Journal of Moral Education
Vol. 37, No. 3, September 2008, pp. 333–356

Routledge
Taylor & Francis Group

Towards a new paradigm of moral personhood

Jeremy A. Frimer* and Lawrence J. Walker
University of British Columbia, Canada

Moral psychology is between paradigms. Kohlberg's model of moral rationality has proved inadequate in explaining action; yet its augmentation—moral personality—awaits empirical embodiment. This article addresses some critical issues in developing a comprehensive empirical paradigm of moral personhood. Is a first-person or a third-person definition of moral behaviour more appropriate? Is operative moral judgement better understood as deliberative or intuitive? What is the essential nature of the moral self? Two basic constructs of moral personality which have been posited to help span the judgement–action gap—moral centrality and integrity—are critically reviewed and some criteria are proffered for evaluating competing models of moral personhood. Significant directions for future research are noted with the hope of moving the field towards a new paradigm of moral personhood. While the content of this paradigm will differ markedly from Kohlberg's, we contend that the spirit of his enterprise will be manifest with vigour redoubled.

Introduction

Moral psychology is controversial stuff. Every psychologist who studies moral topics has had the experience of trying to share some fascinating nuance of his or her research with a (non-academic) friend only to encounter stiff resistance. The friend is not ready to concede the existence of morality in the first place, what morality means in the second and what gives the scholar any authority to determine right or wrong in the third. Controversial issues pervade the academic study of moral topics just the same and have since the field's inception.

Fifty years ago, Kohlberg launched the field of moral psychology with bold claims about the nature of morality and he simultaneously endowed the field with the empirical means by which to investigate these claims. His structural-developmental model (1969, 1981, 1984) outlined a stage sequence of moral cognition, with one of the main premises being that moral rationality encompassed the heart of moral functioning. The field has since largely moved past this core premise that vaulted

*Corresponding author: Department of Psychology, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada. Email: jeremyfrimer@gmail.com

ISSN 0305-7240 (print)/ISSN 1465-3877 (online)/08/030333-24

DOI: 10.1080/03057240802227494

moral thought upon a pedestal.[1] That is, most now agree that Plato's famous dictum that 'to know the good is to do the good' is empirically unsubstantiated (Blasi, 1980; Walker & Hennig, 1997), what has become known as the 'judgement–action gap'. Contemporary moral psychology is univocal in recognising that a complete account of moral personhood requires looking beyond the single variable of moral cognition in search for the glue that might hold moral thought and action together in a causal way (Lapsley, 2006).

Independent of this shift in subject matter, Kohlberg's bold intellectual spirit remains undiminished as the field faces some of the same issues that cut through his work, albeit in new form. Kohlberg's (1971a) agenda for moral education was to argue for the fundamental importance of deliberative reasoning—as opposed to the more passive indoctrination of character traits espoused by competing models (Berkowitz & Schwartz, 2006). Kohlberg's philosophical agenda was to defeat ethical relativism by demonstrating that there are better (higher-stage) forms of moral thought and worse forms (Kohlberg, 1971b). As we will discuss in this article, these two issues—the role of deliberative reasoning and the battle against ethical relativism—remain of primary concern to the contemporary study of moral psychology. Kohlberg's success with these issues aside, his brilliance was in his ability to link substantive philosophical issues with relevant psychological data and in providing the empirical paradigm with which to do so.

But the nagging question 'Why be moral anyway?' kept resurfacing and Kohlberg (Kohlberg & Power, 1981; Kohlberg & Diessner, 1991) obviously struggled with the implication that his theory failed to encompass moral motivation. In response to this critical problem, Blasi (1983, 1984, 1993, 1995) re-introduced[2] notions of the moral self and identity to the study of moral psychology, thereby offering a more inclusive account of moral personhood. His model has since dominated discussions in this area (Hardy & Carlo, 2005). In no way did Blasi intend for his theory to oust Kohlberg's; rather, Blasi (2004) held that moral cognition remained a necessary component of moral functioning. Meanwhile, other models of moral personhood ignore or minimise the role of moral judgement (Hoffman, 2000; Eisenberg, 2005). To such models, the judgement–action gap is of minimal concern. Thus, the very contents of this review will be controversial. In brief, we argue (after Blasi, 2004) that any model of moral functioning that does not buttress the foundational importance of moral cognition entails a vacuous (empty-headed) conception of morality and lays prone to charges of ethical relativism. We bracket this issue, however, and continue with our investigation of moral personhood, premised on the critical retention of moral cognition and the resulting problem of the judgement–action gap.

Blasi's theory buttresses the existence, and causal importance to moral functioning, of something beyond judgement—something about the individual who forms the judgement. As the story goes, those with a well-developed moral personality are more likely to be motivated to carry out their moral judgements in the face of competing interests than are those bereft of such a moral personality. Thus, to Blasi, moral personhood—that is, the owner of the full scope of agentic moral

psychological processes—entails both sophisticated moral judgement and a well-developed moral personality. Blasi goes on to detail the specifics of what a moral personality means. Unlike Kohlberg, however, Blasi did not provide an empirical paradigm by which to investigate his conceptual model, leaving this task to others. Some research has since been done on certain aspects relevant to his model (viz. moral centrality) whereas other aspects (e.g. integrity) remain without a credible operationalisation. Our main contention is that, with respect to an empirical framework, the field remains in a pre-paradigmatic state. This is where we wish to take up the task.

The goal of this article is to present a selective and critical review of the 'state of the art' of aspects of moral personality that may bridge the judgement–action gap and to use what we now know to lay out critical future directions for research. By focusing particularly on those aspects of moral personality theory that are the most contentious, we intend to direct attention to the vulnerable points of extant theory and evidence. Our hope is that this article will instigate research that will move the field towards a new empirical paradigm of moral personhood. First, we lay out the complex problem of the judgement–action gap. Second, we review rival approaches to capturing the construct of moral centrality that have been posited to help span the judgement–action gap. Third, we consider the essential nature of the self and its implications for an understanding of moral functioning. Fourth, we address the critical problem of integrity for the enterprise of moral psychology and its relevance for bridging moral judgement and action. Finally, we step back and proffer some criteria by which we can evaluate competing models of moral personhood.

In sum, this article will honour Kohlberg's legacy, not by directly giving credence to the content of his work, but instead by examining the current state of moral psychology and by framing future agendas for the field in a way that will end up bringing us full circle to Kohlberg's vision.

The gappiness of moral life

The breach between judgement and action in moral life represents a fundamental conundrum for psychological theories of moral functioning. Before delving into how a bridge over this gap might be constructed, it will be helpful to explore the abutments on either side, beginning with moral behaviour and then turning to moral judgement.

How to define moral behaviour?

The danger associated with drawing ostensively amoral components (e.g. will, personal desires) into a model of moral functioning is that without careful delimiting the meaning of such components can become ambiguous. In a parallel vein, Kohlberg charged the 'bag of virtues' approach with the crime of ambiguity in that 'one person's "*integrity*" is another person's "*stubbornness*"' (Kohlberg & Mayer, 1972, p. 479, italics in original). Behaviour is prone to a similar ambiguity. What

should count as moral as opposed to non-moral behaviour? Perspective once again enters. Thoughtful outside observers, philosophers and psychologists (domain theorists notwithstanding) can produce nuanced but plausible illustrations of moral implications for just about any (ostensively non-moral) behaviour (e.g. wearing a tie, choosing a hobby). Thus, most behaviour could qualify as morally relevant from (what we call) a 'third-person perspective' in that behaviour almost always impacts the self and/or others non-neutrally. However, from the 'first-person perspective' of moral behaviour (whether or not the *actor* construes his/her action in moral terms) the same act may appear to be categorically non-moral; the actor may not be aware of any moral implications and may thus construe the behaviour in entirely non-moral (i.e. conventional, personal) terms. Is a first-person moral appraisal or judgement *necessarily* required for behaviour to 'count' as moral or is a third-person's philosophically viable argument sufficient? Opinions obviously diverge on this pivotal issue.

On the one hand, Blasi fashioned his Self Model as a bridge of the 'judgement–action gap', arguing that personality provides the motivation for the carrying through of specific moral thought to the specified action. Therefore, the action that is relevant to Blasi is that which the actor sees as having moral relevance (Blasi, 2005)—the first-person perspective. The benefit of such a stringent criterion is that it ensures that the behaviour in question is indeed morally relevant. However, there are a couple of inherent practical shortcomings to requiring a first-person moral judgement here, ones that may turn out to be unnecessary. Researchers trying to elicit specific moral judgements regarding actions (especially in instances of moral failure) will undoubtedly be swimming against the current of social desirability, cognitive dissonance and cognitive distortions (Bandura, 2002), each of which has the effect of invalidating the actor's reports. A second shortcoming with a first-person definition is that it conflates moral behaviour with sensitivity to the moral domain. An act would count as moral not only if the researcher succeeds in avoiding defence mechanisms but also if individuals' memory and moral sensitivity cues them to the moral thought that had preceded the act. To illustrate this problem, Walker (2004) reported that, upon prompting research participants for a recent moral dilemma from their personal experience, some were able to recall several from that day whereas other (morally insensitive) persons had to reach back many years, even decades, for a situation with moral overtones. For the latter type, apparently little or none of their behaviour, spanning significant portions of their life, would qualify for the moral domain under a first-person definition of behaviour; that seems fallacious.

One of the more pertinent contemporary examples of where first-person perspectives commonly fail to account for morally relevant behaviour is the issue of climate change. When considering travelling, most people (especially energy-profligate North Americans) likely do not see their choice of mode of travel (viz. driving, flying or travelling itself in the first place) as being a moral concern. Instead, they often focus predominantly on time efficiency and personal comfort. But Al Gore in his 2006 film, *An inconvenient truth* (David *et al.*, 2006), popularises the compelling case for how climate change (and its causes) are indeed moral issues,

whether or not the climate changer is aware of them. Human-induced climate change has far-reaching negative impacts on persons and ecosystems worldwide; ignorance of the connection between one's action choices and this global crisis does little to remove its moral gravity. Ignorance may be a reason, but it is by no means an excuse.

In contrast to the first-person approach where the explicit moral judgement of the participant is elicited, many researchers have measured moral behaviour without specifically assessing participants' moral judgement of their action—a move that may appear specious. Illustrative examples include bravery in rescuing others (Walker & Frimer, 2007), extraordinary care (Hart & Fegley, 1995; Matsuba & Walker, 2004; Walker & Frimer, 2007), social activism (Colby & Damon, 1992), honesty (Derryberry & Thoma, 2005), ecological behaviour (Kaiser & Wilson, 2000), community service (Hart *et al.*, 2006) and volunteerism (Aquino & Reed, 2002). By adopting some criterion of what constitutes morally laudable behaviour, these researchers implicitly endorsed an often-unstated, non-neutral, normative conception of the good (e.g. bravery is good). All non-neutral conceptions of the good favour some particular worldview over another; thus, endorsing one leaves the study's conceptualisation prone to any critique aimed at its associated conception of the good.

On first blush, this may seem avoidable. By relying on first-person judgements of moral behaviour, a researcher could argue to have remained neutral with respect to conceptions of the good in that no *telos* is (or appears to be) imposed upon the individual. In such a framework, for example, a moral personality that bridges conformity-oriented moral judgement to conformity-oriented action would have to be scored as being just as well-developed as a moral personality that bridges caring-oriented moral judgement to its corresponding action (the development of moral reasoning being equal). That is, superior moral functioning within a value-neutral theory cannot—by definition—give preference to any particular value (viz. conformity over caring or vice versa). What distinguishes 'good' moral functioning from 'bad' moral functioning must be some other (value-neutral) quality. Kohlberg resided squarely in this camp with his adherence to Rawls' (1971) and Kant's (1785/1964) neutralist theories in that he argued that the structure of moral reasoning was separable from content and, thus, value-neutral.

But Sher (1997) makes the compelling argument against the very notion of value-neutrality—Rawls' (1971) championing of autonomy and Kohlberg's (1981) heralding of sophistication in reasoning are both endorsements of non-neutral values from the start. Thus, after Sher, we argue that our theories of moral personhood must endorse not only goals for how individuals should (deontologically) reason right from wrong but also what they should (teleologically) value as being good from bad. The new challenge becomes the highly controversial task of articulating and defending an account of which values are preferable in sufficiently broad terms to avoid ethnocentrism. To be clear, our point is that this task is *categorically unavoidable* for any theory of moral psychology. Thus, we come full circle to argue that moral psychological theories should incorporate some

third-person (non-neutral) definition of what constitutes good moral behaviour. Just which non-neutral conception of the good is most appropriate remains important conceptual work for the field.

Whereas first- and third-person perspectives of moral behaviour are conceptually distinguishable, the two appear to overlap considerably in real life. Many definitions of moral behaviour (e.g. fairness) are sufficiently normative that most people share them (Rawls, 1971). Arguably, the reason why some people score higher than others on normative metrics of moral behaviour is a reflection of individual differences in the prioritisation of (competing) values and other personological factors; not that individuals typically disagree on whether the behaviour is good and right in the first place. An advantage of relying on third-person metrics of moral behaviour is that doing so avoids the inherent problems, previously noted, in eliciting valid moral judgements. Meanwhile, the drawback of relying on a third-person definition is that it introduces the possibility that the action may serve an ulterior motive. For example, Hart *et al.* (2006) noted that, whereas many students engage in community service for moral reasons, others do so merely for the sake of building a résumé for university admission.

Obviously what constitutes moral behaviour is critical in attempts to bridge the judgement–action gap, to explain moral functioning more fully and to articulate goals for development. First- and third-person perspectives may be tapping somewhat different phenomena—descriptive versus normative forms of moral behaviour. Independently, each is important to our understanding of moral functioning. The ways in which they converge and the ways in which they diverge are considerations to which researchers should focus more attention in the future.

What is the role of deliberative judgement?

A parallel issue concerns the opposing abutment of the judgement–action bridge, namely the nature of moral judgement. Some (e.g. Blasi, 1984) maintain the centrality of effortful and explicit reasoning as the origin of moral psychological processing. To Blasi, personality structures are meant to give life to these judgements and, thus, bring them forth to fruition. Blasi's moral exemplar is one who regularly gives pause to consider moral implications, reasons through problems in sophisticated ways, judges responsibility for action based on the significance to the self and has the prerequisite moral desires, willpower and integrity to carry judgement to action. The point to be taken from this illustration is the significant degree of effortful/active/explicit cognitive processing involved.

In contrast to this deliberative approach, a competing position—which we will label the intuitive approach—posits that much less explicit processing is normally operative in moral functioning. Haidt (2001), for example, argues that intuition (largely conditioned by innate moral emotions) is responsible for most moral judgements and that deliberative reasoning factors in primarily as a means of *ex post facto* justification and social persuasion. Irrespective of its validity, Haidt's claim is helpful in pointing out that many moral judgements are fast. Only rarely do

individuals engage in Kohlbergian-type reasoning in real-life circumstances; and even if some deliberative reasoning does occur, it becomes quicker with development to the point that it approaches automaticity. Lapsley and Narvaez (2004), although diverging from Haidt on several points, present a social-cognitive perspective that emphasises automaticity of moral judgements. Their theory goes something like this: exposure to, deliberation on and the pursuit of personal goals relevant to moral issues activate, elaborate and refine moral schemas. For those well along on a moral path, these schemas become chronically accessible and affectively charged to the point where they approach automaticity. Persons with a moral personality often automatically orient to the moral features of a context and reach a judgement without engaging in deliberation. Thus, the moral exemplar here is one who has been deeply socialised into a richly moral path, has consciously worked through a range of moral issues and has come to have highly elaborated and accessible moral schemas that collectively represent as moral expertise. In this state, the exemplar cues to the moral undertones in a given context and intuitively formulates a moral judgement (presumably, the same processes that Blasi implicates, such as integrity and agency, then factor in to bridge the judgement–action gap).

In deriving a coherent model of moral functioning that meaningfully connects judgement and action, the primary point of contention between the deliberative and the intuitive accounts lies in the nature of *operative* moral cognition. The issue can be stated as follows: is moral cognition necessarily reflective, explicit and slow or can it be reflexive, implicit and fast? On the one hand, the deliberative position may be the safer bet of the two in that it requires the individual to engage in, and in some sense to be aware of, his or her moral processing for the psychological activity to qualify as the moral thought abutment of the judgement–action bridge. The intuitive account, in contrast, requires a modicum of moral cognition but grants it permission to fly below the radar. Thus, the intuitive account partially removes the I-self or conscious from the story of moral functioning. On the other hand, the intuitive position seems better able to account for the role of implicit processing in formulating moral judgements as well as individual differences in moral sensitivity.

From a phenomenological and, thus, an empirical perspective, the task of distinguishing deliberative from intuitive judgements, *in vivo*, will be difficult. The heavy cognitive demands of making moral judgements and the short operative time envelope present as two obstacles to observation. Schrödinger's (1935) famous thought experiment involving a cat seems relevant here; the upshot of which is the notion that one cannot observe a phenomenon without interfering with the phenomenon itself as it would have existed independent of the observation. In the course of forming a moral judgement, self-observation will likely interfere with the process of judgement formation. That is, simply asking individuals in real time to state whether or not they formed a judgement intuitively or deliberatively will have the effect of contaminating the judgement itself. An enticing solution is to ask individuals after the fact, but this approach presupposes that individuals sufficiently encoded to memory not only the contents of their judgement, but also the process by

which they formed it—a presupposition that may be impervious to empirical scrutiny. Thus, it remains a significant challenge to establish a means of distinguishing between explicit and implicit judgements. Without such a demonstration, the historical bifurcation of these processes may break down leaving the distinction between them merely quantitative in nature.

The issues that arise out of these competing accounts of the definition of the judgement–action gap have important conceptual and practical implications. What role does rationality play in mature moral functioning? This is the very question on which Kohlberg took a hard-line stance. Does the development of a moral personality involve the shedding of deliberative moral judgement and the nurturing of the intuitive kind (a qualitative change in process)? Or does development entail increasingly sophisticated deliberation, applied more broadly and perhaps with greater efficiency (a quantitative change in speed)? These questions are not only of profound conceptual significance but are also relevant to practical applications in moral education and other interventions and present exciting directions for future research.

The judgement–action bridge

Having taken a stance on the nature of the judgement–action gap, the task then turns to its spanning. Divergence of opinion exists regarding the personological components (sensitivity, emotions, moral centrality, integrity, responsibility, agency) that a model takes as primary and which are secondary in the sense of being by-products of the primary components. Thus, the very subject matter of a review will itself inevitably be controversial. We have chosen to use Blasi's Self Model as the framework for our discussion as his is the most formulated account of moral functioning on the market and has the most intuitive appeal to us. The Self Model posits that three personological components intersect to bridge the judgement–action gap: moral centrality, integrity and responsibility. Of these components, moral centrality has received the most empirical attention and implicates concerns regarding the essential nature of the self. Meanwhile, the component of integrity presents as a critical problem to the field, one that has deep philosophical and practical implications. Integrity, as we will discuss it, also subsumes the notion of responsibility to a considerable extent. Thus, our focus will be on the two core components of the model, moral centrality and integrity.

Moral centrality

Blasi (1983, 1984, 1993, 1995) and Colby and Damon (1992, 1993) posited that the fusion of moral goals with personal ones is a hallmark of a moral personality and thus deny the common equating of morality with onerous duty, self-sacrifice and other things that seemingly run counter to human inclinations. With aims of capturing this abstract contention, researchers have operationalised the construct of *moral centrality* in multifarious ways. Broadly speaking, the various approaches can

be characterised either as being explicit and thus assess self-concept, or as being implicit and thus are more projective in nature.

Three empirical paradigms can be taken to illustrate the explicit approach. First, Aquino and Reed's (2002) self-report measure of moral identity has garnered some attention. After eliciting naturalistic conceptions of the characteristics of a moral person, they used list-reduction techniques to condense the list to nine defining traits of moral personhood: caring, compassionate, fair, friendly, generous, helpful, hardworking, honest and kind. They then had participants rate their degree of agreement with ten statements that addressed the self-relevance of the collection of traits. Summing across items, they found that individual differences in moral centrality were predictive of self-reported volunteerism and of a behavioural measure of generosity (food bank donation), albeit only weakly so.

Another attempt to capture explicit moral centrality has relied on values theory and has yielded mixed results. For example, Bardi and Schwartz (2003) had participants rate the self-relevance of 45 value items. These value items cluster into ten diverging value types: achievement, power, tradition, conformity, security, benevolence, universalism, self-direction, stimulation and hedonism. They also assessed self-reports and peer reports of behaviours relevant to each value type and found that the prosocial and sustainability-oriented values of universalism (tolerance and protection for all people and nature) and benevolence (concern for the welfare of close others) predicted associated self-reported behaviours moderately, but peer-reported behaviours only weakly.

The third empirical paradigm for assessing explicit moral centrality has relied on personality trait theory and has again indicated meagre findings. For example, Matsuba and Walker (2004) found that only one factor (agreeableness) of the Five-Factor Model of personality distinguished caring exemplars from a matched comparison group; and Walker and Frimer (2007) found that neither the dominance nor the nurturance dimension of the interpersonal circumplex distinguished their moral exemplars from comparison participants (although, of their moral exemplars, the caring type was more nurturant than the brave type).

For the explicit moral-centrality approach to succeed it will have to overcome three problems. First, a general trend in the empirical literature is that the stronger correlations between self-endorsed traits and moral behaviour are based on less robust measures of moral behaviour. Bardi and Schwartz (2003) found the strongest correlations when they relied on self-reports of behaviour ($r \approx .45$); but when they used peer-reported behaviour, the correlations dropped to $r \approx .20$. And when Walker and Frimer (2007) used moral exemplarity—a far less susceptible measure of moral behaviour—the relationship between dispositional traits and behaviour more or less evaporated. This raises the question regarding the nature of the construct tapped by self-endorsed traits. Trait theory presumes that the way that individuals endorse traits reflects the degree to which these traits are accessible and operative in day-to-day life. However, Smith *et al.* (2007) found this not to be the case: rating a value as important does not necessarily reflect its chronic accessibility (as assessed by spontaneous production). McClelland *et al.* (1989) explain the weak intercorrelation

between divergent behavioural correlates of self-report and projective measures by arguing that the two methods tap different motivational systems. Given that recognition and production are not tapping the same thing, it may be either that both cannot be accessing the construct of moral centrality or that the singular notion of moral centrality may be inadequate. The question for future research thus becomes: what are the motivational nature and behavioural outcomes of each of the two methods/systems—endorsement (explicit) versus production (implicit)—and what, if any, is the relationship between them?

The second problem is to overcome Kohlberg's objection to the 'bag of virtues' approach. In Kohlberg's view (Kohlberg & Mayer, 1972), the task of constructing a list of primary virtues involves a kind of dipping one's hand blindly into the bag and arbitrarily pulling out virtues until one has a list that seems to capture the breadth of moral character. Thus, one challenge is that such a list is arbitrary. Several research teams (Walker & Pitts, 1998; Lapsley & Lasky, 2001; Aquino & Reed, 2002; Smith *et al.*, 2007) have performed such a dipping by assessing naturalistic conceptions of moral maturity and then using various data reduction techniques. This war of attrition yields victors that capture the prototypic epicentre of moral personhood. If Kohlberg were right, little cross-dipping consistency should be present and, indeed, the extant evidence indicates that this may be the case (both Lapsley & Lasky, 2001, and Smith *et al.*, 2007, compared their lists of virtues with that of Walker & Pitts, 1998, and reported only partial overlap).

A third problem for the explicit moral-centrality approach is that such inventories of traits gloss over important inter-virtue complexities and how various virtues are balanced. In the course of participating in research, rating oneself as high on trait items comes with no particular cost; one is free to rate oneself uniformly high on all. On the other hand, such virtues come with associated costs (in terms of other virtues) in real life. For example, one virtue may conflict with the demands of another in concrete situations (e.g. being honest may be at odds with being compassionate). Also, such lists of desirable virtues may entail a set of characteristics that no single person could possibly exude; such an agglomeration within one person would be incoherent. Illustrating the complexity of virtues is the observation that they can take on maladaptive forms, as Hennig and Walker (2008) have demonstrated with some types of care that have apparently gone awry. None of the self-endorsed trait approaches to date have addressed these critical nuances regarding virtues.

We have identified three significant problems confronting the explicit moral-centrality approach: (1) the meaning and predictive validity of self-endorsed traits, (2) the arbitrariness and cross-list variability of collections of virtues and (3) the inter-dependencies among virtues and the complexities of their expression. Each of these three issues is amenable to empirical investigation that will clarify the usefulness (or lack thereof) of the virtue approach to moral functioning.

The competing approach to capturing moral centrality filters the individual's *implicit* sense of self for moral themes. Hart and Fegley (1995) introduced two such methods to the study of moral personality: self-concept as content and self-concept

as hierarchical set relations. To tap self-concept as content Hart and Fegley asked adolescent participants to produce descriptions of the self and these responses were then matched to one of 28 categories. They found that the content of the self-descriptions of caring adolescent exemplars included more references to moral and caring personality traits and goals than did the self-descriptions of comparison adolescents. Notice that this approach escapes two of the critiques levelled against the explicit trait approach. By tapping production, self-concept as content may avoid the problem of arbitrariness associated with endorsement of a preset menu of virtues and by coding statements ipsatively, moral self-statements come at some price, partially addressing the problem of inter-virtue complexities. By coding ipsatively, we mean that, in this procedure, individuals can only self-attribute prosocial traits (e.g. 'caring') in some sense by not claiming other traits (e.g. 'competitive'). However, the approach relies on 28 categories, constructed seemingly *ad hoc*; future work could focus on establishing the validity of the category structure itself.

The self-concept as content approach also fails to clearly map out the relationship between individuals' moral sensibilities and their actual self (the conceptual hallmark of moral centrality). Self-concept as hierarchical set relations succeeds in this regard. Hart and Fegley's (1995) methodology here entails increased methodological demands, requiring multiple sessions with participants and complex analyses. In the first session, participants are asked to generate descriptors of various representations (or sets) of the self, which included (among others) the actual self, ideal self and ought self. In the second session, participants rate the applicability of each descriptor to each of the other representations of self in an 'identity matrix', a demanding task which typically requires thousands of ratings. Analyses determine which sets are contained within other sets. Hart and Fegley found that the ideal self was contained within the actual self more so for the caring adolescent exemplars than for comparisons, suggesting that the two are fused in a way previously theorised (and Derryberry & Thoma, 2005, replicated this finding). Due care is required in interpreting this result, however: the ideal self is a *first-person* rendition of the moral self and thus is free to wander away from philosophically valid notions of morality. This promising approach could be improved by taking into consideration some third-person coding of the ideal self, to distinguish ideal selves that are less than exemplary (e.g. competitive, powerful, conformist) from those that better adhere to defensible moral definitions (e.g. just, caring or brave).

Another approach to capturing implicit moral centrality comes from our own research, currently in progress. To avoid transparency of the task, we interviewed university students about various aspects of their lives (such as relationships, activities, goals, personality). We then coded their responses thematically for implicit endorsement of the ten values of the Schwartz value typology—a category structure that is universally applicable, having been validated in 44 countries (Schwartz, 1994). Among the implicit values predicting a battery of measures tapping various manifestations of moral functioning (including ecologically sustainable behaviour, prosocial action, honesty and a paucity of materialistic values) were (1) universalism (a generalised concern for others and for the environment, $\beta = .39$); (2) power (desire

for material wealth and social status, $\beta = -.33$); and, perhaps most interestingly, (3) an interaction between power and benevolence (concern for known others, $\beta = .39$). We interpret the latter theme as being indicative of the way that agentic or self-advancing themes, on the one hand, fuse with communal or other-advancing themes, on the other hand, in a well-developed moral personality. Another advantage of this approach is that moral centrality is tapped from life-narrative, which is arguably the mode of self-understanding that is most operative in moral functioning (McAdams, 1993; Walker & Frimer, 2007).

In summary, at least six empirical methods for tapping moral centrality exist, some explicit and others implicit. A number of concerns regarding these approaches have been identified and discussed. In that the two types may not be tapping the same construct, they may be assessing different facets of moral centrality. Future research should aim at clarifying the construct(s) and predictive validity of each approach with the intent of resolving the issue of which should become the 'industry standard' in the construction of the bridge over the judgement–action gap.

An essential self or a dog's breakfast?

The previous section illustrates the extensive and at times ingenious ways in which researchers have tapped moral centrality. Tacit in this line of research is the premise that the moral self is forwards operative in causing moral action and not simply observing behaviour and documenting an identity to match—an assertion that turns out to be the focus of heated contest. Moral psychology introduced the topic of the self into the discussion on moral functioning (Noam & Wren, 1993); in doing so, it wittingly or unwittingly joined a debate that cuts to the core of personhood. Namely, *what is the nature of the self*? If the topic sounds less like under-pressed tofu and more like the cut of meat into which all scholars dream of sinking their teeth, consider that a moral psychology advancing the notion of a moral self cannot afford to remain aloof from or neutral on this topic. Before building the case for the construct of moral personhood, let us consider the current against which the enterprise will have to swim.

The dimension of the self of primary concern is that of unity versus disunity, or what Blasi (2005) refers to as the *integrity of identity*. Modernity's self is unified, internally consistent and has an essence that exhibits agency across contexts. Commonly speaking, the will and consciousness are instances of the phenomenon (or epiphenomenon, as the case may turn out to be) of the essential self. Its elegance and parsimony aside, the problem with this account is that it seems to be full of empirical holes. The seminal investigations of Hartshorne and his colleagues (1928, 1929, 1930) found a disappointing result for the notion of the essential self: people's moral actions in different contexts showed little in the way of consistency along the lines of personality traits. The contemporary trouble for the essential self is manifest as the oft-expressed disillusionment with the trait approach within the field of personality psychology (Mischel, 1984; cf. McAdams & Pals, 2006).

In response to the horrors of World War II, a movement of scholars emerged who were characterised by pervasive doubt and deconstruction. Unable to explain how the centuries of progression within the 'Enlightenment Project' could crescendo into the Third Reich, Auschwitz and Hiroshima, postmodernism abandoned the very notion of progress, the good and truth:

> Post-modernism actively eschews all ontological, epistemological and ethical absolutes; disallows the existence of anything Platonic; is antiessentialistic; rejects all notions of human nature; and consigns to the dustbin all ideas of deep structure or the possibility of causally generative mechanisms. Most of all it decisively disallows all prospects of enduring sameness and the possibility of directed growth. As such, post-modernism amounts to a eulogy read out in memorial to all suspect illusions about intrinsic value or disinterested truth… (Chandler, 1997, p. 7)

Unsurprisingly, when aimed at psychology, postmodernism's characteristic rejection has been interpreted as relegating social psychology to an historical discipline (Gergen, 1973) and rejecting the very possibility of postmodern developmental psychology (Chandler, 1997) or postmodern psychology whatsoever (Kvale, 1990). Kohlberg (1971b) made a philosophical goal of defeating such moral relativism or nihilism and of doing so using empirical means. Fascinatingly, the new paradigm in moral psychology—as challenging as it is to Kohlberg's thinking—is just as distasteful to Kohlberg's adversaries as his model was. Like the Kohlbergian era, the post-Kohlbergian one is, and should be characterised by, deep, divisive, philosophical rifts. Behold the old (and renewed) adversary!

In the weakest reading,[3] the postmodern claim against the self denies any metaphysical essence transcending the physical (e.g. an immortal soul, disembodied mind). Such a claim has religious or spiritual significance, but resides squarely beyond the realm of scientific inquiry and is independent from the question of the moral self (and thus presents little interest to the present discussion). But in a stronger, more common and louder reading, postmodernists cast aside talk of persons being governed by some unified inner agent as reflecting an 'ethnocentric Western view of personhood' (Hermans *et al.*, 1992, p. 23), deny the self the ontological status of a natural object and, instead, advance the notion of a narrative self where 'talk about the self cannot be separated from what the self is' (Danziger, 1997, p. 156). Thus, many postmodernists took up the typical deconstructionist stance and took aim at the whole, coherent self of modernity with the intent of shattering it into the socially constructed, context-dependent, fleeting and disunified 'dog's breakfast' of postmodernity (Holland, 1997).[4]

In that the postmodern retraction did not go so far as to deny the very existence of the self in the way that the behaviourists did, social constructionists (e.g. Shweder, 1991; Dennett, 1992; Hermans *et al.*, 1992) faced the problem of delineating some self without an intrapsychic essence. A typical approach was to describe the self as socially constructed 'through the mediation of powerful discourses and their artifacts—tax forms, census categories, curriculum vitae and the like' (Holland, 1997, p. 170). In the 'extreme ephemeralist' position:

> daily life, especially in the postmodern era, is a movement from self to self—at one moment a self resists or is successfully positioned as lacking a soul; the next, it is an emotional woman; the next, an organ donor; two minutes later, a high-risk driver; and then, later in the day, 'a bad neighbor'. (Holland, 1997, p. 170)

Most relevant to our discussion on moral personhood is perhaps the signature feature of the social constructionist's self—a glaring *absence*. The social constructionist's self removes more than a *metaphysical* being; the postmodern self is without a *psychological* process[5] that actively integrates the contents of its own narrative or that provides a psychological basis for diachronic unity. This is not to say that it lacks James's (1890) self-as-knower (or I-self); it is to say, however, that the I-self takes, at most, the role of a detached observer. Thus, the postmodern self may watch itself engage in prosocial action (e.g. consent to organ donation) at one moment and act irresponsibly (e.g. drive dangerously, to continue with Holland's illustration) at another and, yet, the observing self is entirely unperturbed by any inherent incongruity therein. After all, why should there be any congruity between such behaviours when the social pressures of the different contexts bear no resemblance to one another (and the very concept of morality and the term 'should' are simply modernist conspiracies in the first place)?

Thus, an alternate version of a moral self (quite unlike Blasi's) might appear to not require notions of self-unity or integrity. But, on closer inspection, can moral psychology afford to concede a shattered self? In short, we argue that the answer is 'no'. Accountability, as we understand it, rests on the notion that persons possess continuity through time, 'a necessary, constitutive and, therefore, universal design feature of selves of any description' (Chandler & Sokol, 2003, p. 206). Moral psychology places the locus of accountability for an individual's actions on the self—a *psychological* unit. In allowing the self to be at the whim of ever-changing social forces, the social constructionist can retain the notion of accountability but must assign it to some other, *non-psychological* aspect of the individual—one that transcends time and context (e.g. his/her biology).[6] Thus, any viable *psychological* theory of accountability (and thus of morality) must provide an argument for intrapsychic persisting sameness.

Behavioural and identity integrity

What does moral psychology need to demonstrate to defend the causal importance of the moral self? On first blush, the demonstration of complete self-unification may seem necessary. But we argue that, whereas a completely unified self is a seemingly attractive idea, it rests at odds with the unmistakable reality that having complexities, internal inconsistencies and imperfect insight into one's own functioning are unavoidable aspects of being a person. We contend that advancing the developmental goal of *complete* unification is unnecessary. The moral self can have the quality of accountability with imperfect unification so long as there exists an agentic, unifying process whose job it is to cross-reference the contents of the self (beliefs, values, behaviour), *attempt* to mend inconsistency and have some (incomplete) success. The social constructionist charge denies the existence of any such process,

despite the empirical evidence suggesting its existence. Faced with evidence of the incongruity of one's words and actions, individuals experience cognitive dissonance (Festinger, 1957; Festinger *et al.*, 2007), an unpleasant and motivating experience that can be neutralised by (1) adopting a more negative self-concept, (2) engaging in cognitive distortion or (3) changing future behaviour. We argue that cognitive dissonance is indicative of the existence of the contested integrity process.

Moral psychology is interested in more than just establishing the existence of a unifying process—the field aims to demonstrate that developmental/individual differences in this phenomenon help to bridge the judgement–action gap. But strikingly little can be said about empirical paradigms that capture the state of an individual's integrity. Blasi (2005) has discussed the importance of two types of integrity: identity integrity and behavioural integrity. Whereas the former—identity integrity—is the one more applicable to the previous discussion, we contend that it also holds both conceptual and empirical advantages over the latter. Before making these points, some discussion of the latter type is in order.

Behavioural integrity references the degree of correspondence between intra-psychic action (thought, intention) and behaviour. Behavioural integrity alone does not necessarily imply the uniformly positive connotation of far-reaching typical folk conceptions of integrity; rather, under this technical definition, behavioural integrity requires *only* correspondence between one's thought and one's actions, independent of how morally defensible the thought is in the first place.[7] The field still awaits a measure that taps individual differences in behavioural integrity. Such a measure could use one of two approaches: it could either tap the *actual* (in)congruence between thought and action or it could tap individuals' *perception* of the (in)congruence. If behavioural integrity is intended as a variable to help bridge the judgement–action gap, the first option simply will not do. To assert that (actual) behavioural integrity helps span the judgement–action gap is a tautology—the degree of congruence between judgement and action (actual behavioural integrity) is, *by definition*, the very outcome construct that it means to predict in the first place—namely the bridging of the judgement–action gap.

In contrast, the perception of congruence is indeed of theoretical interest but an armoury of defence mechanisms (e.g. self-serving reconstrual of actions) will ultimately hinder researchers from accurately measuring perceived (in)congruence. Future research in this area will need to develop clever ways of bypassing these defence mechanisms. One promising starting point for a paradigm comes from the social-cognitive literature. Cervone's (2004) knowledge-and-appraisal personality architecture is an empirical paradigm that relies on the premise that individuals appraise their traits in light of their beliefs about given situations. If a trait is deemed relevant to a particular situation, then, and only then, does it functionally come into play. To extend Cervone's paradigm for the discussion at hand, we propose that the less individuals connect their (moral) traits to behaviours, the less behavioural integrity they have. Such an approach within moral psychology remains to be developed. Nisan's (1995) model of moral balance may provide a second, and very different, launching point for capturing behavioural integrity. The idea behind

Nisan's paradigm is that people carry with them something of a moral ledger, wherein they record morally upright actions in the credits column and transgressions in the debits. So long as the credits less the debits remain above some cut-off, people can live with themselves. One possibility is that, in this paradigm, behavioural integrity is manifest as individual differences in the cut-off or threshold that the individual finds acceptable.

We contend that Blasi's first type, identity integrity, is of primary philosophic concern and entails intriguing conceptual nuance. Identity integrity connotes the degree to which an identity is internally consistent and unified. Philosophically, Baumeister (1998) argued that the basic notion of the self necessarily entails a study of the self's unity. Whereas the notion of multiplicity is a useful metaphor, he argues, without this notion of unity we have lost our subject matter. Conceptually, the study of identity integrity leads us to consider the subject's account of (in)congruencies among elements of the self. Owing to their highly self-central and evaluative nature, these experiences of (in)congruity are highly motivating (Steele, 1988) and thus provide a launching point for the study of moral motivation. One approach is to borrow from the social psychology literature, specifically the Meaning Maintenance Model (Heine *et al.*, 2006), which posits a universalised account of human motivation. This model proposes that persons universally form mental representations or schemas (i.e. expected relations among possibly disparate entities). These expected relations encapsulate a sense of meaning for an individual; in the event of a challenge to a representation, individuals engage in 'fluid compensation', whereby they defend or enhance some other representation to restore a net sense of meaning. Thus, the model captures a kind of investment that persons have in expectations about the self, about the outside world and about the relationship between the self and the outside world. Identity integrity may be a special case of these expected relations (e.g. the expected relation of self-unity). Of particular relevance to the present discussion is how the Meaning Maintenance Model can be applied to representations of the self. How might fluid compensation provide an *in vivo* mechanism for moral motivation?

A second point of conceptual interest has to do with the presumption implicit in Blasi's (1983) and Colby and Damon's (1992) thinking that, namely, when it comes to moral functioning, the more integration the better. Nuanced reasons suggest reconsideration of that view: is an ideal moral identity necessarily a fully integrated one? Given that 'messing up' in action, even with the best of intentions, is a simple fact of life, a fully agentic and integrated self would necessarily reach a troubling dilemma. Any claim that a non-deliberate part of the self caused the action (cf. the self is purely agentic) or that a part of the self that is at odds with the 'good' part caused it (cf. the self is entirely integrated) is out. The individual faces the only remaining option: that the immoral action was the direct result of the integrated and agentic self, one whose goal could never have been purely good (e.g. 'getting ahead in life, by any means'). By this account, the maintenance of perfect integrity, a sense of agency and a perfectly moral core represents a meta-stable state; the slightest perturbation in action will be the state's undoing. Thus, the moral personality

appears to entail a paradox: whereas it operates so as to sustain and even increase moral centrality, integrity and agency, the attainment of the three tasks appears to be a practical impossibility.

Illustrating this problem, Proulx and Chandler (2007) prompted young adults to defend their psychological self-unity after disclosing both their good and bad actions, behaviours that indicate contradictory desires and intentions. Few students understood themselves in an utterly disjointed way; indeed, most proffered some self-unity warranting strategy (e.g. a self with multiple desires that are reactivated by changing circumstances or a more conscious and deliberate self that occasionally gets disrupted by otherwise repressed desires). But the strategy of most interest here involved the claim of a singular and active self with all behaviours aimed toward attaining a common goal—that is, both self-unified and agentic. These unified and agentic persons were cornered into formulating some singular motive that could justify both their good and bad actions (e.g. 'I am always trying to get a leg up on my neighbour'), undermining the notion of a moral centre. Thus, future theory and research should pay careful consideration to the bright as well as the shadow side of these adaptive constructs, especially when they interact.

This review of empirical efforts to capture identity integrity serves to illustrate a further advantage of this type of integrity over the behavioural kind. That is, the notion of self-unity as a span across the judgement–action gap lends to the development of an independent construct. Behavioural integrity runs tautological risks (as discussed earlier), whereas the perception of coherence among, and psychological investment in, facets of the self is a promising construct that is independent of both judgement and action.

Returning full circle to Kohlberg's dispute with ethical relativism—a position that has now morphed into postmodernism—the quarrel is manifest today on the topic of integrity. What evidence would count to defeat the postmodern assertion that the self is nothing more than a dog's breakfast of attributes or narrative? One possibility is that the demonstration of *any* self-unity (significantly different from complete disunity) should suffice, especially if such a demonstration could be made in a non-modern culture (one where individuals have not been subjected to the modernist conspiracy). Cognitive dissonance has indeed been demonstrated cross-culturally (Hoshino-Browne *et al.*, 2005). This topic presents a valuable direction for future conceptual work, one that will inform and be informed by work of the empirical kind.

Moral personhood—towards a new paradigm

In this article, we have discussed the judgement–action gap as well as the two core components of Blasi's Self Model. We have paid little attention to other personological aspects of moral functioning, such as moral sensitivity, responsibility and emotions. In the coming years, we anticipate and hope that a number of (competing) models will vie for currency in the field. Examples of viable candidates include Rest's (1984) Four Component Model and Hart's (2005) model of moral

identity formation. How then do we go about making sense of competing accounts of moral functioning in this pre-paradigmatic era?

We propose the following four criteria: (1) person-based,[8] (2) comprehensive, (3) parsimonious and (4) predictive. First, a model's components must remain within the boundaries of the construct of moral functioning. By definition, moral functioning implies a strictly *personological* (as opposed to contextual) basis for moral action. As an illustration, two components of Hart's model (social influence and opportunity) remain outside the construct of moral functioning, whereas the other three components (personality, moral cognition and self) remain inside. This is not to diminish or ignore the powerful ways that culture and context shape personhood and behaviour. Rather, the notion that there exists *any* person-based foundation for morality is controversial enough. Kohlberg (like Piaget, 1932/1977) aimed to refute the behaviourist notion that learning occurs through passive infusion in social contexts. The enterprise of moral psychology rests upon the premise that there exists some personological explanation for moral functioning—one that is non-reducible to contextual determinants. Hence, the onus is on the field to account for some moral functioning, using person-based variables that are irreducible to contextual ones.

The second criterion by which models of moral personhood should be measured is *comprehensiveness* in accounting for moral functioning. All of the (primary) variables that are necessary for an account of moral functioning must be included in a model. Kohlberg claimed that his model satisfied this criterion. Even though his model captured only moral cognition, he posited that other facets of moral functioning (e.g. emotions, behaviour) are either merely the offshoots of moral thought or relatively inconsequential aspects of the domain. Kohlberg's bold move fell flat both empirically and conceptually (Blasi, 1980; Campbell & Christopher, 1996), opening the floodgates to other variables. New research programmes are studying personality, identity, emotions and so forth, advocating the non-redundant (to moral thought) role of these variables. For example, Walker and Frimer (2007) found that, although stage of moral reasoning partially distinguished caring moral exemplars from a matched comparison group, personality variables predicted considerable unique variance beyond reasoning alone. Thus, the post-Kohlbergian era is one of construct pluralism, wherein variables that Kohlberg saw as either categorically amoral (e.g. virtue) or the simple offshoot of moral cognition (e.g. emotion) are receiving vigorous exploration.

The expansive move past cognition alone brings with it greater explanatory power of other aspects of moral functioning and the promise of greater predictive power for behaviour. However, there are oft-neglected reasons to be cautious about what is looking like unfettered pluralism. Given that the primary components of a model stay on side of the boundaries of personhood and capture the extent of moral functioning, the third evaluative criterion is *parsimony*. Occam's Razor is applicable here: all else being equal, the best model is the one with the fewest primary variables. Theorists and researchers should be explicit regarding which variables they see as being primary (viz. independent and causal) and which variables they see as

secondary (viz. a direct by-product of primary variable functioning and perhaps the mechanism through which primary variables cause behaviour). Kohlberg held reason and reason alone to be the sole primary variable of the moral realm. His strong stance on the central aspect of moral functioning is praiseworthy in that the single-variable solution to moral functioning had clear heuristic value, optimally suited for applications such as education. In contrast, consider the indigestion for educators that a (competing) multivariate solution would impart. An example of the latter type is Hart's (2005) model of moral identity formation. His model posits five constructs (personality, social influence, moral cognition, self and opportunity) that cause moral behaviour. Each of these constructs is reflected by a number of factors, yielding a total of 13 factors that are held to govern and be governed by moral behaviour. Whereas the model has enjoyed some success in predicting community service (Hart *et al.*, 2006), it has achieved that goal at the expense of concision and heuristic clarity. Does some sub-13 factor model sufficiently capture the breadth of moral functioning with equal predictive power?

Researchers should now feel the tension between adopting more variables with the goal of comprehensively explaining moral functioning on the one hand, and of shaving off variables in order to preserve parsimony on the other hand. Whereas these criteria would appear to negate one another, this is not necessarily the case. A model can hold some set of variables as primary or foundational and demote other variables to secondary, consequential status. To achieve this, however, both conceptual and empirical arguments are required. What is the fewest number of variables that are sufficient and necessary to capturing the foundational and functional core of moral personhood? By most accounts, the most precise answer available these days would seem to be 'more than one'.

Once the primary variables of a model are defined and defended, the task then turns to fleshing out the way in which these variables interact in order to produce different sorts of persons and different sorts of outcomes (in terms of secondary variables). For example, presuming that emotion is not a primary variable in some model (a contentious presumption, cf. Hoffman, 2000; Haidt, 2001), how do the primary variables interact to produce moral emotions and does knowing something about emotional disposition add anything to the predictive power of the model? If emotions do add predictive power, then emotions could not count as a secondary variable. A second example pertaining to the interaction of primary variables was discussed earlier (recall the meta-stability of agency, moral centrality and integrity).

The fourth and final criterion is *prediction*. One of the motivating factors for moving past Kohlberg's model was its lack of predictive validity, typically accounting for about 10% of the variability in moral behaviour (see Blasi's, 1980, review and Buchanan's, 1992, meta-analysis). Clearly, predicting more than 10% of the variability is required, but just how much is enough? To account for all moral behaviour leaves little for powerful contextual factors. Then again, does exemplary moral functioning have the power to channel all situational pressures towards moral ends, thus making near-perfect prediction the goal of a personological account of moral functioning? For the time being, the most specific conclusion that we can

draw is that a model ought to predict significantly more than 10% of the variability in moral behaviour.

Concluding thoughts

Moral psychology is between paradigms. In the coming years, we anticipate the emergence of a new empirical paradigm of moral personhood, one that will reawaken many of the deep conceptual issues that Kohlberg first raised to the field. The goal of this article was to point out some critical issues that require attention, pertaining to the nature of the judgement–action gap, including the measurement of moral centrality, the nature of the self and the critical problem of integrity. Finally, we laid out some guidelines by which the field ought to evaluate competing models of moral functioning as we move towards the next paradigm. While the content of the new paradigm will differ markedly from that of Kohlberg, we contend that the spirit of his enterprise will be manifest with vigour redoubled.

Notes

1. Indeed, Kohlberg also came to recognise the limitations of the moral rationalistic approach. 'The second direction in which our study of moral responsibility and moral action carries us is to the resurrection of the notion of moral character, conceived as the development of the ego or the ''moral self.'' The moral self is the first-person side of the organization or unity of moral behavior postulated by the notion of character. It is the sense of the self's ''integrity'' or ''identity'' that becomes at stake in moral action' (Kohlberg & Diessner, 1991, p. 231). As Kohlberg died in 1987, he never pursued such notions to any extent; but it seems reasonable to assume that he would be sympathetic to the paradigm shift.
2. Western civilization has had talk of the self for millennia but it has not formed part of the conversations within moral psychology until recently.
3. Presenting considerable challenge to this (and any) discussion on what defines the postmodern self is the inherent haziness of what 'postmodern thought' advances in the first place (for a thorough discussion, see Chandler, 1997).
4. Interestingly, some domain theorists (e.g. Nucci, 2004) have implicitly endorsed the postmodernist parade in emphasising the necessarily context-dependent nature of the self.
5. To be abundantly clear, the existence of this process is *entirely* independent of the existence of any kind of metaphysical, disembodied mind or immortal soul. In this essay, we mean to argue for the former and bracket the latter for the scrutiny of philosophers, metaphysicists and spiritual leaders.
6. This is not to say that a narrative-based notion of the self is incompatible with moral psychology; the key consideration is whether or not the self—be it essential or narrative—has some unifying functionality, the business of which is the examination of the various components of the self (e.g. beliefs, values, actions) and the corrective task of working towards internal consistency. We argue that such a feature is possible in both an essential and a narrative self, but not in an *exclusively* socially constructed one.
7. Thus, a Nazi SS officer could have behavioural integrity under this strict definition. Like Blasi, we see at least some aspects of moral personality as being morally neutral while being morally motivating (e.g. a moral amplifier). Buttressing a morally neutral construct is not a problem for the model as retaining the foundational importance of moral reasoning provides the defence against ethical relativism.

8. A model being entirely person-based is as much a classification as it is a criterion. Some extant self-proclaimed models of moral personality confuse this matter; thus, for the pragmatic sake of covering all cases, we include being person-based as a criterion.

References

Aquino, K. & Reed, A., II (2002) The self-importance of moral identity, *Journal of Personality and Social Psychology*, 83(6), 1423–1440.

Bandura, A. (2002) Selective moral disengagement in the exercise of moral agency, *Journal of Moral Education*, 31(2), 101–119.

Bardi, A. & Schwartz, S. H. (2003) Values and behavior: strength and structure of relations, *Personality and Social Psychology Bulletin*, 29(10), 1207–1220.

Baumeister, R. F. (1998) The self, in: D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds) *Handbook of social psychology* (vol. 1, 4th edn.) (New York, Oxford University Press), 680–740.

Berkowitz, M. W. & Schwartz, M. E. (2006) Character education, in: G. G. Bear & K. M. Minke (Eds) *Children's needs III: development, prevention, and intervention* (Washington, DC, National Association of School Psychologists), 15–27.

Blasi, A. (1980) Bridging moral cognition and moral action: a critical review of the literature, *Psychological Bulletin*, 88(1), 1–45.

Blasi, A. (1983) Moral cognition and moral action: a theoretical perspective, *Developmental Review*, 3(2), 178–210.

Blasi, A. (1984) Moral identity: its role in moral functioning, in: W. M. Kurtines & J. L. Gewirtz (Eds) *Morality, moral behavior and moral development* (New York, Wiley), 128–139.

Blasi, A. (1993) The development of identity: some implications for moral functioning, in: G. G. Noam & T. E. Wren (Eds) *The moral self* (Cambridge, MA, MIT Press), 99–122.

Blasi, A. (1995) Moral understanding and the moral personality: the process of moral integration, in: W. M. Kurtines & J. L. Gewirtz (Eds) *Moral development: an introduction* (Boston, Allyn and Bacon), 229–253.

Blasi, A. (2004) Moral functioning: moral understanding and personality, in: D. K. Lapsley & D. Narvaez (Eds) *Moral development, self, and identity* (Mahwah, NJ, Erlbaum), 335–347.

Blasi, A. (2005) Moral character: a psychological approach, in: D. K. Lapsley & F. C. Power (Eds) *Character psychology and character education* (Notre Dame, IN, University of Notre Dame Press), 67–100.

Buchanan, T. (1992) Why is the literature examining the moral cognition–moral action relationship inconsistent? A meta-analytic investigation of five moderating variables, paper presented at the meeting of the *Association for Moral Education*, Toronto, November.

Campbell, R. L. & Christopher, J. C. (1996) Moral development theory: a critique of its Kantian presuppositions, *Developmental Review*, 16(1), 1–47.

Cervone, D. (2004) The architecture of personality, *Psychological Review*, 111(1), 183–204.

Chandler, M. J. (1997) Stumping for progress in a post-modern world, in: E. Amsel & K. A. Renninger (Eds) *Change and development: issues of theory, method and application* (Mahwah, NJ, Erlbaum), 1–26.

Chandler, M. J. & Sokol, B. W. (2003) Level this, level that: the place of culture in the construction of the self, in: C. Raeff & J. B. Benson (Eds) *Social and cognitive development in the context of individual, social, and cultural processes* (New York, Routledge), 191–216.

Colby, A. & Damon, W. (1992) *Some do care: contemporary lives of moral commitment* (New York, Free Press).

Colby, A. & Damon, W. (1993) The uniting of self and morality in the development of extraordinary moral commitment, in: G. G. Noam & T. E. Wren (Eds) *The moral self* (Cambridge, MA, MIT Press), 149–174.

David, L., Bender, L., Burns, S. Z. (Producers), & Guggenheim, D. (Director). (2006) *An inconvenient truth* [motion picture featuring Al Gore] (United States, Paramount Classics and Participant Productions).

Danziger, K. (1997) The historical formation of selves, in: R. D. Ashmore & L. J. Jussim (Eds) *Self and identity: fundamental issues* (New York, Oxford University Press), 137–159.

Dennett, D. C. (1992) The self as a center of narrative gravity, in: F. S. Kessel, P. M. Cole & D. L. Johnson (Eds) *Self and consciousness: multiple perspectives* (Hillsdale, NJ, Erlbaum), 103–115.

Derryberry, W. P. & Thoma, S. J. (2005) Moral judgment, self-understanding, and moral actions: the role of multiple constructs, *Merrill-Palmer Quarterly*, 51(1), 67–92.

Eisenberg, N. (2005) The development of empathy-related responding, in: G. Carlo & C. P. Edwards (Eds) *Nebraska Symposium on Motivation. Volume 51: moral motivation through the life span* (Lincoln, University of Nebraska Press), 73–117.

Festinger, L. (1957) *A theory of cognitive dissonance* (Evanston, IL, Row, Peterson).

Festinger, L., Carlsmith, J. M. & Bem, D. J. (2007) Issue 4: does cognitive dissonance explain why behavior can change attitudes?, in: J. A. Nier (Ed.) *Taking sides: clashing views in social psychology* (2nd edn) (New York, McGraw-Hill), 74–91.

Gergen, K. J. (1973) Social psychology as history, *Journal of Personality and Social Psychology*, 26(2), 309–320.

Haidt, J. (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment, *Psychological Review*, 108(4), 814–834.

Hardy, S. A. & Carlo, G. (2005) Identity as a source of moral motivation, *Human Development*, 48(4), 232–256.

Hart, D. (2005) The development of moral identity, in: G. Carlo & C. P. Edwards (Eds) *Nebraska Symposium on Motivation. Volume 51: moral motivation through the life span* (Lincoln, University of Nebraska Press), 165–196.

Hart, D., Atkins, R. & Donnelly, T. M. (2006) Community service and moral development, in: M. Killen & J. G. Smetana (Eds) *Handbook of moral development* (Mahwah, NJ, Erlbaum), 633–656.

Hart, D. & Fegley, S. (1995) Prosocial behavior and caring in adolescence: relations to self-understanding and social judgment, *Child Development*, 66(5), 1346–1359.

Hartshorne, H. & May, M. A. (1928) *Studies in the nature of character. Volume 1: studies in deceit* (New York, Macmillan).

Hartshorne, H., May, M. A. & Maller, J. B. (1929) *Studies in the nature of character. Volume 2: studies in self-control* (New York, Macmillan).

Hartshorne, H., May, M. A. & Shuttleworth, F. K. (1930) *Studies in the nature of character. Volume 3: studies in the organization of character* (New York, Macmillan).

Heine, S. J., Proulx, T. & Vohs, K. D. (2006) The meaning maintenance model: on the coherence of social motivations, *Personality and Social Psychology Review*, 10(2), 88–110.

Hennig, K. H. & Walker, L. J. (2008) The darker side of accommodating others: examining the interpersonal structure of maladaptive constructs, *Journal of Research in Personality*, 42(1), 2–21.

Hermans, H. J. M., Kempen, H. J. G. & van Loon, R. J. P. (1992) The dialogical self: beyond individualism and rationalism, *American Psychologist*, 47(1), 23–33.

Hoffman, M. L. (2000) *Empathy and moral development: implications for caring and justice* (Cambridge, Cambridge University Press).

Holland, D. (1997) Selves as cultured: as told by an anthropologist who lacks a soul, in: R. D. Ashmore & L. J. Jussim (Eds) *Self and identity: fundamental issues* (New York, Oxford University Press), 160–190.

Hoshino-Browne, E., Zanna, A. S., Spencer, S. J., Zanna, M. P., Kitayama, S. & Lackenbauer, S. (2005) On the cultural guises of cognitive dissonance: the case of Easterners and Westerners, *Journal of Personality and Social Psychology*, 89(3), 294–310.

James, W. (1890) *The principles of psychology* (New York, Holt).
Kaiser, F. G. & Wilson, M. (2000) Assessing people's general ecological behavior: a cross-cultural measure, *Journal of Applied Social Psychology*, 30(5), 952–978.
Kant, I. (1785/1964) *Groundwork of the metaphysic of morals* (H. J. Paton, Trans.) (New York, Harper & Row).
Kohlberg, L. (1969) Stage and sequence: the cognitive-developmental approach to socialization, in: D. A. Goslin (Ed.) *Handbook of socialization theory and research* (Chicago, Rand McNally), 347–480.
Kohlberg, L. (1971a) Stages of moral development as a basis for moral education, in: C. M. Beck, B. S. Crittenden & E. V. Sullivan (Eds) *Moral education: interdisciplinary approaches* (Toronto, University of Toronto Press), 23–92.
Kohlberg, L. (1971b) From is to ought: how to commit the naturalistic fallacy and get away with it in the study of moral development, in: T. Mischel (Ed.) *Cognitive development and epistemology* (New York, Academic Press), 151–235.
Kohlberg, L. (1981) *Essays on moral development. Volume 1: the philosophy of moral development* (San Francisco, Harper & Row).
Kohlberg, L. (1984) *Essays on moral development. Volume 2: the psychology of moral development* (San Francisco, Harper & Row).
Kohlberg, L. & Diessner, R. (1991) A cognitive-developmental approach to moral attachment, in: J. L. Gewirtz & W. M. Kurtines (Eds) *Intersections with attachment* (Hillsdale, NJ, Erlbaum), 229–246.
Kohlberg, L. & Mayer, R. (1972) Development as the aim of education, *Harvard Educational Review*, 42(4), 449–496.
Kohlberg, L. & Power, C. (1981) Moral development, religious thinking and the question of a seventh stage, in: L. Kohlberg (Ed.) *Essays on moral development. Volume 1: the philosophy of moral development* (San Francisco, Harper & Row), 311–372.
Kvale, S. (1990) Postmodern psychology: a contradictio in adjecto? *Humanist Psychologist*, 18(1), 35–54.
Lapsley, D. K. (2006) Moral stage theory, in: M. Killen & J. G. Smetana (Eds) *Handbook of moral development* (Mahwah, NJ, Erlbaum), 37–66.
Lapsley, D. K. & Lasky, B. (2001) Prototypic moral character, *Identity*, 1(4), 345–363.
Lapsley, D. K. & Narvaez, D. (2004) A social-cognitive approach to the moral personality, in: D. K. Lapsley & D. Narvaez (Eds) *Moral development, self and identity* (Mahwah, NJ, Erlbaum), 189–212.
Matsuba, M. K. & Walker, L. J. (2004) Extraordinary moral commitment: young adults working for social organizations, *Journal of Personality*, 72(2), 413–436.
McAdams, D. P. (1993) *The stories we live by: personal myths and the making of the self* (New York, Guilford).
McAdams, D. P. & Pals, J. L. (2006) A new Big Five: fundamental principles for an integrative science of personality, *American Psychologist*, 61(3), 204–217.
McClelland, D. C., Koestner, R. & Weinberger, J. (1989) How do self-attributed and implicit motives differ? *Psychological Review*, 96(4), 690–702.
Mischel, W. (1984) Convergences and challenges in the search for consistency, *American Psychologist*, 39(4), 351–364.
Nisan, M. (1995) Moral balance: a model for moral choice, in: W. M. Kurtines & J. L. Gewirtz (Eds) *Moral development: an introduction* (Boston, Allyn and Bacon), 475–492.
Noam, G. G. & Wren, T. E. (Eds) (1993) *The moral self* (Cambridge, MA, MIT Press).
Nucci, L. (2004) Reflections on the moral self-construct, in: D. K. Lapsley & D. Narvaez (Eds) *Moral development, self, and identity* (Mahwah, NJ, Erlbaum), 111–132.
Piaget, J. (1977) *The moral judgment of the child* (M. Gabain, Trans.) (Harmondsworth, Penguin).
Proulx, T. & Chandler, M. (2007) Jekyll & Hyde in the East & West: cross-cultural variations in conceptions of self-unity, *Revue Internationale de Psychologie Sociale*, 20(2), 57–77.

Rawls, J. (1971) *A theory of justice* (Cambridge, MA, Harvard University Press).

Rest, J. R. (1984) The major components of morality, in: W. M. Kurtines & J. L. Gewirtz (Eds) *Morality, moral behavior, and moral development* (New York, Wiley), 24–38.

Schrödinger, E. (1935) Die gegenwärtige Situation in der Quantenmeckanik [The present situation in quantum mechanics], *Die Naturwissenschaften*, 23, 807–812, 823–828, 844–849.

Schwartz, S. H. (1994) Are there universal aspects in the structure and contents of human values?, *Journal of Social Issues*, 50(4), 19–45.

Sher, G. (1997) *Beyond neutrality: perfectionism and politics* (Cambridge, Cambridge University Press).

Shweder, R. A. (with Bourne, E. J.). (1991) Does the concept of the person vary cross-culturally?, in: R. A. Shweder (Ed.) *Thinking through cultures: expeditions in cultural psychology* (Cambridge, MA, Harvard University Press), 113–155.

Smith, K. D., Türk Smith, S. & Christopher, J. C. (2007) What defines the good person? Cross-cultural comparisons of experts' models with lay prototypes, *Journal of Cross-Cultural Psychology*, 38(3), 333–360.

Steele, C. M. (1988) The psychology of self-affirmation: sustaining the integrity of the self, in: L. Berkowitz (Ed.) *Advances in experimental social psychology* (vol. 21) (San Diego, Academic Press), 261–302.

Walker, L. J. (2004) Gus in the gap: bridging the judgment-action in moral functioning, in: D. K. Lapsley & D. Narvaez (Eds) *Moral development, self and identity* (Mahwah, NJ, Erlbaum), 1–20.

Walker, L. J. & Frimer, J. A. (2007) Moral personality of brave and caring exemplars, *Journal of Personality and Social Psychology*, 93(5), 845–860.

Walker, L. J. & Hennig, K. H. (1997) Moral development in the broader context of personality, in: S. Hala (Ed.) *The development of social cognition* (East Sussex, England, Psychology Press), 297–327.

Walker, L. J. & Pitts, R. C. (1998) Naturalistic conceptions of moral maturity, *Developmental Psychology*, 34(3), 403–419.