifically, it's sometimes said that only people who were coerced would 'volunteer' to sell a body part, or to provide 'sexual services', or to be surrogate mothers. For, as Kligman and Culver put it,

an offer of payment can take on some of the qualities of manipulation, or even coercion; for example, if the offer is grossly unfair, or one party has no way of accurately assessing the value of his services, or one's need for the offered sum is so pressing that one is not genuinely free to refuse the offer.<sup>3</sup>

### 6.1 Introducing coercion

Unfortunately, confusion over what the term 'coercion' stands for has not stopped people from using it.<sup>4</sup>

Coercion is ethically significant in a number of different ways.<sup>5</sup> First, 'coercion', more often than not, is used as a negative moral term.<sup>6</sup> It suggests a particular kind of moral criticism, and coercion's *victims* regularly

*complain*, in moral terms, about having been coerced. Second, being the victim of coercion normally 'serves to nullify or mitigate one's legal or moral responsibility for one's actions'.<sup>7</sup> We regard coerced actions as forced and involuntary. Hence, someone who is coerced into breaking a moral or legal rule can, in her defence, reasonably say something like 'I couldn't help it – I had no real choice'. The actions of the person who is coerced are, as Dworkin puts it,

in one sense hers because she did them [but] . . . in another sense attributable to another. It is because of this that such infringements may excuse or (partially) relieve a person of responsibility for what she has done.<sup>8</sup>

Finally, the use of coercion to gain consent typically renders any resultant consent invalid, usually with important ethical implications.

Let's start our analysis by considering the following example.

An armed woman breaks into my house and threatens to kill my beloved cat, Tiddles, if I don't give her \$5,000. I decide to give her \$5,000, since I would undoubtedly prefer to be \$5,000 worse off than to watch poor Tiddles be murdered.

In this case, there clearly *is* a sense in which I freely decide to pay \$5,000 rather than watch Tiddles die. I do, after all, act on a genuine preference and could have decided to keep the money instead. However, we probably *would* want to say of the Tiddles case that my decision to hand over the cash wasn't truly voluntary because I was coerced, and that (a) the cat killer's actions made me less free than I otherwise would have been, and (b) I was not fully responsible for my actions, so that even if giving the cat killer money were wrong (for example, because she would spend it on weapons, or on 'hard' drugs for children) I shouldn't be held (fully) responsible.<sup>9</sup> We would also probably want to say that any consent given by me under such coercive conditions was invalid. This can be seen clearly in our intuitive responses to the following questions. First, if I took the cat killer to an ATM to get the \$5,000 that I 'promised' her and got a chance to escape without paying up, would I be morally obliged to refrain from escaping and to give her the money? Second, does my coerced 'consent' to her taking the \$5,000 in any way justify her taking it? In both cases, we would (rightly) say *no*. This is because of the invalidity of my consent – a 'minimal' consent which neither justifies the cat killer's actions nor morally constrains mine.

So what exactly is coercion? In fact there are two different, though intimately related, questions here. One concerns the *coercer* and is about

what counts as *coercing* or, more accurately, as *acting coercively*; the other concerns the *coercee* and is about what counts as having been successfully coerced (which is typically contrasted with being free or acting freely).

The most plausible answer to the first question (though one that will have to be refined later) is that A coerces B when A threatens to harm B if B doesn't do what A wants. As Wertheimer puts it,

In general, A coerces B to do X if A proposes (threatens) to make B worse off with reference to some baseline condition unless B does X.<sup>10</sup>

Hence, we can say that the cat killer coerces me because she threatens to harm me (by killing my cat) if I don't do what she wants (hand over the cash). On this view, which I regard as fundamentally correct, the account of coercion is based principally on a distinction between threats and offers, with coercion being equated with threatening. We'll look soon at whether *all* threats are coercive, or whether instead threatening is merely a necessary condition for coercing, while, in section 6.3, we'll be taking a closer look at how to draw the required distinction between offers and threats.

A plausible answer to the second question (the one about the *coercee*) is offered by Gerald Dworkin. A puzzle which concerns Dworkin is

why coercion is thought of as a way of getting someone to do what he doesn't want to do rather than a way of getting someone to want to do something else.<sup>11</sup>

This relates to the idea (mentioned above) that when I hand over the \$5,000 I do so because I *prefer* life without the \$5,000 to life without the beloved Tiddles. This is clearly true, for I *could* have chosen Tiddles's death instead if, say, I valued her life at only \$3,000. Nonetheless, the intuition that the cat killer is making me hand over the money *against my will* remains strong.

What Dworkin (and I) think this shows is not that all actions (even coerced ones) are voluntary but rather that our account of being coerced cannot depend on a distinction between acting on one's own desires and acting on something other than one's own desires – because even if my desires are caused by coercion, they are still *mine*. In 'Acting freely' (1970), quoted above, Dworkin's alternative account of coercion is based on the negative attitudes of the *coercee* to the desires on which she acts. More specifically, the account is 'in terms of the resentment or aversion men have to acting for certain reasons'.<sup>12</sup> So, of the cat killer case, we could (employing Dworkin's view) say that I am successfully coerced because I

am caused to act on reasons which I *resent* having to act on (for example, that I prefer preventing Tiddles's murder to saving \$5,000).

In Dworkin's later work, notably *The Theory and Practice of Autonomy* (1988), this account of coercion is made more general and fitted into a wide-ranging theory of autonomy. On this later view, coercion is one of a number of ways in which someone's autonomy can be disrespected or undermined. Coercion does this in one or both of two ways. It may cause the *coercee* to act on desires which she does not reflectively endorse (at the level of her 'second-order' desires). Alternatively, it may lead to 'failure of procedural independence', which Dworkin describes in general terms as follows:

Spelling out the conditions of procedural independence involves distinguishing those ways of influencing people's reflective and critical faculties which subvert them from those which promote and improve them. It involves distinguishing those influences such as hypnotic suggestion, manipulation, coercive persuasion, subliminal influence, and so forth, and doing so in a non ad hoc fashion. Philosophers interested in the relationships between education and indoctrination, advertising and consumer behaviour, and behaviour control have explored these matters in some detail, but with no finality.<sup>13</sup>

Since autonomy, coercion, and Dworkin are not the principal concerns of this book, there won't (sadly) be space here to explore in detail these important and interesting matters (although coercion is explored a little more in the next few sections). I can, however, briefly make the following points. First (as Dworkin admits), although resenting (or otherwise not reflectively endorsing) the reasons on which I 'have to act' is probably a *necessary* part of having been coerced, it's certainly not *sufficient* for having been coerced – since one may resent 'having to act' on an addictive desire, or on a deeply held moral conviction, without being in any way coerced into acting on them. Second, it seems to me likely (though this is not the place to argue the point) both that the correct account of being coerced will be built around the idea of autonomy and that something like Dworkin's account of autonomy – which sees being autonomous as 'having the capacity to raise the question of whether I will identify with or reject the reasons for which I now act' – is correct.<sup>14</sup>

## 6.2 Descriptive or normative?

At the most general level, there are two views about coercion. One view holds that coercion claims are essentially value-free, that whether one *is* coerced into doing something is an ordinary



empirical question. Another view holds that coercion claims are moralized, that they involve moral judgements at their core.<sup>15</sup>

In this section, and indirectly in the next, I want to consider further coercion's role in ethics, in particular the debate between those who regard it as 'value-free' or 'descriptive' and those who regard it as 'moralised' or 'normative'.

Coercion could be 'moralised' without always being wrong. Nonetheless, it's useful to start by asking whether coercion is ever permissible. Cheyney Ryan argues that it's sometimes permissible as follows:

Janet has a highly contagious disease, which would have disastrous consequences if allowed to spread. The government orders her to stay in the house, but Janet's disease makes her irascible and she refuses to comply. The government seal her house and threaten her with reprisals if she attempts to leave. Janet acquiesces. There is little doubt that . . . the threats of the government have coerced her into staying put, despite her own wishes. But clearly she has no right to leave the house and contaminate the society around her, and the government has the right to keep her quarantined.<sup>16</sup>

This and similar cases seem decisive. The government's actions are both clearly coercive and clearly not wrong. Hence, coercion is not always wrong.

Not everyone agrees, however. Some writers maintain that the concept of coercion is normative in the strongest possible sense, and that actions don't even count as coercive unless they're wrong. Jennifer Greene's paper, 'Coercion: description or evaluation', illustrates this. She asks us to think of threats in terms of loss of *conjunctive options*:

every time an agent faces a demand accompanied by a threat, she loses what I shall call a *conjunctive option*. Although the victim of coercion can still choose between complying and refusing to do so, the conjunctive option of *both* refusing to comply *and* avoiding the threatened penalty has been made impossible.<sup>17</sup>

She then goes on to claim that A coerces B *only if* 'A's intervention *illegitimately* closes off a conjunctive option to which B is entitled' – and (conversely) that those threats which *justifiably* remove conjunctive options aren't cases of coercion.<sup>18</sup> What, then, should Greene say of Ryan's Janet case (above)? She'll have to choose between saying that the government's actions are wrong, or denying that Janet is really coerced.

And, of these, one would think that the latter is preferable, although it does seem at odds with ordinary usage of the term 'coercion'.

Greene, however – and about this much she's right – doesn't take the fact that her theory departs from everyday language to be a decisive objection. On the contrary, her theory of coercion is unashamedly *revisionist*, as the following passages demonstrate:

confusion over the meaning of 'coercion' . . . is in large part responsible for the ubiquitousness of the term. Charging someone or something with 'coercion' seems to be a national pastime. But . . . the charges would not come up so frequently if we did not ultimately consider coercion to be a very serious form of injustice.<sup>19</sup>

those who cry 'coercion' at every turn often seem to want to take advantage of the implicit wrongfulness of the act of coercion and yet make it impossible for the term to designate any particular wrong-doing. Sloppy overuse of the concept has led to its complete impoverishment.<sup>20</sup>

So there may be grounds for taking a revisionist line about the Janet case. This would involve admitting that many users of the word 'coercion' in fact view Janet as *both* coerced *and* fairly treated, but also claiming that these 'coercion' users are mistaken and ought not to be classifying this as a case of coercion (*precisely because* Janet has not been treated unjustly).

As we'll see shortly, I have a certain amount of sympathy with this sort of revisionism. Furthermore, the strategy outlined above for dealing with the Janet case and, more generally, Greene's view are not incoherent. There do, however, seem to be two things which count against her position. The first is that she presents us with a false dichotomy: *either* we must believe that coercion is not a normative concept at all, *or* we must believe that coercion is always wrong. This stark choice is what leads her to be such a radical revisionist. But these are not the only options, and it's possible to maintain that coercion is essentially normative without adopting Greene's style of absolutism about coercion. Her highly counter-intuitive position is therefore unfounded, since it's possible to get what she wants (a normative account of coercion) without departing so extremely from standard intuitions.

The second problem with her position is that it forces us to say odd things about Janet-type cases. Janet has 'a *highly* contagious disease, which would have *disastrous* consequences if allowed to spread'. But what if her disease were only *fairly* contagious, with *quite bad* consequences if allowed to spread? Or a *little bit* contagious, with a threat of *mild inconvenience* if allowed to spread? What I'm getting at here is that

there is a set of cases ranging from 'catastrophic/extremely contagious' through to 'mildly inconvenient/hard to catch'. At some point on the continuum, the risk to public health becomes so great that threatening Janet becomes justified. Greene's view is that, at the very same point on the continuum, threatening Janet ceases to become coercion. Consider, for example, the following case:

An excessively cautious government agency uses threats to make Janice, who poses a low-moderate risk to public health, stay indoors. The risk to public health is (*ex hypothesi*) not great enough morally to justify the agency's actions. However, Janice's infectiousness unpredictably rises to a moderate-high level and, on Tuesday, her condition becomes so contagious that the agency's ongoing detention strategy becomes fully justified. Given, however, that no one could have predicted her post-Tuesday state in advance, the fact that she is now a moderate-high risk does not retrospectively justify the agency's pre-Tuesday actions.

What I am trying to bring out in this case is the peculiarity of saying that the agency coerced Janice until Tuesday, but on Tuesday it stopped coercing her and instead justifiably threatened her. Of course, one *could* say this, and Greene would have to. But this is a major drawback for her position. When it comes to conceptual and linguistic questions like this, occupying a highly counter-intuitive position isn't ruled out in advance, but it's best to steer clear of such positions unless there is a compelling reason to adopt them.

The next question I want to look at is whether, on the assumption that coercion is not always wrong, coerciveness might nonetheless always be wrong-making. This question probably sounds a bit odd to non-philosophers (and perhaps to some philosophers too). What it means is: is coerciveness the sort of property which always counts as a moral reason not to perform those actions which have it? Or if we prefer to talk in terms of value, we could ask instead whether coerciveness is the sort of property in virtue of which actions are always morally 'more bad' and/or 'less good'. Even if coerciveness were universally wrong-making, there would (or could) be some cases of permissible coercion. In such cases, the fact that the action in question was coercive would count as a moral reason not to perform it, but this 'anti' reason would be *outweighed* by some other 'pro' reason, such as the fact that it would save many lives. This is how, in Chapter 2, I suggested that we should think of exploitation. What I said there was that exploiting persons is always wrong, provided that other (sufficiently strong) countervailing moral considerations are absent. But another equally good way of making the

same point would have been to say that the property *being exploitative* is universally wrong-making (in the sense just outlined).<sup>21</sup>

The view that coerciveness is always wrong-making is compatible with Ryan's Janet example. This is because the best analysis of this case is that sealing Janet's house and threatening her is *prima facie* wrong (partly because it's coercive and partly for other reasons). However, sealing the house, etc. is not wrong *all things considered* because the 'anti' considerations, such as coerciveness, are *outweighed* by countervailing 'pro' considerations – in particular, the fact that Janet has a communicable disease, one which would have catastrophic consequences if spread. That this is the best analysis is confirmed by modifying the case slightly and thinking about what we would say (a) if she didn't have a communicable disease, or (b) if a less coercive way of stopping her from spreading the disease were available. In (a) we would say that sealing her house, etc. is wrong because of the *prima facie* wrongness of acting coercively and because of the absence of any countervailing 'pro' considerations (i.e. there is no contagious disease to justify the coercion). In (b), on the other hand, we would regard less coercive ways of ensuring her compliance as morally better than more coercive ways, provided that all the ways were roughly equal in terms of their effectiveness.

A trickier case, also from Ryan, is the following:

You arrive home one evening to find that an intruder has broken into your house and is assaulting your wife. Before you are noticed, you grab your pistol from the desk drawer and threaten to shoot him if he does not immediately leave. He leaves.<sup>22</sup>

In this case, thoughts such as 'the intruder *deserves* to be threatened' and 'the householder was *acting within his rights*' do a lot of intuitive work. If they are to be believed, and I don't see any reason why they shouldn't be, then arguably what's shown is not merely that threatening the intruder wasn't wrong, but that any coerciveness involved in the householder's actions was *not even wrong-making*. But why should we take this extra step? Why not just use the same analysis as in the Janet case and say that the coerciveness is wrong-making, but is simply outweighed by much weightier 'pro' considerations such as the fact that the householder's wife is in danger and the fact that the intruder has already violated the couple's rights?

One way of answering this is to think, as we did in the Janet case, about varying the case slightly and asking: if a less coercive but equally effective (and otherwise relevantly similar) method of making the intruder leave had been available, would this have been morally preferable? If we answer yes, then it looks as if the coerciveness involved is seen as a bad feature of the action, because an action with less of it would have been

better. But if we answer no, then the coerciveness is not viewed as bad. Unfortunately for our purposes, people's intuitions about this case vary substantially. To a great extent, this depends on what they think justifies the use of a threat in the first place. People who think that the justification is self-defence tend also to think that the less coercive option is preferable, because self-defence only justifies *necessary* levels of coercion and force; for these people, the coercion is a necessary evil. Other people think instead in terms of the householder's rights, or in terms of the intruder's deserving to be threatened. These people don't think of coercion here as an evil at all because the householder *has every right to coerce*, or the intruder *deserves* to be on the receiving end of coercion. Finally, some people don't even regard this as a case of coercion. For example, Ryan (who provided the example) uses it as a counter-example to the non-normative view of coercion. He says that the non-normative view would make us describe the householder's threat to the intruder as having '*coerced him into not raping your wife*', which is '*absurd*'.<sup>23</sup>

This takes us back to the question of whether *all* threats are coercive. Greene would say no because, for her, only wrongful threats are coercive. Ryan, whose 'normativist' view is more complex, would also say no; whereas someone who regards coercion as 'value-free' would say yes, because there's no more to coercing than threatening. My own view, which I shall develop in the next section, is that things are not quite as straightforward as these answers suggest and that we need both to look more carefully at the distinction between threats and offers and to return to the idea of welfare baselines in order to answer this question properly.

### 6.3 Threats and offers

When are proposals coercive? The intuitive answer is that threats are coercive whereas offers are not . . . The crux of the distinction between threats and offers is quite simple: A *threatens* B by proposing to make B *worse* off relative to some baseline; A makes an *offer* to B by proposing to make B *better* off relative to some baseline. More precisely, A makes a threat when, if B does *not* accept A's proposal, B will be worse off than in the relevant baseline position. A makes an offer when, if B does *not* accept A's proposal, he will be *no* worse off than in the relevant baseline position. A's proposal may include both a threat and an offer (what Michael Taylor calls a '*thoffer*').<sup>24</sup>

Wertheimer's answer to the coercion question (above) is fundamentally correct, but more needs to be said. Take, for example, a case which is slightly different from the Tiddles case discussed earlier. In this new case, Tiddles is suffering from a life-threatening illness and is taken to the vet,

who says that Tiddles's life can be saved, but only if I'm willing to hand over \$5,000 in vet's fees. Structurally, this seems just like the case in which my home is invaded by a hostile cat killer. For, in both cases, the choice is between being \$5,000 worse off or losing Tiddles. But in the second (vet) case, we probably don't want to say that the vet is threatening me, provided that her fees are not exorbitant and that she is behaving otherwise reasonably.<sup>25</sup> Why do we feel inclined to draw this distinction?

The key to understanding the distinction between offers and threats is contained in our earlier discussion of harm: in particular, the discussion of harm and relevant baselines. Coercing always involves making a threat. And to threaten someone is to tell her (though not necessarily directly or explicitly) that you will harm her if she doesn't (or does) do X. This is to be contrasted with making someone an offer, which means telling her that you will benefit her (or, in cases where people's desires are not self-interested, that you will provide something that they want) if she does (or doesn't) do X. Finally, for completeness, a *thoffer* is a proposal which contains both a threat and an offer (for example, 'sell me your car for \$1,000,000, otherwise I'll kill you').

This explains why we want to (and should) view the cat killer and the vet differently. The would-be cat killer acts coercively because (given a few plausible assumptions) she threatens to harm me, relative to all three possible baselines. She threatens harm relative to a *pre-interaction* baseline because she proposes to take me from a state with Tiddles to a state without Tiddles. She threatens harm relative to a *closest possible world* baseline on the (plausible) assumption that the relevant possible world is not one that contains an even worse incident (for example, a cat killer who proposes also to assassinate my goldfish). And she threatens harm relative to a *normative* baseline on the assumption that she ought not to kill Tiddles.

To turn now to the vet, why should we say that she is offering rather than threatening? The reason is that (again, given a few plausible assumptions) she is not proposing to harm me if I don't hand over the cash, but rather offering to benefit me if I do hand it over. If we use a normative baseline here, what's doing the work is the thought that, unless she is paid, the vet has no obligation to treat Tiddles. If she did have such an obligation and proposed not to act on it, then I could rightly view her proposed behaviour (in this case, an omission) as harmful and her proposal to do nothing as a threat. Then her position would be rather like that of a public servant who, despite being paid an adequate salary by the state to provide its citizens with a free service, insists also on receiving bribes from individual service users.<sup>26</sup> Such a public servant would be threatening to harm service users relative to a normative baseline, since she proposes to withhold from them something to which they have a right, and which she has a duty to provide.

In the vet case, we assume that the vet has no positive obligation to provide free treatment. Hence, it looks as if saying that she won't treat Tiddles unless she's paid is an offer, not a threat (so long as the amount of money requested is reasonable). To put it in a slightly different way, the normative baseline is a possible world in which the vet acts in accordance with all of her *freestanding* (which, here, means something like *pre-contractual*) obligations to Tiddles. That world is (let's assume) a world in which the vet does nothing to help Tiddles. What the vet proposes then is, in return for money, to bring about a future in which Tiddles and I are *better off* than we would be in this ('pre-contractual') normative baseline world. So she is *offering* (to benefit us) rather than *threatening* (to harm us). We can then track our intuitions about coercion by understanding the distinction between offers and threats in terms of *harm relative to a normative baseline*. As just suggested, if the vet has no freestanding obligation to treat Tiddles for free, then her proposal is an offer (and not a threat) because she proposes to *benefit* me, relative to the normative baseline, in return for money. If, on the other hand, the vet *does* have a freestanding obligation to treat Tiddles for nothing, then her proposal is a threat (and hence coercive) because she proposes to *harm* me, relative to the normative baseline, if I don't give her money.

This (like public servants demanding bribes) would be a case of *omissive coercion*, because the vet would be threatening me with *omissive harm* (see section 4.4). Generalising, we can say that A omissively coerces B to do X if and only if

- 1 A has a duty to do Y for B and A's failing to do Y would cause or constitute omissive harm to B; and
- 2 A proposes (threatens) *not* to do Y for B unless B does X; and
- 3 the existence of A's duty to do Y for B does not depend on B's doing X (i.e. it is 'freestanding').

Omissive coercion is presumably quite a widespread phenomenon. Imagine, for example, that an employee already deserves a pay-rise (based on past performance) but that the employer says that she won't get it unless she does extra work. Surely, this is omissive coercion – a threat not to give her that to which she already has a right. And, as we have already seen, corrupt public servants who demand bribes are often guilty of omissively coercing (amongst other things).

However, saying whether (and why) someone's behaviour counts as *omissively coercing* X rather than merely *not offering to benefit* X can be far from straightforward. One of the most interesting discussions of this problem (and, more generally, of the threat-offer distinction) is Robert Nozick's paper 'Coercion'. Consider, for example, what he has to say about his widely discussed *drowning case*:

Q is in the water far from shore, nearing the end of his energy, and P comes close by in his boat. Both know there is no other hope of Q's rescue around, and P knows that Q is the soul of honesty and that if Q makes a promise he will keep it. P says to Q 'I will take you in my boat and bring you to shore if and only if you first promise to pay me \$10,000 within three days of reaching shore with my aid.' *Is P offering to take Q to shore if he makes the promise, or is he threatening to let Q drown if Q doesn't make the promise?*<sup>27</sup>

In many respects, Nozick's view of this and similar cases is like the account sketched so far here:

whether someone makes a threat against Q's doing an action or an offer to Q to do the action depends on how the consequence he says he will bring about changes the consequences of Q's action from what they would have been in the normal or natural or expected course of events. If it makes the consequences of Q's action worse than they would have been in the normal or expected course of events, it is a threat; if it makes the consequences better, it is an offer. The term 'expected' is meant to straddle *predicted* and *morally required*.<sup>28</sup>

If I say 'I expect all of my students to turn in their assignments on time', that can mean two entirely different things. *Either* they are fantastically hardworking, reliable, etc., and so I *believe that they will* turn in their assignments on time. *Or* they are (morally, or according to the regulations, or whatever) *obliged* to turn in their assignments on time. Nozick's view is that there are two usable baselines, corresponding to these two different senses of 'expect': one *predictive*, the other *normative*. If this is applied to the drowning case, there are on his view two plausible answers to the 'offer or threat' question. One involves asking what Q can rationally expect of P in the predictive sense. Hence, we might ask what a (statistically) normal person from P's (or Q's) society would do in these circumstances. The other involves asking what P's moral duties are.

Unless one adopts a very crude form of moral relativism (such as the view that P's duties are necessarily the same as what a normal person from P's society would do), it's obvious that these two baselines can generate different results. For example, it might be true both that P has a freestanding duty to rescue Q *and* that members of P's society are incredibly reluctant to allow strangers onto their boats and would only do so for a handsome reward. If so, then use of the normative baseline generates the conclusion that P threatens Q, while use of the predictive baseline generates the view that P makes Q an offer.

My own view is that the predictive baseline looks rather implausible in the drowning man case and that the normative one is to be preferred. The reason for this is rather like the reason why one might want to reject crude forms of moral relativism. The objection to moral relativism that I have in mind is that whether A's actions are right or wrong ought not to depend (directly) on whether or not A's actions conform to the norms of A's society. For it's possible in principle for *all* members of A's society to have false moral beliefs and/or to behave very badly from an objective moral point of view. Much the same can be said of the 'offer or threat' question (which I take to be a morally significant question, at least sometimes). That is, whether or not P is threatening Q ought not to depend (directly) on what other members of P's society would have done in similar circumstances. To see this point more clearly, think back to the cat killer example. Would her 'offer' be any less of a threat, or any less coercive, if we both lived in a society full of cat killers? Obviously not – for what we should say of such a society is just that it contains *lots of* coercion and *lots of* felicide (cat murder). Yet assessing her 'offer' relative to a predictive expectation baseline gives the conclusion that she really is offering, not threatening – since behaviour like that is only to be expected (in the predictive sense) in a society of cat killers.

Interestingly, this also suggests a 'knock-down' argument against the *universal* use of predictive expectation baselines. Such use would – absurdly – rule out *a priori* the existence of societies which were *both predictably and extremely coercive*. For in order to count as coercive, what's 'threatened' would always have to be worse than the norm. Hence, even if everyone in an alternative society (S) acts 'coercively' (i.e. in ways which *we* would regard as coercive) S won't count as a more coercive society than ours because the behaviour of its members isn't worse than what one would rationally expect.

So given these problems, why not just use normative baselines? Nozick considers and rejects this possibility:

One might think that in deciding whether something is a threat or an offer, the (morally) expected course of events always takes precedence over the normal or usual course of events, where these diverge. It is not obvious that this is so. I have in mind particularly the example . . . where your normal supplier of dope says that he will continue to supply you if and only if you beat up a certain person. Here, let us suppose, the morally expected course of events is that he doesn't supply you with drugs, but the course of events which forms the background for deciding whether he has threatened you or made you an offer is the normal though not morally expected course of events (in which he supplies you with drugs for money); it is against this

background that we can obtain the consequence that he has threatened you.<sup>29</sup>

Wertheimer rightly suggests that Nozick's analysis of this case may not be right because even if the drug relationship is 'immoral from an external perspective, it is arguable that *within* that relationship, A is morally required to continue to supply B with drugs'.<sup>30</sup> Nonetheless, I think that we can find cases that will do what Nozick wants, such as the following:

#### The laundry case

A is a 40-year-old bachelor. He lives with his mother, B. B does all of A's laundry and has done so throughout his life. A makes only a modest financial contribution to the household despite now having a well-paid job. B asks A if he would be willing to increase his contribution to the household budget by 20 per cent. A says no, because he is saving money for a new sports car. After persuasion and reasoned argument fail, B *threatens* to stop doing A's laundry if he doesn't pay the extra money.

This is clearly a threat. But, equally clearly, B doesn't have a duty to carry on doing the laundry. So this must be a case where the predictive baseline takes precedence. We are happy to call this a threat because a strong *predictive expectation* has been established by B's behaviour towards A during the last forty years.

The concepts of threat and coercion, then, are rather like the concept of harm. All three are *multiple baseline* concepts. What I mean by this is that while, in some (most) cases, we have to work out whether behaviour is threatening, or coercive, or harmful, by comparing what is proposed or done to what *ought to be* proposed or done, in other cases such assessments have to be carried out using other comparators, such as *the way things were* ('pre-interaction'), or the way things *would have been* if things had been relevantly different ('relevant possible world'), or the way we could *reasonably expect them to be* ('predictive expectation'). I would suggest that these complexities lead to a good deal of confusion about how exactly to define and apply the terms in question – and, in particular, that it's responsible for spawning the debate discussed in section 6.2 about whether coercion is a 'normative concept' or a 'descriptive concept'. In fact, this is the wrong question, or an overly simple or general question, because whether or not a particular coercion claim has normative significance depends principally upon which baseline is used.

For example, in the laundry case (above) the fact that the mother's proposal is a threat has no normative significance at all. It's not wrong to

threaten her son in this way, and what's more, the fact that her proposal is a threat isn't even *wrong-making*. The main reason for this is that her proposal only counts as a threat if we use a predictive expectation baseline (i.e. she threatens to make him worse off than he could reasonably have predicted given her previous behaviour). If we'd used a normative baseline (i.e. compared what she proposed to what she ought to do), then her proposal would have turned out not to be a threat, on the plausible assumption that she has no open-ended obligation to provide domestic services. There's nothing wrong with using the predictive expectation baseline in cases like this. On the contrary, it fits very well indeed. But, under this interpretation, the assertion that the mother's proposal is a threat is without intrinsic normative significance. By contrast, in the cat killer case (discussed earlier) the fact that the cat killer's proposal ('pay me or I kill Tiddles') is a threat *does* have normative significance – the reason for this being that her proposal is a threat relative to a normative baseline. In short, she ought not to slay Tiddles but nonetheless proposes to (if I don't comply).

So is coercion a normative concept? The answer is that those coercion claims which make use of normative baselines have intrinsic moral significance, whereas those which use other baselines do not. More specifically, in all cases of the first kind, the coerciveness in question is *wrong-making* – although the act of coercion need not be wrong because of the possibility of countervailing moral considerations. That coercion relative to a normative baseline should always be *prima facie* wrong should hardly be surprising given that, when such baselines are used, the threatener is necessarily threatening *either* to do what she ought not to do or not to do what she ought.

Finally, what about the relationship between coercion and threatening? It seems to me – and this is just a 'linguistic intuition' – that we use these two terms a little differently. We are more willing to use the word 'threat' non-morally than we are the word 'coercion', and conversely 'coercion' has stronger negative moral connotations than does 'threatening'. Unfortunately – and about this Greene is quite right – our ordinary linguistic practices are neither clear nor consistent, and there's no doubt that 'coercion' can be used non-morally and that 'threatening' often does have negative moral overtones. However, one can envisage an only slightly revisionary account of coercion according to which coercing is a subset of threatening. More specifically, coercion claims would be those threat claims which utilised moral baselines, with other threats – like the one in the laundry case – being classed as non-coercive threats. In substance, this position is exactly the one I have defended thus far, the only difference being the addition of a linguistic change. On this view, which I support, coerciveness (in the sense just specified) is necessarily *wrong-making*, but threatening(ness) isn't.

#### 6.4 Summary and conclusions

The concept of coercion is ethically significant in a number of ways. 'Coercion' is routinely used to express moral disapproval. Coerced actions tend to be thought of as forced and involuntary in ways which can relieve coerced actors of legal and/or moral responsibility. And the use of coercion to gain consent typically renders any resultant consent invalid. Furthermore, concerns about coercion often underlie people's opposition to controversial practices including the sale of body parts, prostitution, and commercial surrogacy. This chapter has aimed to give an account of coercion which will help us to understand these concerns. The (slightly revisionary) view argued for here is that we should reserve the word 'coercion' for a particular kind of threat: those cases in which (1) A proposes (threatens) to harm B if B doesn't comply with A's wishes, and (2) the proposed harm to B is harm *relative to a relevant normative baseline*. (See Chapter 4, especially section 4.3, for an explanation of harm and relevant baselines.)