

Article

Psychopathy: what apology making tells us about moral agency

Ayob, Gloria Leila and Thornton, Tim

Available at <http://clock.uclan.ac.uk/7255/>

Ayob, Gloria Leila ORCID: 0000-0001-5374-2161 and Thornton, Tim ORCID: 0000-0002-0137-1554 (2014) Psychopathy: what apology making tells us about moral agency. Theoretical Medicine and Bioethics, 35 (1). pp. 17-29. ISSN 1386-7415

It is advisable to refer to the publisher's version if you intend to cite from the work.

<http://dx.doi.org/10.1007/s11017-014-9279-3>

For more information about UCLan's research in this area go to <http://www.uclan.ac.uk/researchgroups/> and search for <name of research Group>.

For information about Research generally at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Psychopathy: what apology making tells us about moral agency

Abstract

Psychopathy is often used to settle disputes about the nature of moral judgement. The Trolley Problem is a familiar scenario in which psychopathy is used as a test case. Where a convergence in response to the Trolley Problem is registered between psychopathic subjects and non-psychopathic (normal) subjects, it is assumed that this convergence indicates that the capacity for making moral judgments is unimpaired in psychopathy. This in turn is taken to have implications for the dispute between motivation internalists and motivation externalists, for instance. In what follows, we want to do two things: firstly, we set out to question the assumption that convergence is informative of the capacity for moral judgement in psychopathy. Next, we consider a distinct feature of psychopathy which we think provides strong grounds for holding that the capacity for moral judgement is seriously impaired in psychopathic subjects. The feature in question is the psychopathic subject's inability to make sincere apologies. Our central claim will be this: convergence in response to Trolley Problems doesn't tell us very much about the psychopathic subject's capacity to make moral judgements, but his inability to make sincere apologies does provide us with strong grounds for holding that this capacity is seriously impaired in psychopathy.

Psychopathy: what apology making tells us about moral agency

I. Introduction

In philosophical debates about the nature of moral judgement, the case of psychopathy is often used to test various claims. For instance, it has been used to attack motivation internalism, as well as certain versions of moral rationalism. Claims made about moral judgement in psychopathy are based on empirical tests, which usually consist of questionnaires that feature hypothetical scenarios involving a conflict between two moral imperatives.¹ A commonly used scenario in philosophical discussions features the so-called Trolley Problem. Where a convergence in response to a hypothetical dilemma is registered between psychopathic subjects and non-psychopathic (normal) subjects, this convergence is generally taken to indicate that the capacity for making moral judgments is unimpaired in psychopathy. From this, the motivation externalist, for instance, draws the conclusion that the amoral behaviour observed in psychopathy shows that moral judgement and the motivation to act are merely contingently related. [12]

The ascription of the capacity for moral understanding to psychopathic subjects on the basis of convergence in response to hypothetical dilemmas (such as to Trolley Problems) assumes that such convergence is indeed revelatory of this capacity. In what follows, we want to do two things: firstly, we set out to question this assumption. Next, we turn to consider a distinct feature of psychopathy which we think provides strong grounds for holding that the capacity for moral judgement is seriously impaired in psychopathic subjects. Our central claim will be this: although convergence in response to Trolley Problems doesn't tell us very much about the psychopathic subject's capacity to make moral judgements, there is something distinctive about psychopathic behaviour which does shed light on this capacity. The feature in question is the psychopathic subject's inability to make sincere apologies.

This aspect of moral cognition has so far received little attention in the literature, but it seems to us to be essential in determining whether or not a person is capable of forming and retaining moral beliefs. The practice of apology-making is an important feature of our moral lives, and we intuitively take it as a measure of

¹ This experimental paradigm was originally articulated by Lawrence Kohlberg, in the context of studying the ontogenetic development of socialization and the capacity for moral judgement. [3] More recently, philosophers and cognitive scientists have placed the Trolley Problem within this experimental paradigm to explore the patterns of judgement made by competent adult subjects. A pioneering paper in this more recent exploration is Greene et al 2001. [5]

a person's moral awareness.² The psychopathic subject's inability to make sincere apologies invites us to look more closely at the pre-theoretical importance that we place on the practice of apology-making in determining a person's capacity for moral judgement. In Section III of this paper, we will try to explicitly connect apology-making with moral understanding. We think that the importance that we ordinarily place on the ability to make apologies in determining a person's capacity for moral understanding will be shown to be warranted. This shows therefore that the psychopathic subject's capacity for moral understanding is seriously impaired.

II. Trolleyology: what does the convergence in judgement tell us?

The Trolley Problem comes in two key variants, one which requires the agent to inflict so-called 'personal harm' and the other which requires the agent to inflict 'impersonal harm'. The two scenarios are summarised in Greene et al 2001 as follows:

A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Ought you to turn the trolley in order to save five people at the expense of one? Most people say yes. Now consider a similar problem, the footbridge dilemma. As before, a trolley threatens to kill five people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Ought you to save the five others by pushing this stranger to his death? Most people say no. [5]³

Because the second scenario described requires the agent to come into direct physical contact the victim, the agent is described as inflicting 'personal' harm on the victim. The first scenario does not require any direct physical contact between agent and victim, and so is described as a case of 'impersonal' harm.

A recent paper reports that psychopathic and non-psychopathic (normal) subjects respond in similar ways to the two variants of the Trolley Problem just sketched: like most subjects in the normal population, subjects in the psychopathic population tend to judge that we should hit the switch in the first scenario but not push the man off the bridge in the second. [2]^{4,5} A convergence in judgement between psychopathic subjects and non-psychopathic (normal) subjects—in the Trolley scenarios as well as other types of scenario—is generally taken to indicate that the capacity for making moral judgments is unimpaired in psychopathy. On the basis of this observed convergence, various substantive conclusions have been drawn about the nature of moral judgement, e.g. that emotional processes are not causally necessary for moral judgement [2], or that the motivation to act is merely contingently related to moral judgement (since psychopathic subjects display amoral behaviour). [12]

These conclusions rest on one important assumption: that convergence between the normal (non-psychopathic) population and the psychopathic population provides us with sufficient grounds for ascribing to the psychopathic subject the capacity to make moral judgements. This in turn presumes that the judgements made by the majority of subjects in the normal population are the *morally right* ones to make. In other words,

² That we do this is most clearly exemplified by judicial remarks. In delivering their judgements, it is not uncommon for judges to remark on the absence of remorse shown by the defendant, in cases where the defendant has shown no remorse throughout the trial.

³ Note that Greene et al 2001 don't touch on the subject of psychopathy; they are interested in explaining the difference in judgement that ordinary subjects give across the two scenarios.

⁴ This is the latest finding in a fairly long history of findings that report a similarity in patterns of moral judgement between psychopathic and non-psychopathic populations, although note that earlier studies did not use the Trolley Problem (see e.g. [1],[8], [14]).

⁵ It should be noted, however, that contrary findings were reported in slightly earlier studies, i.e. in contrast with the majority of subjects in the normal population, psychopathic subjects tended to judge both that the switch should be hit in the first scenario and that the man ought to be pushed off the bridge in the second scenario. [6], [7]

it presumes that the majority of subjects in the normal population are, by and large, successful at tracking moral norms. This assumption is questionable, at least with respect to the trolley problems. Quite independently of the merits of consequentialist views of morals, the inconsistency in judgement that normal subjects make—between the ‘personal harm’ case and the ‘impersonal harm’ case—should be a cause for concern. Why should pushing someone directly into harm’s way, as opposed to hitting a switch that will certainly run him over, be a morally significant difference? There may be psychologically significant differences, but unless we are prepared to identify these psychologically significant differences as morally significant ones, the inconsistency seems puzzling — and it seems to suggest that most normal subjects do not always make the best moral decisions.

Pertinent to the present point is that most, but not all, subjects in the normal population manifest this inconsistency in responding to the two trolley problems. A minority of subjects in this population do in fact judge that the trolley should be stopped from killing the five people in both cases, i.e. regardless of whether doing so involves personally or impersonally harming another person. A consequentialist would argue that this is the morally right decision to make. In this vein, Peter Singer argues as follows:

In some people, reasoning can overcome an initial emotionally based response. ...The answer that these [minority] subjects gave is, surely, the answer that a rational being would give. We ought to minimize the loss of innocent life, unless there is some valid reason against doing so. [13:196-7]

The fact that we happen to be biologically programmed to respond emotionally aversely to pushing a person, but not to hitting a switch that will ultimately cause the same person to be killed, is not, for Singer, a morally salient fact - even if it is a psychologically interesting fact. So pushing someone directly onto the path of the trolley does not, for Singer, constitute a valid reason against minimizing the loss of innocent life in the ‘personal harm’ version of the trolley scenario.

Singer puts his point in terms of emotional responses and moral reasoning, which might look problematic in light of the findings reported in Cima et al, 2010. [2] Recall that, according to this study, psychopathic and the majority of normal subjects converge in their judgements in the two versions of the trolley problem. The authors conclude on the basis of this convergence that these judgements cannot consist in, or be caused by, emotional responses since psychopathic subjects have deficits in emotional processing. So Singer’s division between emotional responses and moral reasoning to account for different patterns of judgement in the normal population (majority vs. minority) might seem to be erroneous for empirical reasons. But there is a deeper point that Singer is making, that is not hostage to the empirical question of whether the majority of subjects in the normal population are responding merely emotionally or otherwise in the ‘personal harm’ variant of the trolley problem. To bring out the deeper point, we’ll label the response to the ‘personal harm’ variant—given by both the psychopathic population and the majority of subjects in the normal population—the “X response”. Now Singer’s point is that by the consequentialist’s lights, the X response is morally wrong even if the majority of subjects in the normal population manifest this response. There is a conceptual gap between the moral judgements that people do in fact make (in this case, the X response) and the judgements they ought to make, as determined by a normative ethical theory. Therefore, even if psychopathic subjects converge in their judgements with the majority of the normal population, this does not yet show that they are capable of making moral judgements. It may just show that a large number of subjects in the normal population aren’t very good at making moral judgements.

Against this, it may be said that the ordinary subject’s moral understanding—i.e. a subject who belongs to the majority subset in the normal population—is a good enough understanding, and that we cannot expect ordinary subjects to be perfect in their moral decision-making. The capacity for making moral judgements is not, nor should it be thought of as being, *infallible*. However, this isn’t a happy response for those wanting to ascribe to the psychopathic subject the capacity for moral judgement—on the basis of convergence between the psychopathic subject and the majority of normal subjects—because it concedes that the majority of normal subjects make the wrong moral judgement in the ‘personal harm’ case, and a convergence in an erroneous judgement can surely tell us nothing about the capacity to make moral judgements either in psychopathic or in the majority of normal subjects.

In any case, there is a more basic difficulty about reasoning from convergence that Singer’s discussion brings out, which is best seen in the following counterfactual case: suppose the psychopathic subject’s judgement in the ‘personal harm’ variant of the trolley problem turned out *not* to resemble the judgement made by the

majority of normal subjects (suppose, that is, that the psychopathic subject judges that we should push the man over, just as we ought to hit the switch, to save five people). Then, since convergence is being offered as the reason for ascribing moral understanding to the psychopathic subject (importantly, as the sole reason for doing so), the failure of convergence should lead us to retract the ascription of moral understanding to the psychopathic subject. But that is precisely what we shouldn't do, according to Singer, because there is a conceptual gap between the X response (i.e. the response displayed by the majority of normal subjects) and the response that one ought to give, as determined by a normative ethical theory. This difficulty may be spelt out as follows:

- (I) if there is convergence, it indicates a similarity in the X response between psychopathic and the majority of normal subjects (and the X response is not, for normativists like Singer, the same thing as making a moral judgement) and
- (II) If there is a failure of convergence, it may indicate: either (i) a dissimilarity in the X response or (ii) a similarity in the tendency to make, or see the attraction of, the X response but the capacity on the part of the psychopathic subject to override that response using reason (although we cannot say yet whether that reason involves the exercise moral understanding. That is, we cannot say just yet whether the psychopathic subject is exercising *moral* reason specifically)

For the normativist, it is because the X response does not necessarily constitute an exercise of moral judgement that nothing interesting about the psychopathic subject's capacity for moral judgement follows from convergence (or the lack thereof). Crucially, for the normativist, a failure of convergence may in fact demonstrate that the psychopathic subject *both* shares the attraction to make the X response with most subjects in the normal population *and* is capable of making moral judgements (Option (II)(ii)). It is this overlooked possibility that renders any attempt to draw conclusions on the basis of convergence (or otherwise) invalid. The only way to block option (II)(ii) would be by identifying the X response with moral judgement as a matter of principle, so that a moral judgement just is whatever the X response is. This would render incoherent the idea that a judgement could fail to converge with the X response and nonetheless constitutes a successful exercise of the capacity for moral judgement. And it is only by ruling out this possibility that convergence—or the failure thereof—tells us anything interesting about the psychopathic subject's capacity for moral judgement. But blocking option (II)(ii) by identifying the X response with moral judgement as a matter of principle is a substantive commitment, and a controversial one too. Much ink has been spent defending the normativist conception of ethics, according to which moral 'oughts' are distinct from facts about how we do in fact respond to morally charged situations. We are sympathetic to the normativist position, and thus we find it unattractive to identify the X response with moral judgement as a matter of principle. (For recent defences of the normativist position, see [9], [13]). For our present purposes, the following observation suffices: drawing any conclusion about the capacity for moral judgement in psychopathy from the convergence (or lack thereof) between psychopathic and normal subjects in response to trolley problems requires making a controversial assumption, namely that moral reasoning is simply a function of the patterns of judgement that are in fact observed in the majority of the population. If we reject this assumption, and hold instead that moral reasoning involves adhering to moral norms that aren't simply a function of the patterns of judgement displayed of the majority (in other words, if moral norms aren't just a statistical norm), then convergence (or the failure thereof) with the majority in solving trolley problems tells us nothing about a person's capacity to make moral judgements. In other words, trolley problems tell us very little, if anything, about whether or not the capacity for moral judgement is impaired in psychopathy. We'll now turn to a distinct consideration which we think *does* show that the psychopathic subject's capacity for moral judgement is seriously impaired.

III. Being apologetic: an exercise of the capacity for making moral judgements

Whilst the Trolley Problem has been widely discussed in (both its own right but also) its application to psychopathy, there is a more direct connection between psychopathy and moral understanding which is much less explored. This is the connection between psychopathy, a lack of remorse, and moral understanding.

Psychopathy is defined in part by an inability to feel remorse and this, we think, reveals that the psychopathic subject's capacity for moral understanding is seriously impaired.⁶

The feeling of remorse is an essential part of making a sincere apology. When someone issues an apologetic remark without seeming remorseful, we naturally doubt that a (sincere) apology has been made. This feature of our everyday practice of apology making warrants the following observation:

- (A) Making an apology requires appropriate affective attunement, the most relevant affect being the feeling of remorse.

We will restrict ourselves to a relatively weak reading of this claim, by granting that a subject need not necessarily feel remorse every single time she make a sincere apology. What is required is that the subject is *capable* of feeling remorse, and must engage this capacity on most occasions when she makes an apology. So a subject who utters a sentence that is conventionally used to make an apology, but who is dispositionally incapable of feeling remorse, does not succeed in making an apology.

Broadly, there are two conceptions of moral understanding, a cognitivist conception and a non-cognitivist one, and it is a short step from (A) to the failure of moral understanding on the latter conception. The non-cognitivist holds that moral understanding consists not in making any judgement with objective purport, but is rather an expression of our affective responses. Of course not all our affective responses have moral significance; only those that figure in the attitudes that we recognise as moral will have this significance. Insofar as apology-making is a morally significant practice (and it does seem to be as central to our moral lives as intending to act in accordance with moral principles), the affective attunement that is essential to apology-making will consist in affective responses that have moral significance. A failure to be appropriately affectively attuned when an apology is called for thus arguably constitutes a failure of moral understanding, according to the non-cognitivist conception of moral understanding. Regardless of how moral understanding is articulated—whether as an error theory, projectivism or quasi-realism—it is anchored in the feeling of remorse in the case of apology making. Whatever the precise role of this feeling is in the theoretical account, its centrality suggests a key deficit in the moral understanding of those suffering from psychopathy, according to the non-cognitivist picture. A lack of remorse implies an inability to make sincere apologies and hence a key lack in moral awareness.

But the inability of the psychopathic subject to feel remorse and thus to make sincere apologies arguably implies a serious impairment in the capacity for moral understanding even according to the cognitivist conception of what this understanding consists in. Furthermore, the cognitivist conception underscores certain further features of apology making which help to illuminate the failure of moral understanding manifest in psychopathy. Consider, first, the fact that knowing when an apology is, or would, be called for is having the capacity to recognise some of the demands that are placed on moral agents by their social surroundings. Sincere apology doesn't seem to be just a matter of an affective response of the right subjective colour, as it were. The affective response has to be appropriately wrung from the subject by their circumstances. We may state put the point as follows:

- (B) Apology making essentially involves recognising, of a relevant particular situation, that an apology is called for or required.

An important reason why apology making seems to involve the engagement of recognitional capacities is that the appropriateness conditions for the affective response wrung from the subject are considerably complex. To appreciate the complexity of the recognitional demands placed on an apology maker, notice that subjects have to distinguish between those situations which elicit empathic responses and which do call for an apology, from those situations in which an apology is not called for but which nonetheless do evoke an empathic response. Let's consider the following scenario: in a fit of rage, Alex punches Ben, as a result of which Ben's

⁶ According to the ICD-10, one of the diagnostic conditions of dissocial personality disorder (which we commonly refer to as "psychopathy") is that the subject is incapable of feeling guilt (ICD-10, p.152); and the upcoming DSM-5 takes a "lack of remorse after hurting or mistreating another in order" to be a diagnostic marker of this disorder.

ICD-10: <http://www.who.int/classifications/icd/en/GRNBOOK.pdf>; DSM-5 (proposals): <http://www.dsm5.org/ProposedRevision/Pages/proposedrevision.aspx?rid=16>

nose starts to bleed. As it happens, there happen to be two onlookers, Chris and Dan. Alex, Chris and Dan can all see that Ben has been physically hurt. But Dan, let us suppose, is psychopathic and so, unlike Chris and Alex, does not see anything more than this. Affectively, he perceives this scene no differently to a scene consisting of movements involving just inanimate objects (for instance, a wave crashing into a boulder on the shore). But intuitively, there is more to be seen. One obvious thing that is missing from Dan is an empathic response. But whatever a subject sees or grasps when a situation elicits an empathic response from the subject, even this content alone is not sufficiently discriminating to characterise what a subject recognises when he feels apologetic, since Chris does not feel apologetic towards Ben (nor should he feel apologetic) in spite of the fact that he feels empathically for Ben, i.e. he “feels Ben’s pain”.⁷ Moral agents discriminate between (i) conditions that elicit an empathic response on their part but for which no apology is called for, and (ii) conditions that elicit an empathic response on their part and for which an apology is called for. What distinguishes Alex and Chris, and places Alex in the latter condition, seems to be a self-recognitional element: Alex recognises not just that Ben is suffering—recognising this in a way that elicits his empathy—but that something he has done has brought about this suffering (the point applies equally well for omissions as it does for commissions). Thus the recognitional component (B) does not just depend on features of a moral agent’s social environment, features having to do with the suffering of others, but also the relevant responsibility of the agent herself. For this reason, we’ll call it a dyadic trope.

This example highlights the complexity of the recognitional demands that are placed on a moral agent who makes an apology: the content the agent grasps involves a reflexive pairing of agent and ‘object’ of apology. Because apology making involves recognising that an apology is called for (when indeed it is called for), and because what is recognised is something with a complex structure (i.e. the reflexive dyadic trope), apology making thus seems to engage the subject’s recognitional capacities. For this reason, we think that making an apology constitutes an exercise not just of moral understanding (as a non-cognitivist might put it) but of moral *judgement*. There is another feature of apology making reinforces this claim: apology making carries with it certain consequences for the subject who makes the apology. Of particular interest to our present concerns is the fact that the subject accrues certain obligations in virtue of making the apology (one obligation that frequently arises is the obligation to provide restitution to the injured party who is the target of the apology). A subject who frequently attempted to make apologies but was deaf to the relevance, sometimes at least, of restitution could not be said to have grasped the content of apology. Thus, in grasping what is owed by way of an apology, a sincere subject has to understand some of the consequences that are *rationally* connected to the making of the apology itself. This indicates that, in making an apology, subjects grasp a *content* that is inferentially potent.

Humean commitments may motivate the concern that the judgement involving nature of apology summarised in (B) stands in conflict with (A). But to accept that they are incompatible will require providing a revisionist account of at least certain aspects of the practice of apology-making, since on the face of it both (A) and (B) hold of sincere apology. Crucially, the intuitive inseparability of feeling remorse and recognitional judgement suggests that claims (A) and (B) must be related non-accidentally. If the two claims were related merely accidentally to each other, then it would be conceptually possible to recognise that an apology is called for, and to exercise the recognitional capacities involved in making an apology, and yet be incapable of feeling remorse. But it is precisely this point that seems counterintuitive, which indicates that feeling remorse is a constitutive part of the exercise of the recognitional capacities involved in making an apology. In other words, claims (A) and (B) seem to be related to each other non-accidentally. The most obvious way to respect this relation is to accept that what it is to recognise that one stands in the appropriate circumstances (paradigmatically including responsibility for suffering) is to have the right affective response. (Equally, we think that no merely affect-based account can account for element of judgement. An apparently sincere avowal of remorse stripped of the other recognitional capacities would not count as genuinely one of remorse). Consider another, late, arrival to the mêlée described above: Elvis, who has kicked the fallen Ben full in the face. Suppose that this were an accident, the result of where Ben lay, partly hidden. Elvis may grasp Ben’s hurt and his own partial responsibility. He might regret the fact that he entered the room at that moment. In the UK, at least, he would be likely to utter a brisk “I’m so sorry, old chap”. But whilst his regret and sympathy might be sincere, he could feel no appropriate remorse and without a corresponding sense of fault, his utterance could not be a proper apology. Suppose instead that Elvis’ kick were not so much an accident as a deliberate opportunistic exploration of what kicking another feels like. Perhaps Elvis is

⁷ For simplicity, we’re assuming that Alex acted alone, i.e. he wasn’t goaded on by any of the others.

psychopathic. He might still recognise Ben's pain (though not value it so highly as not to kick him) and know himself to be its agent. But he feels no remorse. We contend that this absence of affective attunement to the situation just is his failure to recognise its nature.

A non-revisionist cognitivist view of apology (i.e. one aiming to uphold both (A) and (B)) thus has to make sense of this relation in something like the following terms: it is the subject's affective attunement that places her in a position to grasp the moral facts germane to apology-making. In being incapable of attuning himself affectively in the appropriate way, particular moral facts are blocked off completely from the psychopathic subject's view. And it is because he is blind to certain moral facts that the psychopathic subject is unable to make certain moral judgements, i.e. those moral judgements that are constitutively involved in apology-making. Since the dyadic trope sketched out above is intensionally characterised using the affectively-charged term 'suffering', the grasp of this dyadic trope does seem to consist, at least in part, in an appropriate affective attunement on the subject's part.

The cognitivist picture of apology-making just sketched out thus posits the existence of moral facts the grasp of which require affective attunement. To describe what a moral agent grasps—in recognising that he owes another person (or party) an apology now—as a *moral* fact thus requires us, firstly, to acknowledge the existence of particular moral facts, in addition to general moral principles. That such facts exist is a claim that has long been argued for by moral particularists such as John McDowell and Jonathan Dancy. [4], [10] But even if it is accepted that particular moral facts do exist, the very idea that there could exist such facts the grasp of which require affective attunement is striking. We are familiar with physical facts, the grasp of which require subjects to be appropriately perceptually attuned. What is less familiar but has been discussed by philosophers such as McDowell is the idea of logical or mathematical or even moral features of the world, the grasp of which require subjects to be appropriately intellectually or rationally attuned [10]. McDowell likens moral features of the world to secondary qualities, though the analogy is contested by Crispin Wright, for example [11], [15]). On this picture, there is an essential relation between the moral features which a knowledgeable agent is able to recognise and his or her responses to them. But apology seems to suggest a novel twist on even this picture. We have already suggested that when a subject is sincerely apologetic, it seems right to describe the subject as *recognising* that an apology is called for but also that sincerity, generally at least, requires a feeling of remorse. On the broadly cognitivist model of these two features we favour, they come together in the grasp of a reflexive dyadic trope.⁸

Summarising the discussion in this section, we may say the following: claim (A) should be acceptable to both cognitivist and non-cognitivist accounts of moral understanding alike. So, in virtue of his inability to feel remorse, the psychopathic subject suffers a serious impairment in moral understanding according to both of these two conceptions of moral understanding. But in addition to claim (A), the cognitivist endorses claim (B), and the latter claim helps shed light on certain further aspects of apology making (i.e. its inferential import,

⁸This conclusion, i.e. that affective attunement is integral to moral understanding, might concern the cognitivist who wishes to hold that affective attunement is wholly irrelevant to moral understanding. An austere Kantian might be motivated to uphold this claim, but it seems to us that a Kantian of this sort is faced with a choice between two options, neither of which seems obviously palatable. On the one hand, the Kantian could accept that we have a moral duty to make apologies when called for, and then try to block the necessity of affective attunement for moral understanding by denying that the feeling of remorse is essential to apology making. The difficulty with this option is that it requires us to adopt a revisionary stance towards the concept of apology making, i.e. of what we *mean* by an apology.

In order to avoid this difficulty, the austere Kantian might instead opt to deny that any moral *duty* attaches to apology making — so that it doesn't matter for our capacity for moral understanding that apology making requires the feeling of remorse essentially. But this claim does not seem palatable either because a subject's resistance towards making an apology when called for, in a particular instance, surely cannot be generalised into a universal law. A moral practice in which this resistance is generalisable would look very alien to us. Now this in itself may not be fatal to the austere Kantian's position, since how things appear to us may not be the final arbiter in determining what the moral law is. However, if the austere Kantian does proceed along this path, he owes us an account of how no specifically *moral* duty attaches to apology making (i.e. he needs to demonstrate that the practice of apology making forms merely a part of social niceties or pragmatics, and thus carries with it no moral duty or obligation).

but especially its recognitional demands), which turn out to be helpful in suggesting where the source of the failure of moral understanding manifest in psychopathy may lie.

III. Conclusion

We have argued that the convergence in patterns of judgement between the normal population and the psychopathic population, in response to trolley problems, tells us nothing very much about the capacity for moral judgement in psychopathy; but that the psychopathic subject's inability to feel remorse does provide us with grounds for holding that the capacity for moral understanding is seriously impaired in psychopathy. The latter claim rests on the connection between the ability to feel remorse, which is essential to apology-making, and apology-making as manifesting the capacity for moral understanding. The purpose of Section III of this paper has been to suggest two ways of spelling out this connection—as constrained by the two main conceptions of moral understanding—therein rendering intelligible the importance that we pre-theoretically place on a person's ability to make apologies in ascertaining his capacity for moral understanding. If this measure of moral understanding is well grounded, it follows that we ought to conclude from the psychopathic subject's inability to feel remorse that his failing is not simply one of motivation, but of understanding. The psychopathic subject doesn't simply fail to be motivated to act morally, having grasped how he ought to act; more fundamentally, his capacity for moral understanding is seriously impaired.

The discussion in Section III underscores one further difficulty about the argument from convergence which we criticised in Section II, a difficulty which we shall formulate in cognitivist terms. According to the cognitivist characterisation of apology making proposed in Section III, the psychopathic subject's moral understanding is impaired because he fails to grasp particular moral facts. This draws attention to the fact that mere convergence in judgement doesn't tell us anything about why the convergence happens, and in particular about whether the convergence occurs because of anything to do with the moral sensibilities of subjects in the two populations (the normal population and the clinical population). If we are right about what follows from the inability to feel remorse, then the psychopathic subject does not see what ordinary moral subjects see, in the two trolley situations. The ordinary moral subject may see certain features of the second trolley situation (the 'personal harm' situation) which causes her to feel revulsion at the very idea of pushing someone directly into the path of the trolley (say); whereas perhaps for the psychopathic subject, the difference between pulling a switch and pushing someone over is simply a matter of effort: it takes more effort to push someone over (which is why he judges that he wouldn't push the person over) than it does to pull a switch (which is why he judges that the right thing to do would be to pull the switch). In short, while the pattern of judgement displayed by most normal subjects may be determined by those subjects' moral sensibilities, the very same pattern displayed by a psychopathic subject might in fact be determined by considerations that aren't specifically moral (i.e. considerations of expediency). Mere convergence in judgement alone leaves open this possibility, and the discussion in Section III gives us reason to suspect that there are indeed disparate reasons why the subjects in the two populations make the judgements that they do. Thus, from convergence alone, we can draw no substantive conclusion about the capacity for moral understanding in psychopathy.^{9, 10}

⁹ Because psychopathic subjects are noted for their selfishness, it would be of interest to the empirical investigation to introduce a selfish motive into the trolley problem, and then to see if there would still be a convergence in judgement especially in the 'impersonal harm' case. Suppose we added to this version of the problem the following detail: "Pulling the switch and derailing the trolley so that it only hits one person (rather than five) will entail severe delays to the train you are waiting for, because that single person is standing close to the platform that your train is due to arrive at. This will mean that you will miss most of the film you were travelling to see." With this detail added, it would be interesting to find out whether the psychopathic subject would still say that he would pull the switch.

¹⁰ A general observation about the experimental paradigm as a whole: we should be wary of findings about psychopathy made on the basis of interviewing psychopathic subjects about hypothetical scenarios since psychopathic subjects are known not to be especially concerned about speaking truthfully. They are also known to be capable of turning on the charm.

Acknowledgements

We would like to thank an anonymous referee for helpful comments on an earlier draft of this paper.

References

1. Blair, R.J. 1995. 'A cognitive developmental approach to morality: investigating the psychopath.' *Cognition*, 57, 1-29.
2. Cima, M., Tonnaer, F., & Hauser, M. D. 2010. 'Psychopaths know right from wrong but don't care.' *Social Cognitive And Affective Neuroscience*, 5(1), 59-67.
3. Colby, Anne and Lawrence Kohlberg. 1987. *The Measurement of Moral Judgment, Volume 1: Theoretical Foundations and Research Validation*. Cambridge: Cambridge University Press.
4. Dancy, J. 1993. *Moral Reasons*, Oxford: Blackwell
5. Greene, J. D., Sommerville, R., Nystrom, L. E., Darley, J. M., & Cohen, J. D. 2001. 'An fMRI Investigation of Emotional Engagement in Moral Judgment.' *Science*, 293(5537), 2105-2108.
6. Greene, J.D. 2007. 'Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains.' *Trends In Cognitive Sciences*, 11322-323.
7. Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. 2007. 'Damage to the prefrontal cortex increases utilitarian moral judgements.' *Nature*, 446(7138), 908-911.
8. Lee, M., Prentice, N.M. 1988. 'Interrelations of empathy, cognition, and moral reasoning with dimensions of juvenile delinquency.' *Journal of Abnormal Child Psychology*, 16, 127-39.
9. Parfit, D. 2011. *On What Matters*. Oxford: Oxford University Press.
10. McDowell, J. 1979. 'Virtue and Reason' *Monist* 62: 331-350.
11. McDowell, J. 1985. 'Values and Secondary Qualities.' in Honderich, T. (ed) *Morality and Objectivity: A Tribute to J.L. Mackie*, London & Boston: Routledge & Kegan Paul.
12. Roskies, A. A. 2003. 'Are ethical judgments intrinsically motivational? Lessons from 'acquired sociopathy'.' *Philosophical Psychology*, 1651-66.
13. Singer, Peter. 2011. 'Afterword' in *The Expanding Circle: Ethics, Evolution, and Moral Progress*. New Jersey: Princeton University Press.
14. Trevathan, S., Walker, L.J. 1989. 'Hypothetical versus real-life moral reasoning among psychopathic and delinquent youth.' *Development and Psychopathology*, 1, 91-103.
15. Wright, C. 1988. 'Moral values, projection and secondary qualities' *Proceedings of the Aristotelian Society* supplementary volume 62.