**‘Explanation and Condemnation’**

Ward E. Jones

**Forthcoming in *Judgement and Understanding*, edited by Pedro A. Tabensky (Ashgate)**

**Editor’s Introduction**

**Ward Jones discusses explanatory strategies endorsed by Raimond Gaita and Martha Nussbaum that allegedly weaken our need to condemn in order to make explicit what sorts of things that would have to be true for the proverb ‘”*tout comprendre, c’est tout pardonner*” (“to understand all is to forgive all”)’ to contain at least the seed of truth. In Jones’ words, ‘It is a philosophical and legal truism that what kind of explanation we offer of [someone’s] behaviour will affect the kind of judgement that we make of [that person] in virtue of that behaviour, but *which* explanatory strategies have *which* effects, and why?’. Jones’ focus is on condemnatory attitudes rather than on condemnation itself, for his concern is to analyze some of the modes of understanding that would incline us to lessen the strength of the reactive attitude of condemnation. Following Socrates, Gaita argues that evil emerges out of ignorance and he thinks that understanding this should incline us to judge perpetrators less harshly than we otherwise would. Nussbaum, on the other hand, following Seneca, argues that sensitivity to human vulnerability to adverse circumstances will incline us to be more merciful towards perpetrators than we would otherwise be. Jones, without committing himself to either**

**explanatory strategy, gives us reasons for thinking that, if we adopt their explanatory strategies, then the attitudes towards condemnation that they recommend would follow. So, it does seem that at least with regard to some possibly better forms of understanding, understanding is in tension with condemnatory judgment.**

The origin of the French proverb '*tout comprendre, c'est tout pardonner*' ('to understand all is to forgive all') is uncertain. Its earliest known appearance in exactly these words is in a Russian novel, Tolstoy's *War and Peace*.[1] The proverb is sometimes attributed to Madame de Staël, but she had written something slightly different: '*tout comprendre rend très-indulgent*' ('understanding everything makes one very charitable').[2] De Staël may have been inspired by Goethe, who twenty years earlier had written '*was wir verstehen, das können wir nicht tadeln*' ('what we understand we cannot reproach').[3]

My suspicion is that, like all good proverbs, this one—along with its earlier historical variants—contains an important truth, hidden in both simplification and exaggeration. As a first stab—which we will subsequently need to modify—we can say that this proverb and its variants all make the following claim:

> [T] There is a tension between, on the one hand, understanding or explaining someone's wrongdoing, and, on the other, condemning her for it.

I take [T] to be a contingent claim about our moral psychology, and as such must in the end be supported or falsified by empirical work.[4] However, just as in any other area of

our contingent moral psychology, there is room here for non-empirical discussion. In particular, a certain amount of theoretical work is required to find a plausible way of describing the phenomenon at hand. It is a philosophical and legal truism that what kind of explanation we offer of Jones's behaviour will affect the kind of judgement that we make of Jones in virtue of that behaviour, but *which* explanatory strategies have *which* effects, and why?

This paper is intended as a start on the descriptive project of finding a more precise statement of [T], as well as its sources. In the first half of this paper, I will argue that the tension is this: *improving* one's understanding of someone's wrongdoing is likely to result in an abating of our *negative moral sentiments* towards the wrongdoer. Drawing from work by Raimond Gaita and Martha Nussbaum, I will outline two ways in which understanding of moral wrongdoing can be improved. While I will not defend either account of moral wrongdoing, I will argue that we can expect both ways of improving an explanation of a wrongdoing to lessen the explainer's condemnatory attitudes towards a wrongdoer.

## 1. The Tension: Understanding and the Reactive Attitudes

I take it that any tension that exists between understanding and condemnation is a tension between attitudes: one or more of the ways in which we gain an *understanding* of someone's wrongdoing lead us away from having an unequivocal or unambiguous *condemnatory attitude* towards her for what she has done.

On the one side of the tension is understanding. *Explaining* and *understanding* are interdependent notions: to explain $\phi$ to someone (including oneself) is to attempt to

increase her understanding of ϕ. A successful explanation increases someone's understanding of ϕ by informing her of some aspect of why ϕ occurred. I will be concerned solely with explanations of human behaviour, although the phenomenon at hand may be applicable to nonhuman behaviour, or to mental states like beliefs, desires, or intentions.

On the other side of the tension are the various emotional attitudes that we have in response to another person's wrongdoing. A number of writers, beginning in recent years with P.F. Strawson, have emphasized the ethical importance of our emotional responses to another person's behaviour—the so-called 'reactive attitudes' or 'moral sentiments'.[5] A condemnatory moral sentiment is an emotion that arises when a moral expectation that we have of someone is violated. When the violator is myself, I may feel *guilt*, *shame*, or *remorse*; when the violator is someone else, I may feel *indignation* or *disapprobation*.

My condemnatory emotion towards you incorporates a *moral judgement* of you for what you have done. As R. Jay Wallace writes:

> The reactive emotions have the backward-looking focus characteristic of the attitude of blame, and their connection with moral expectations one accepts properly situates blame in the nexus of moral obligation, right, and wrong.[6]

When, for example, I feel indignation towards you given something you have done, the content of that emotion both (i) is targeted at you as *responsible* for the wrongdoing and (ii) involves a *moral criticism* of you.

While we might be able to *imagine* creatures who judge without feeling the

condemnatory attitudes, our disposition to the condemnatory emotions is a central feature of our moral lives. As William Neblett writes, in his discussion of indignation,

> Feeling (other-regarding) feelings of indignation over injustices suffered by others is a symptom of sympathetic concern for others, and in general, of a special sensitivity to matters of morals. Feeling (self-regarding) feelings of indignation over wrongs we ourselves suffer is central to a proper sense of dignity and self-respect.

And later, he adds,

> Certainly, an individual absent of all moral anger and indignation must be indifferent to the existence of moral evil, must be absent of all moral sensitivity.[7]

Condemnatory attitudes like indignation represent one mode of identifying someone as accountable for a moral transgression; without them, our ethical personalities would be very different indeed.

As I conceive of them, the condemnatory attitudes are laid out on a spectrum of strength, in terms of how strongly they lead us to want to *act towards* the wrongdoer before us. At their weakest, they do not lead us to respond in anyway before the wrongdoer; I will simply *feel* indignation, disapprobation, or some form of righteous anger towards the perpetrator. This may or may not include an inclination to see the wrongdoer suffer; for some reason or another, I may not feel that the wrongdoer's

suffering is appropriate or justified in the light of her wrongdoing. Stronger condemnatory attitudes will incline me to verbally reproach, rebuke, or otherwise censure the wrongdoer. At its strongest, a condemnatory attitude may lead me to make *restitutional* demands of the wrongdoer; I may desire an apology or compensation from her, or I may support her imprisonment or physical punishment. Jeffrie Murphy describes such an emotion, which he calls 'retributive hatred':

> The desire to hurt another, to bring him low, is … sometimes, I suggest, … motivated by feelings that are at least partly *retributive* in nature—e.g., feelings that another person's current level of well-being is undeserved or ill-gotten (perhaps at one's own expense) and that a reduction in that well-being will simply represent his getting his just deserts.[8]

The sort of condemnatory attitude that Murphy describes is *essentially* retributive, and has clear behavioural manifestations. My feeling retributive hatred towards you would involve my undertaking or supporting some form of punishment on your behalf.[9]

Like all mental states, the condemnatory attitudes are at least partly constituted by their manifestations in our behaviour or other mental states, and so I will not distinguish in this paper between the weakening of the manifestations (behavioural or otherwise) of the attitude and the weakening of the attitude itself.

## 2. *Verstehen* Explanations: Reasons and Identification

Our question is whether there are any explanations of a person's wrongdoing such that

when I accept them my condemnatory attitudes are likely to weaken. A basic and undefended assumption of this paper is that there are two ways in which we can explain someone's behaviour, which I will designate using the German words '*erklären*' and '*verstehen*'.[10]

An *erklären* explanation of someone's behaviour involves showing—in accordance with science—that the behaviour instantiates some sort of generalization. Not all *erklären* explanations involve causes; there are *erklären* explanations that cite non-causal correlations. Imagine, for example, someone who adds orange juice to milk, and then explains the result by claiming that the juice of all orange-coloured fruit curdles milk, knowing full well that it is not the *orangeness* of the fruit that causes the curdling of the milk.[11] What makes this person's explanation *erklären* is that it depends upon a generalization; he explains what happens by showing how it instantiates the way of the world.

A *verstehen* explanation of Jones's behaviour, on the other hand, works by making Jones's behaviour appropriate from his point of view; it appeals to what appear to be Jones's own reasons for doing what she did. Just because an explanation cites reasons, however, does not in and of itself make it *verstehen*. The differentiating component of *verstehen* explanations is their lack of dependence upon generalizations. While *erklären* explanations need generalizations, *verstehen* explanations get nothing from them; any generalization that *could* meaningfully be added to a *verstehen* explanation—like 'people tend to act in accordance with their desires'—will be a truism, and thus will add no explanatory power to the explanation.

For some explanatory statements, it will not be clear where they fall in the dichotomy between *verstehen* and *erklären*. On the one hand, when someone says,

'Jones ϕ-ed because he is selfish', is he saying that 'Jones *tends* to perform selfish acts, and that explains why he performed the selfish act ϕ' (i.e., an *erklären* explanation), or is he saying something like 'Jones was seeking his own benefit when he ϕ-ed' (i.e., a *verstehen* explanation)? We would need to know more about the speaker, the hearer, and the context of their interaction, in order to know which interpretation is correct. On the other hand, and as we will see below, an explanation may have both *erklären* and *verstehen* features. However, even in such 'mixed' explanations, *erklären* and *verstehen* elements contribute their explanatory power in very different ways. Introducing an *erklären* element into an explanation of a piece of behaviour works by showing that some aspect of the behaviour instantiates a pattern, while introducing a *verstehen* element works by showing that the behaviour was undertaken because it was, from the agent's point of view, reasonable or appropriate, and does not depend upon any pattern in the world.

I will look at each of these explanatory strategies in turn. In this and the next section, I will focus on *verstehen* explanations, and in Section 4, I will return to explanations with *erklären* elements.

*Verstehen Explanations and Rationalization*

Those who write on explanations in the *verstehen* category tend to emphasize one of two characteristics that such explanations apparently possess. Some writers present *verstehen* explanations as 'rationalizing'; offering a *verstehen* explanation of someone's behaviour involves showing that the actor behaved as he did because he had a reason to do so. Other writers present *verstehen* explanations as involving a process of

'identification' or 'replication'; offering a *verstehen* explanation of someone's behaviour involves seeing the behaviour from the actor's own point of view.[12] I am not here concerned with defending either rationalization or identification as a necessary aspect of *verstehen* explanations. Rather, I will be concerned with showing that *neither* apparent aspect of *verstehen* explanations would, by itself, lead us to expect there to be a tension between offering a *verstehen* explanation of someone's wrongdoing and at the same time condemning her for it.

Many accounts of *verstehen qua rationalizing* explanations claim that such explanations present the agent as behaving because there is some sense in which it is *suitable* or *fitting* for him to do so. This view of mental state explanations has been prominent in influential work from Donald Davidson, Daniel Dennett, and John McDowell. McDowell, for example, claims that propositional attitudes 'have their proper home' in rationalizing explanations, that is, in 'explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be.'[13] Explicitly contrasting *verstehen* explanations with *erklären* explanations, Phillip Pettit writes that the former include

> a norm at which [an agent] aimed, rather than just a datum about how [she behaves] … the explanandum [is] made intelligible, not by being shown to exemplify the world's regular mode of operation, but by being depicted as something that had to happen if the [agent] was to continue to satisfy the principle that represents its norm.[14]

Each of these writers emphasizes that *verstehen* explanations treat an agent as behaving

in such a way because he had reason to do so.

Conceiving of *verstehen* explanations as rationalizing has the consequence that they have a positive partiality inherent in them. In a rationalizing explanation, an agent is said to have behaved in such a way because he *should*—in a minimal sense—have behaved as he did. The qualification 'in a minimal sense' must be added, because an explainer who gives a rationalizing explanation of Jones's behaviour need not *endorse* Jones's behaviour; she need not agree that Jones ought to have behaved as he did. Nonetheless, the explainer must make some sort of positive evaluation of the behaviour. Rationalizing explanations work precisely by showing that the agent's behaviour was, from her point of view, proper.

Bernhard Schlink emphasizes this aspect of *verstehen* explanations in order to diagnose the tension between explaining and condemning. *Verstehen*, Schlink declares, has enough 'positive normative connotation … to make condemning difficult.'[15] The thought here is that the positive partiality inherent in rationalizing explanations, while leaving the explainer plenty of room for *praising* Jones for his behaviour, does not seem to leave as much room for *condemning* Jones; this is because, in the process of offering a rationalizing explanation of Jones's behaviour, we have already established that he saw himself as having a reason to do what he did. Schlink writes that 'once we understand that another person's behaviour *makes sense* in light of our own normative considerations, it becomes, of course, hard to judge his or her behaviour.'[16] As the lens through which we see Jones's action, rationalizing explanations seem to give a head start to a positive judgement of Jones, while a basis from which to reproach Jones looks harder to come by. This, Schlink suggests, is the source of tension [T].

While there seems to be a positive assessment built into *verstehen* explanations,

I do not think that this can be the source of our tension. To see this, imagine an agent, Jones, who performs an action, ϕ, which we judge to be a wrongdoing. Perhaps the simplest explanation of Jones's ϕ-ing would portray him as, simply and straightforwardly, *intending* to commit a wrong: Jones's ϕ-ing was a malicious committing of harm. Note that this explanation of Jones's action is a rationalizing explanation, as it presents Jones as intending or desiring harm, and then acting in such a way as to bring that harm about. It recognizes that Jones acted in such a way as to fulfil his (malicious) desires. I will refer to explanations which present an actor as intentionally malicious as '*verstehen*$_{\mathrm{M}}$ explanations'. *Verstehen*$_{\mathrm{M}}$ explanations claim that in full awareness of what he was doing, an agent sought and acted so as to bring about harm.

*Verstehen*$_{\mathrm{M}}$ explanations are a counterexample to Schlink's suggestion that rationalizing explanations *simpliciter* are in tension with condemnation, for there seems to be no tension whatsoever between offering a basic *verstehen*$_{\mathrm{M}}$ explanation of someone's action and feeling condemnatory emotions towards her. There is nothing in a *verstehen*$_{\mathrm{M}}$ explanation that either detracts from our thinking that the agent is responsible for her action or which would lead us to be compassionate with her. On the contrary, a straightforward *verstehen*$_{\mathrm{M}}$ explanation states that the agent is both responsible and malicious; indeed, that is *all* that a *verstehen*$_{\mathrm{M}}$ explanation tells us, so there is nothing within such an explanation to lessen our condemnatory feelings. So, *pace* Schlink, there is nothing in a rationalizing explanation *per se* that is in tension with condemnation.

*Verstehen Explanations and Identification*

Perhaps the other alleged aspect of *verstehen* explanations—as involving 'identification' with the agent—will lead us to predict a tension between understanding and condemnation. The idea here is that a *verstehen* explanation involves adopting, to some extent, the agent's own attitudes and thought processes. Jane Heal describes the process—which she calls 'replication'—thus: 'I place myself in what I take to be [the agent's] initial state by imagining the world as it would appear from his point of view and I then deliberate, reason and reflect to see what decision emerges.'[17] And Robert Gordon writes that, in essence, the explainer predicts or explains what the agent does by herself 'deciding what to do'.[18] According to this conception of *verstehen*, the *verstehen* explainer explains by having *herself* run through the agent's decision process.

Is the process of identifying with a wrongdoer in tension with condemning her? Again, *verstehen*$_{\mathrm{M}}$ explanations seem to provide us with a counterexample to this suggestion. I can adopt your point of view in my *verstehen*$_{\mathrm{M}}$ explanation of your wrongdoing. That is, I can imagine having your malicious desires, and imagine being confronted with the situation that was in front of you before you committed your wrongdoing. Then, I can deliberate with those desires and perceptions, and decide to commit the same action that you did. Doing so would be a replication of your decision process. However, my running through this replication process does not seem to detract from my inclination to condemn you for your malicious act. I understand why, from your own point of view, you have acted as you did, but it is far from obvious that my doing so will weaken my condemnatory attitudes towards you.

Brian Penrose, a defender of the identification view, writes that 'when we

consider what we *would* have done, most of us, I think, find it hard to say with any confidence that we would have [acted] any differently' from a wrongdoer; Penrose adds that an explainer's recognition that he would have done the same as the wrongdoer Jones will weaken his condemnation of Jones.[19] I do not find the second claim obvious; is it really true that my thinking 'there but for the grace of God go I' will weaken my condemnation of a wrongdoer?[20] Even if I agree with Penrose that it would, however, the existence of *verstehen*$_{M}$ explanations throws significant doubt on Penrose's first thought, namely that understanding a wrongdoing leads me to doubt whether I would have acted differently in the wrongdoer's situation. In a *verstehen*$_{M}$ explanation, I attribute malicious desires to a wrongdoing agent, and in doing so I discover reasons for his doing what he did. However, it is obvious that I can attribute malicious desires to another agent *without attributing them to myself*. And if this is true, then it follows that I can offer a *verstehen*$_{M}$ explanation of someone's wrongdoing without thinking that I would have done the same thing in the same situation. If one's explanation of a wrongdoing is based on desires the wrongdoer has but one does not, then one can explain a wrongdoing while at the same time thinking that in the same situation one would have acted differently. I can learn a great deal about a wrongdoer, in short, while not thinking that I am relevantly like him to have performed the same action.[21]

*Improved Understanding and the Tension*

As we have seen, *verstehen*$_{M}$ explanations provide us with a counterexample to the thought that either rationalizing or identifying with a wrongdoer is in tension with feeling condemnatory attitudes towards her. More importantly, however, *verstehen*$_{M}$

explanations, which are perfectly *familiar* and *usable* explanations, show us that if there is a tension between understanding and condemnation, it is not *just any* instance of understanding which will create that tension. Rather, the tension must lie between having *improved* understanding—e.g., compared to that achieved with a *verstehen*$_M$ explanation—and condemnatory attitudes; the *better* we understand, the more likely our condemnatory attitudes will weaken. That is, we must change the phenomenon we are looking for into one that is essentially normative:

> [$T_N$] While there may be no tension between accepting a *verstehen*$_M$ explanation of someone's behaviour and condemning her for it, having a *better* understanding of her immoral behaviour will lessen one's condemnatory emotions towards her.[22]

Thus, in order to capture any tension between understanding and condemnation, we need to find ways in which we can improve upon a *verstehen*$_M$ explanation, and we need to then determine whether it is plausible that the condemnatory attitude(s) that we feel towards the agent in virtue of what she has done will be weakened by any of these explanatory improvements. In the next two sections, I will examine two views of the nature of wrongdoing; each gives us a standard according to which one explanation will be better than another, and each, I will suggest, gives us reason to think that my accepting a better explanation of a wrongdoing may indeed weaken my condemnatory attitudes towards a wrongdoer.

### 3. *Verstehen*$_{\text{I}}$ Explanations: Wrongdoing and Ignorance

Some *verstehen* explanations present Jones's wrongdoing as not being a matter of malicious intent, but rather as the result of Jones's ignorance. Jones threw away Rachael's latest poem, but he did not know that the poem was scrawled on the crumpled piece of paper on the floor. In such cases, Jones's ignorance is seen to *excuse* him—although, as we will see, perhaps not completely—from being guilty of a malicious act. Wallace writes:

> Suppose I do something that happens to be of a kind *x* [e.g., bring about a harm]. The first class of excuses [i.e., inadvertence, mistake, and accident], defeats a presumption that I did *x* intentionally, by showing that I did not know that I would be doing something of kind *x* at all when I chose to do whatever it was that turned out to be of kind *x*.[23]

The inclusion of this kind of excusing condition portrays the agent's harming not as a consequence of his malicious intent, but the result of an epistemic failure: Jones did not know that his ϕ-ing might or would result in harm.

Ignorance *verstehen* explanations, or *verstehen*$_{\text{I}}$ explanations, make reference not to Jones's reasons for harming (for he may have had none), but to Jones's reasons for doing something else altogether; perhaps Jones wholly intended to throw the piece of paper away as a part of his larger activity of cleaning the house. Thus, *verstehen*$_{\text{I}}$ explanations *are* a kind of *verstehen* explanation. However, a *verstehen*$_{\text{I}}$ explanation recognizes that Jones did not know (with certainty in some cases, or at all in others) that

the harm would result. Jones was not malicious in the harm he brought about; Jones's throwing away Rachael's poem was not intentional.

Introducing the presence of ignorance *may* mean that Jones does not deserve any condemnation; perhaps Jones took every sensible precaution against such harm from occurring, or perhaps he could not have expected such harm to have resulted from his action. But in other cases the presence of ignorance will not let Jones completely off the hook, as Wallace reminds us:

> [Ignorance] may not be accepted [as an excuse] if the ignorance that makes what one did unintentional is itself culpable. In that case, it will be taken not as a valid excuse, but evidence for one of a different family of faults that includes negligence, carelessness, [and] forgetfulness … [24]

Jones did not know that the piece of paper he threw away contained Rachael's latest poem, but he *should* have known that a piece of paper on the floor in her room could be important, or he *should* have checked that there was nothing on the piece of paper before he threw it away. So, while it is true that introducing Jones's ignorance into our explanation of his action leads us to recognize that harming was not something Jones desired or did intentionally, we may reproach him nonetheless: perhaps he should have known that harm could result from his action (i.e., he was culpably ignorant) or perhaps he did know that such harm was *possible*, but he did not take this possibility into account (i.e., he was negligent).

Importantly, even if it does not get Jones completely off the hook, introducing the presence of ignorance changes what otherwise might look to be a malicious action

into one that is not malicious. If Rachael thinks that Jones is guilty of negligence or culpable ignorance, then she may feel he deserves her resentment or reproach. However, her resentment will be *less* than it would be were she to believe a *verstehen*$_{M}$ explanation of Jones's throwing her poem away. In the latter case, Rachael will see Jones's action as expressive of an ill will towards her; she will see him as having sought to cause her harm. However, changing her mind to a *verstehen*$_{I}$ explanation would remove Rachael's thought that Jones had any ill will towards her. So, attributing an element of ignorance to Jones—that is, by changing her *verstehen*$_{M}$ explanation into a *verstehen*$_{I}$ explanation—Rachael's condemnatory attitude towards Jones for what he has done is likely to weaken.

*Verstehen* explanations of actions of wrongdoings can be divided into *verstehen*$_{M}$ explanations, in which the agent is said to have intentionally committed a malicious act, and *verstehen*$_{I}$ explanations, in which the agent is said to have been negligent or ignorant.[25] Furthermore, I have suggested that the latter are going to arouse less strong condemnation—and perhaps no condemnation at all—than the former. If this is right, then we now have one place in which switching from one kind of understanding to another lessens condemnatory attitudes: while *verstehen*$_{M}$ explanations are not in tension with condemnation, introducing ignorance into our portrayal of an agent's wrongdoing, it seems, is likely to decrease condemnation.

*Gaita: The Ignorance at the Bottom of Maliciousness*

Drawing a simple division between two categories of *verstehen*$_{M}$ and *verstehen*$_{I}$ explanations, and showing that the latter will arouse less condemnation than the former,

however, does not establish that [$T_N$] is correct. [$T_N$] says that a *better* explanation leads to weaker condemnatory attitudes. We have established that *verstehen*$_I$ explanations lead to weaker condemnatory attitudes, but we have not yet established that *verstehen*$_I$ explanations are to be preferred over *verstehen*$_M$ explanations.

Raimond Gaita pursues a train of thought, associated with Socrates, that would make *verstehen*$_M$ explanations less desirable than *verstehen*$_I$ explanations.[26] Gaita argues that an act that is apparently malicious will in reality be an act that is done in ignorance of what one is doing. The ignorance Gaita has in mind here is an evaluative ignorance, a lack of understanding of the value of human beings; one could not, he thinks, wholly comprehend 'the preciousness of each individual human being' and at the same time intend to bring about harm to one. Discussing the example of the sadist, someone who 'clear-sightedly' treats evil as an object of 'fascination and desire', Gaita writes:

> Sadists appear to have a refined sense of human dignity and they take self-conscious pleasure in its violation. But that is quite evidently consistent with the Socratic thought that the sense of human dignity that gives pleasure to their cruelty is a false semblance of a genuine understanding of it. If Socrates is right, then the sadist fails fully to understand what he does, just as the ordinary brute does.[27]

If this is right, then a *verstehen*$_M$ explanation of an act, one which presents the agent as intentionally acting so as to bring about harm, should in reality be a *verstehen*$_I$ explanation, one that acknowledges the agent's ignorance. Following Gaita's Socratic

line of thought would mean that we tend to opt for *verstehen*$_{\mathrm{I}}$ explanations as being 'deeper' than *verstehen*$_{\mathrm{M}}$ explanations. The malicious wrongdoer is such because he has not gained or appreciated a fundamental kind of moral or evaluative understanding.

Gaita emphasizes that

> this failure of understanding is not of a kind that would interest a court. It would not enable one to enter a plea of diminished responsibility for a crime.[28]

He is right, of course. The sort of ignorance he is describing is not a straightforward excusing condition. The sort of *verstehen*$_{\mathrm{I}}$ explanation he has in mind will not mean that we wholly excuse an agent from responsibility for wrongdoing. Nonetheless, it remains true that seeing the wrongdoer in this light tempers our condemnation of him. This is not merely because—as Gaita is at pains to point out—such understanding serves as a reminder that *we* must treat the wrongdoer himself as precious, but also because it moves the agent's wrongdoing into the realm of a 'lack' in his epistemic character. Accepting a *verstehen*$_{\mathrm{I}}$ explanation in place of a *verstehen*$_{\mathrm{M}}$ explanation involves a change of focus, from malicious intent to epistemic lack, and it is not implausible to think that this change of focus would tend to diminish our urge to condemn him. Such an explanation directs our attention to the type of person the wrongdoer has *not* become, rather than upon the action he has performed. We come to see the agent's wrongdoing as a matter of epistemic shortcoming. As this is the change of view lying behind the whole range of ignorance excuses, it is plausible to think that adopting *verstehen*$_{\mathrm{I}}$ explanatory strategy in lieu of *verstehen*$_{\mathrm{M}}$ strategy will to some extent weaken our condemnatory attitudes towards wrongdoers.

If Gaita is right, then it is not just that there are two kinds of *verstehen* explanatory strategies that we can use, but that one is better than another. A better understanding of wrongdoing is to be found through a *verstehen*$_{I}$ strategy than a *verstehen*$_{M}$ strategy. And as we have seen, there is reason to think that this better strategy tends to weaken our condemnatory reactions to the wrongdoer, that using it will lead us towards weaker condemnatory attitudes than we may otherwise have felt.

While I will not here undertake to defend Gaita's ignorance view of wrongdoing, it seems to me that this is precisely the sort of consideration that we need to support [$T_N$], for it gets us to the claim that better understanding is in tension with condemnation. If we were to *correctly* understand how intentional harms happen, says Gaita, we would see them as at bottom a matter of ignorance, and our condemnatory feelings would weaken as a matter of course.

## 4. *Erklären* Explanations

In an *erklären* explanation, a piece of behaviour is explained as being an instantiation of a generalization. Jones, it is said, behaved in a B-like way, because Jones was characterized by C (e.g., has a certain genetic configuration, or a mental illness, or was abused as a child), and people who are characterized by C tend behave in B-like ways. The property of *being-characterized-by-C* tends to be accompanied by the property of *behaving-in-a-B-like-way*. Early in his lecture, Schlink suggests that we could offer an *erklären* explanation of someone's behaviour and still condemn it: 'After having dealt with empirical facts and nothing but empirical facts, we are free to judge and, if appropriate, to condemn'.[29] An *erklären* explanation, Schlink suggests, is not in tension

with condemnation.

At the limit of *erklären* explanations, at least, this is not right. To see someone's behaviour *solely* as an instantiation of a generalization is not to see it *as an action*, as something that an agent does intentionally, in pursuit of a goal or in adherence to a norm. But to see a piece of behaviour in this way is not yet to see it in light of which it can be judged at all. It is only when we see the behaviour as being undertaken in fulfilment of some norm or goal, and not just *as a part of the way of the world*, that we have a position from which to speak of the behaviour as appropriate or unsuitable, as good or bad. When we take a solely *erklären* view of a piece of behaviour, it tends to lose its status as something in the light of which anyone deserves judgement, condemnation, or praise. An *erklären*-explicable piece of behaviour is not, by that fact, in the space of reasons, and, accordingly, if I accept a wholly-*erklären* explanation of Jones's behaviour, then I am not likely to feel any condemnatory attitudes towards him. We need to see someone through *verstehen* lens before we feel the moral sentiments towards her.[30]

More commonly, however, an explanation will not be *wholly erklären*, but will be a *verstehen* explanation with *erklären* elements. That is, the explanation will be a *verstehen* explanation in which the agent's action is seen as being causally influenced by non-rational factors. For example, imagine explaining why a teenager hits his little brother. We may offer a simple *verstehen*$_{\mathrm{M}}$ explanation of the teenager's action: he simply wanted to hurt his little brother to see him cry. But perhaps we know more about him. Perhaps we know that he is often hit in a similar way by his father, and we think that such treatment is likely to have causally influenced his own treatment of others. Or, perhaps we know that he has a genetic make-up which makes him susceptible to

impulsive behaviour. The result of adding either bit of information would be a mixed *verstehen/erklären* explanation, one which includes both kinds of explanatory elements. It is in these sorts of examples—in which intentional actions are influenced by other factors—that we see that there must be no clear boundary between *verstehen* and *erklären* explanations of actions. The former often include elements of the latter.

So, while it is clear that accepting a wholly *erklären* explanation will not bring about condemnation, we need to look at 'mixed' explanations, *verstehen* explanations with *erklären* elements, in order to see (1) whether adding *erklären* elements can be seen as an improvement on other explanations, and (2) whether adding *erklären* elements to an explanation will tend to reduce the condemnatory attitudes of those who accept that explanation.

*Nussbaum: the Necessity of Non-Rational Determination*

One of the core points of Nussbaum's 'Equity and Mercy', as I read it, is that the more one examines an individual who commits a wrongdoing, the more one comes to appreciate the role that non-rational determinants play in determining what he has done. In her sympathetic interpretation of Seneca, Nussbaum writes, 'People who do bad things … are yielding to pressures—many of them social—that lie deep in the fabric of human life', and thus,

> the wise person is not surprised at the omnipresence of aggression and injustice, 'since he has examined thoroughly the circumstances of human life' [Seneca, *On Anger*, II.10]. Circumstances … are at the origins of vice. And when the wise

> person looks at these circumstances clearly, he finds that they make it extremely difficult not to err. … Thus aggression and violence grow not so much inside us as from an interaction between our nature and external conditions … [31]

Nussbaum is here endorsing a phenomenon that most of us will find familiar: looking to a person's upbringing, circumstances, and context can lead us to discover aspects of his past and present context that play a role in determining her behaviour on any one occasion. Importantly, this determination is not a rational determination. Nussbaum's phrases 'yielding to pressures' and 'an interaction between our nature and external conditions' do not describe an agent following reasons; they describe non-rational correlations between our behaviour and the world. To appeal to this sort of determinant in the explanation of someone's behaviour is to introduce *erklären* elements into that explanation.

Nussbaum is not specific about the nature of non-rational determinants of behaviour, nor does she tell us the mechanisms by which they work. I suspect that these omissions are intentional. What determinants I introduce into my explanations will depend upon the 'theory' of behavioural forces with which I work.[32] Some of the generalizations in such a theory will describe *social* influences, like 'in the midst of peer pressure or a crowd, an agent may do things that she would otherwise have avoided as wrong'. Others will be *psychological* or *biological*, as in 'if an agent is under great stress then she is more likely to act in an impulsive, unthinking, or callous way towards other persons' or 'if S was abused as a child, then she is more likely to behave in a callous or less caring way towards other persons.'[33] Indeed, we have every reason to think that our knowledge of the non-rational processes affecting behaviour will continue

to change and grow; as, for example, we come to have more *genetic* knowledge of ourselves, we may begin to incorporate such knowledge into our explanations of each other's behaviour.

Whether or not any of the particular generalizations concerning the non-rational determination of behaviour are correct is irrelevant to the points that Nussbaum wants us to take from her discussion. First, she wants us to see that it is inevitable that as one delves deeply into an agent's life, one will come across features of her action for which the agent has no reasons. At some level of description, an agent will be influenced by emotions and tendencies to behaviour which are not, from her point of view, rational or suitable. While a *verstehen* explanation may, in a particular context, be sufficient, an expanded *verstehen* explanation will eventually become a mixed *verstehen/erklären* explanation. At some point as I fills out the portrayal of a wrongdoer, I will no longer be able to offer a purely *verstehen* explanation of her behaviour. As I add more and more details in my explanation of the agent's behaviour, I will inevitably make recourse to non-rational determination, and at this point my explanation will become partly *erklären*.

Nussbaum also wants to convince us that, all else being equal, a mixed explanation is going to be a *better* explanation than a purely *verstehen* explanation. As she writes, 'good moral assessment, like good medical assessment, is searchingly particular, devoted to a deep … understanding of each concrete case.'[34] The mixed *erklären/verstehen* explanation is one that 'the wise person' accepts. Lying behind this claim is a view of wrongdoing: 'People who do bad things … are yielding to pressures,' she writes, 'circumstances … are at the origins of vice.'[35] Just as Gaita thinks that wrongdoing is, at bottom, the result of a certain kind of ignorance, Nussbaum thinks

that wrongdoing is *at least partly* the result of non-rational pressures; a more particular explanation of wrongdoing, including *erklären* elements, is going to more accurately represent how the agent came to commit the harm, than one that is wholly and simplistically *verstehen*.

As in my discussion of Gaita in the previous section, I am not here concerned to defend Nussbaum's view of wrongdoing. Rather, I want to return to the question of whether the change of agential explication that Nussbaum advocates will lead to the weakening of my condemnatory attitudes towards her. There are two reasons why one might think that such elements will do so; only one, I will argue, is a good reason.

*Non-rational Determination and Responsibility*

Ignorance, we saw in Section 3, is one condition under which we excuse an agent from responsibility. We also excuse an agent in the face of behaviour that she did not initiate and could not prevent. Recognizing this second kind of excuse may lead to the following line of thought. Non-rational determination involves a loss of *control* over one's behaviour; the *passivity* of an agent's behaviour, we might say, is incompatible with her being held responsible for that behaviour. In so far as an explainer sees an agent's behaviour as influenced by non-rational elements, so far will she see the agent's behaviour as having been out of his control. But to see a piece of behaviour as, to some extent, out of one's control is to see it, to that same extent, as not being an action which belongs to the agent. The more I see non-rational elements influencing an agent's behaviour, the less I see her as responsible for that behaviour. Accordingly, and as a result, as I come to see that particular and non-rational causes have influenced an

agent's behaviour, my moral sentiments towards her will be weakened.

This line of thought is related to that behind the traditional problem of free will. The *thesis of causal determinism*—which says that any event, including any action of mine, will have been inevitable given the state of the world at some previous time—inspires an image in which we are mere *liaisons* of change, in that while an event may not have arisen without us, we will not hold an originating place in its being carried out. But if such motions are already determined to occur—as the thesis of causal determination tells us—it is perplexing how I can be seen as something that is a source of my bodily movements, for the intuitions associated with the practice of excusing an agent in the face of passivity suggest that seeing someone as an agent of an action involves taking her to be non-passively involved in change. Thus the traditional problem of agency: how, in the face of determinism, can I be picked out as locus of responsibility for any change?

P.F. Strawson famously pointed out that a commitment to the thesis of causal determinism will not block the moral sentiments; even if I believe determinism, I will still respond with emotions like resentment and indignation in the face of wrongdoing.[36] Strawson is surely right in this. However, his point about the inefficacy of the *generalized* thesis of causal determinism upon our moral emotions towards individuals does not change the possibility that when I come to think that particular and non-rational causes have influenced a particular agent's behaviour, this image of the agent's passivity emerges with a force, one might think, that can weaken our moral sentiments. Seeing an individual wrongdoer as having been socially, psychologically, or genetically influenced to behave in the way in which she did, will affect our moral responses to her in a way that the general thesis of determinism cannot. In short, the suggestion that I am

now considering is that as we add *erklären* elements to our explanation of a wrongdoing, the agent begins to have less and less salience in our portrayal, and with that disappearance goes our condemnatory attitudes.

This is certainly a sensible suggestion, but it is not right. It is important that the *erklären* feature under consideration here *influences but does not necessitate* the agent's bringing about harm. If it did necessitate her behaviour, then there would be a wholly *erklären* explanation of the behaviour. In this case the wrongdoer would be off the hook completely; she would thereby be shown not to have been in control of her behaviour *at all*. However, we are here concerned with *adding erklären elements* to an account of someone's wrongdoing, that is, we are concerned with non-rational *influences* upon behaviour. Such influences may not, and in many cases will not, be seen as entailing a loss of control. In the situation in which Jones was susceptible to a non-rational influence leading him to wrongdoing, the question is bound to arise as to whether Jones could have prevented himself from acting in the way in which he did. And if we cannot see why this person did not 'control herself' in the face of such an influence, then we may not be inclined to condemn her less.

The American lawyer Alan M. Dershowitz advocates a non-excusing response to many non-rational influences on behaviour in his book *The Abuse Excuse, and Other Cop-Outs, Sob Stories, and Evasions of Responsibility*. For example, Dershowitz disagrees with the 1991 acquittal of a woman (from the charge of assaulting a police officer) on the basis that she was suffering from premenstrual syndrome (PMS), writing:

> Though some women who are irritable and hostile during the premenstrual period of their cycle may well suffer from PMS, the vast majority of women

> who suffer from PMS do not behave the outrageous way the [woman] in this case did. … She ought to take responsibility for her own actions.[37]

The last sentence in this passage indicates that while Dershowitz appreciates that PMS can influence a woman's behaviour, he assumes that its influence is controllable, that is, that PMS does not entail its sufferer's complete loss of control. Because of this, he argues, we must condemn behaviour under the influence of PMS no less than that which is under no such influence.

In 'Equity and Mercy', Martha Nussbaum is concerned to have us recognize that erring agents face what she calls *obstacles* to good action.[38] Perhaps the wrongdoer was under great stress, perhaps she was jealously in love, perhaps she was not shown compassion as a child, or perhaps she has a genetic tendency to impulsive behaviour. As the passage from Dershowitz shows, it is not obvious that my recognizing your obstacles will lessen my condemnatory feelings towards you. Importantly, this is true even *if I have faced the same obstacles*. As Michael Stocker writes,

> consider jurors who learn of the difficulties and temptations the defendant faced and remembered that they also faced similar difficulties and temptations. This may increase leniency and mitigation; perhaps these people are moved by thoughts such as: 'There but for the grace of God go I.' But some people become harsher the more they see others as similar to them. Here we might think of those who remember—fiercely, perhaps with pride, perhaps with indignation—that they struggled and overcame their own difficulties without any excuses. They now *demand* the same high performance from others …[39]

Jones's personal experience with the obstacles that a wrongdoer has faced will not necessarily weaken his condemnation of her, because the process of identifying with her involves Jones's bringing his own past to bear in his judgement of her. Such baggage may mean that Jones, as explainer, feels more condemnation towards the wrongdoer, rather than less.

I introduce the above passages from Dershowitz and Stocker not because I endorse either Dershowitz's view or the view of Stocker's demanding jurors. Rather, I see both as illustrating that non-rational influences upon behaviour may not, and in many cases will not, detract from our moral condemnation of the agent. While it may be true that non-rational *necessitation* of Jones's behaviour *completely* excuses him from condemnation, it does not follow that a non-rational *influence* on Jones's behaviour *partially* excuses him from condemnation.[40]

*Pity, Compassion, and Ambivalence*

My condemnatory attitudes towards a wrongdoer will not weaken with my recognition of *just any* non-rational influences. However, it seems likely that they will weaken if the recognition of those influences lead me to have emotions towards the wrongdoer that *conflict* with condemnatory attitudes. If Jones's *erklären* explanation of a wrongdoing leads him to feel pity or compassion towards the wrongdoer, then it is to be expected that his condemnation of the wrongdoer will not be as unambiguous as it would otherwise have been.

We expect conflicting *beliefs* not to exist side-by-side. A rational believer will

give up one of a pair of contrasting beliefs, or suspend judgement on both. In contrast, conflicting emotions can exist side-by-side, without any loss of rationality on the part of the emotive agent. In 'Ambivalence and the Logic of Emotion', Patricia Greenspan imagines the following familiar situation:

> Suppose that a friend and I are in competition for some honorific position … What emotions might I feel, not toward my rival himself, but toward the fact that he turns out to win? I think we might plausibly hold, in some conceivable cases, that I have mixed feelings. I feel both pleased (at least to some extent) and pained—happy 'for' him (as we say)—since I know that he deserves the honour and has been hoping for it, but unhappy on my own account, since my own desire has been frustrated.[41]

Situations like this will be recognizable to all of us, not only in our public careers but in our relationships; think of your feelings towards an ex-partner who excitedly announces to you that he or she is getting married. In such cases, our emotions may be, as Greenspan puts it, 'ambivalent' or 'mixed'; we may find ourselves 'in two minds'. Emotional ambivalence is reflected in our behaviour. We may pause, for example, before congratulating our colleague or ex-partner, or we may be less enthusiastic in our expression of happiness for him.

It is possible to experience a similar ambivalence with our condemnatory attitudes. Condemnatory attitudes conflict with *conciliatory* attitudes like pity, empathy, and sympathy. While the former lead us towards reproach, retribution or punishment, the latter lead us towards amenability, consolation, forgiveness, or mercy. The two

kinds of emotions pull in opposite directions. It is clear, however, that both condemnatory and conciliatory attitudes can be held towards the same subject at the same time. When they do so, we can see that the condemnatory attitudes may be weakened by the co-existence of conciliatory attitudes and the contrary feelings and behaviour to which these conciliatory attitudes incline me.

I suspect that this may be the moral psychological feature driving much of 'Equity and Mercy'. As we have seen, Nussbaum's conception of an improved *erklären* explanation involves focus on the 'obstacles' that each of us faces in our lives. Such an explanation provides us with a story, of an agent and her wrongdoing, which involves events that *happened to* her. Sometimes, as we develop a picture of a wrongdoing agent, we will begin to feel conciliatory attitudes towards her. Our feeling pity or compassion will, of course, depend upon what the picture shows us as having happened to the agent. We are perhaps most likely to pity her if she was herself the victim of abuse, malicious wrongdoing, or neglect, but our empathy or sympathy can arise in the face of other facts about her as well; perhaps someone she loved greatly died, or perhaps she has been the victim of a natural happening or some significant piece of bad luck. In the face of such knowledge about a wrongdoer, we may come to feel an emotional ambivalence towards her. On the one hand, we will feel condemnatory emotions towards her in the face of the fact that she intentionally harmed someone else, but on the other hand, we will feel conciliatory emotions towards her in the face of something we have learned about her life.

Perhaps I can pity a person and still wholly condemn her, if I think that what I pity her for has *nothing* to do with what I condemn her for. However, and more often, we will incorporate pity-inducing features of a wrongdoer's life into our understanding

of the agent's behaviour. I take, for example, Jones's childhood abuse to contribute to his tendency violent action as an adult. The result is a clash between my emotions over Jones's violent behaviour: I feel pity towards him in virtue of his suffering abuse, and yet I also feel condemnation towards him in virtue of his harmful behaviour. An influence upon Jones's behaviour, his past abuse, is something that brings about my compassion. Thus, adding *certain* kinds of *erklären* elements to an explanation of wrongdoing will lead us to feel, at the same time, both conciliatory and condemnatory attitudes towards the wrongdoer, with the overall result that the latter, or the *effects* of the latter, are weakened.

Much of 'Equity and Mercy' is concerned with narratives and their presentation of wrongdoers, and Nussbaum is explicit in endorsing narratives that lead to conciliatory attitudes. At the end of the paper, she describes the sort of explanation she sanctions:

> It is really interested in the obstacles to goodness … It judges these social forces … but, where judgement on the individual is concerned, it yields in mercy before the difficulty of life. This means that it can be in its form a powerful work of narrative art. If you really open your imagination and heart to admit the life story of someone else, it becomes far more difficult to finish that person off with a karate kick. In short, the text constructs a reader who, while judging justly, remains capable of love.[42]

I am suggesting that the best way to understand the reader Nussbaum describes in the final sentence of this passage is 'emotionally ambivalent', caught between feeling as if

she should condemn the wrongdoer, on the one hand, and 'loving' her, on the other.

Nussbaum predicts that some narrative art will embody the sort of explanation that she endorses, a prediction that is borne out perhaps most strikingly in narratives that are concerned to portray real wrongdoers. Several biographical films in recent years have presented the lives of real wrongdoers in such a way that while we have no inclination to excuse the wrongdoer, many viewers do find themselves with conciliatory attitudes towards him or her. *Dance With A Stranger* (dir. Mike Newell, 1985), *Dead Man Walking* (dir. Tim Robbins, 1995), and *Monster* (dir. Patty Jenkins, 2003) portray non-excusing, detailed accounts of intentional wrongdoings—all three are murderers. It is, of course, by no means necessary that a viewer will respond to these narratives with emotional ambivalence, but I suspect that such a response was the aim of some or all of the writers and directors involved. Nussbaum, I suspect, would approve of all of these films; they each bear the mark of their makers' compassion for their subjects.[43]

## 5. Conclusion

I have been concerned with the nature of, and the possible sources of, a tension between understanding a wrongdoing and condemning an agent for it. In Section 2, I claimed that *verstehen*$_{\mathrm{M}}$ explanations—which are not in tension with condemning a wrongdoer—provide evidence for concluding that if there is a tension, then it lies not between *any* explanation and condemnatory attitudes, but between *better* explanations and condemnatory attitudes. In order to discover candidates for a 'better explanation of a wrongdoing', I turned in the next two sections to look at two conceptions of wrongdoing; any theory of wrongdoing is going to give us a standard by which to judge

an explanation of a wrongdoing as better or worse.

First, we saw that Raimond Gaita argues that ignorance is at the bottom of malicious wrongdoing; this entails that *verstehen*$_{\mathrm{I}}$ explanations are better than *verstehen*$_{\mathrm{M}}$ explanations. Secondly, we saw that Martha Nussbaum argues that wrongdoing will, at least partly, be the result of non-rational influences; this entails that explanations of wrongdoings that include certain *erklären* elements will be better than those that do not.[44] The work of both authors suggest that a purely *verstehen*$_{\mathrm{M}}$ explanation is in an important way incomplete, that an understanding that allows us to see an act as malicious is necessarily oversimplified. Neither author, however, defends a purely *erklären* explanation of wrongdoing. On Gaita's picture of wrongdoing, I have suggested, a *verstehen*$_{\mathrm{I}}$ explanation will be an improvement over a *verstehen*$_{\mathrm{M}}$ one, while on Nussbaum's picture of wrongdoing, *verstehen* explanations will be improved by introducing *erklären* elements.

Although I have defended neither Gaita's nor Nussbaum's conception of wrongdoing, I find it plausible that accepting certain explanations of either kind will result in my experiencing weakened condemnation. Were I to follow Gaita, and change my view of wrongdoers from seeing them as malicious to seeing them as ignorant, it seems likely that it this will, indeed, bring about a weakening of my condemnatory attitudes towards them. Were I to follow Nussbaum, and incorporate the circumstances of wrongdoers' lives into my explanation of their wrongdoings, it seems likely that my doing so will similarly result in my pity, empathy, or sympathy towards them; such emotions would, I suspect, weaken my condemnatory attitudes towards wrongdoers. Both of these claims are conjectures, however, and empirical work will be needed to establish that either tendency to weaker condemnation holds.

My conclusion is, in sum, a conditional whose consequent is speculative: *if* either Gaita's or Nussbaum's respective views of wrongdoing is correct, then it *looks* as if there is a tension between having better understanding of someone's wrongdoing and holding condemnatory attitudes towards her.

I began with a French proverb and its variants. I have not yet established that they hold any truth, but if I have been right in this paper, then we know a bit more about what we need to establish in order to discover what truth, if any, this proverb and its variants contain, and why.[45]

---

**Notes**

1 Leo Tolstoy, 1868, Chapter 28.

2 *Corinne*, 1807.

3 *Torquato Tasso*, 1790.

4 At least one other writer in this collection, Thad Metz, takes the tension to be a normative phenomenon, e.g., that the more one understands, the less *reason* one has to condemn.

5 'Freedom and Resentment', 1962, reprinted in Gary Watson (ed.) *Free Will*, Oxford: Oxford University Press, 1982, pp. 59-80.

6 *Responsibility and the Moral Sentiments*, Cambridge, MA, Harvard University Press, 1996, pp. 82-3.

7 'Indignation: A case study in the role of feelings in morals', *Metaphilosophy*, **10,** April 1979, p. 139 and p. 140.

8 Jean Hampton and Jeffrie Murphy, *Forgiveness and Mercy*, Cambridge University Press, 1988, p. 89.

9 My conception of condemnation—allowing as it does for condemnation without

action or the desire for suffering—is looser than that found in other contributions to this collection.

10 The philosophical and sociological use of *verstehen* and *erklären*—and related concepts—to pick out two different modes of explanation, stretches back to Wilhelm Dilthey and Max Weber.

11 This example comes from Fred Dretske and Berent Enç's 'Causal Theories of Knowing', *Midwest Studies in Philosophy*, Peter French et.al. (eds.), Minneapolis: University of Minnesota Press, 1984.

12 Some writers, like Jane Heal, emphasize both of these features.

13 'Functionalism and Anomalous Monism' in *Perspectives on Actions and Events: The Philosophy of Donald Davidson***,** Ernest LePore and Brian McLaughlin (eds.), Oxford: Basil Blackwell, 1986, p. 389.

14 'Broad-minded Explanation and Psychology' in *Subject, Thought, and Context*, in Philip Pettit and John McDowell (eds.), Oxford: OUP, 1986, p. 38.

15 'Why Understanding?', a public lecture given at the Goethe Institute, Johannesburg, in November 2004. I thank Bernard for sharing a copy of this lecture with me.

16 Schlink, 'Why Understanding?', Section II, emphasis added. Although Schlink uses the neutral word 'judge', it is clear from his paper that by this he means *negative* judgement.

17 'Replication and Functionalism', in Martin Davies and Tony Stone (eds.), *Folk Psychology*, Oxford: Blackwell, 1995, p. 47. Originally in J. Butterfield (ed.) *Language, Mind, and Logic*, Cambridge: Cambridge University Press, 1986.

18 Robert M. Gordon, 'Sympathy, Simulation, and the Impartial Spectator', *Ethics*

**105,** July 1995, p. 733.

19 'Understanding "Understanding" in *The Reader*', Chapter 7 in this volume, Section I.

20 For reasons to question the claim that there is a tension between understanding and condemning oneself, see Samantha Vice's contribution to this volume.

21 As I read Andrew Gleeson's 'Humanizing Evil-Doers', Chapter 7 in this volume, he is arguing that the response of Stocker's first person is *better* than that of Stocker's second person, for the first person recognizes the 'vulnerability' common to all human beings. But this normative claim about condemnation does nothing to establish that *verstehen* explanations are in any tension with condemnation.

22 This conception of the tension fits better with two of the quotations with which I began this paper, which make reference to understanding 'all' ('*tout comprendre*').

23 Wallace, *Responsibility and the Moral Sentiments*, pp. 136-7.

24 Wallace, *Responsibility and the Moral Sentiments*, p. 138.

25 I do not claim that these two categories exhaust all *verstehen* explanations.

26 'Evil Beyond Vice', the second chapter of his *A Common Humanity: Thinking About Love and Truth and Justice* (London, Routledge: 2000).

27 Gaita, *A Common Humanity*, p. 43.

28 Gaita, *A Common Humanity*, p. 43.

29 Schlink, 'Why Understanding?', Section II.

30 The thought of this paragraph has a long pedigree, but the early influential 20thC statements of it include G.E.M. Anscombe's *Intention*, Oxford: Blackwell, 1957;

Stuart Hamphire's *Thought and Action*, London: Chatto and Windus, 1959; as well as Strawson's 'Freedom and Resentment', 1962.

31 Nussbaum, 'Equity and Mercy' (****in this collection, Ch 5), p. 100.

32 Such a theory may be worked out in detail (e.g., if you are a professional psychologist or sociologist), or it may be held only in vague brushstrokes.

33 Nussbaum speaks about influences in both categories; she writes, for example, of our 'socially taught scheme of values' giving rise to 'diseased passions', like 'excessive attachment to sex' and 'anger and violence connected with sexual jealousy' ['Equity and Mercy', p. 100].

34 Nussbaum, 'Equity and Mercy', p. 99. I have removed the words 'and internal' from this sentence, having discussed identification in the previous section.

35 Nussbaum, 'Equity and Mercy', p. 100.

36 Strawson, 'Freedom and Resentment'.

37 *The Abuse Excuse*, Boston: Little, Brown and Company, 1994, p. 55.

38 'Equity and Mercy', *Philosophy and Public Affairs* **22**(2), Spring 1993, p. 99. This volume, Chapter 5.

39 'Responsibility and the Abuse Excuse', *Social Philosophy and Policy*, 16(2), 1999, p. 189-90.

40 For philosophical discussion of this point, see the papers collected in Part II of *Genetics and Criminal Behavior*, David Wasserman and Robert Wachbroit (eds), Cambridge**:** Cambridge University Press, 2001, especially contributions by Marcia Baron, Patricia Greenspan, and David Wasserman.

41 In A.O. Rorty (ed.), *Explaining Emotions*, Berkeley CA, University of California Press, 1980, p. 228.

42 Nussbaum, 'Equity and Mercy', pp. 124-5.

43 Bernard Schlink portrays such an emotional tension in his novel *The Reader*. Watching Hanna's trial, Michael feels an intense contempt for her wrongdoing, but as he comes to realize the affect that her illiteracy has had on her, this contempt becomes mixed with a pity towards her. Above all else, *The Reader* is about the pitiable circumstances involved in human wrongdoing, and the mixed emotions that we can feel when we are learn about them.

44 I take it that these two views of wrongdoing are not mutually exclusive, but I will not defend this here.

45 Thanks to Thad Metz, Mac Rebennack, Pedro Tabensky, Samantha Vice, and the members of an audience at the University of Oklahoma.