Author Accepted Manuscript. To appear in a special issue of *Synthese* on *The Cultural Evolution of Human Social Cognition,* edited by Richard Moore, Rachael Brown & Cecilia Heyes.

24 July 2019

IS MORALITY A GADGET?

NATURE, NURTURE AND CULTURE IN MORAL DEVELOPMENT

Cecilia Heyes

All Souls College & Department of Experimental Psychology

University of Oxford

Oxford OX1 4AL

United Kingdom

cecilia.heyes@all-souls.ox.ac.uk

Abstract

Research on ‘moral learning’ examines the roles of domain-general processes, such as Bayesian inference and reinforcement learning, in the development of moral beliefs and values. Alert to the power of these processes and equipped with both the analytic resources of philosophy and the empirical methods of psychology, ‘moral learners’ are ideally placed to discover the contributions of nature, nurture and culture to moral development. However, I argue that to achieve these objectives research on moral learning needs to 1) overcome nativist bias, and 2) distinguish two kinds of social learning: learning *from* and learning *about*. An agent learns *from* others when there is transfer of competence - what the learner learns is similar to, and causally dependent on, what the model knows. When an agent learns *about* the social world there is no transfer of competence - observable features of other agents are just the content of what-is-learned. Learning *from* does not require explicit instruction. A novice can learn from an expert who is ‘leaking’ her morality in the form of emotionally charged behaviour or involuntary use of vocabulary. To the extent that moral development depends on learning *from* other agents, there is the potential for cultural selection of moral beliefs and values.

Keywords: Cultural evolution; social learning; moral psychology; moral learning; moral development; nativism.

Moral psychology - a field in which philosophers are at least as prominent as psychologists - is growing exponentially (Priva and Austerweil 2015). Some would say it is blossoming, others that it is spreading like a weed, but even detractors must admit that, since it emerged 15-20 years ago, moral psychology has told us a great deal about what people consider to be wrong, the types of psychological and neurobiological mechanisms involved in making moral judgements, and where those mechanisms come from (Cushman, Kumar & Railton 2017). This article focusses on the last of these issues, on questions about moral development. I am enthusiastic about moral psychology as an interdisciplinary enterprise but I worry about the way it is tackling moral development. Moral psychology seems to be proceeding as if human psychological development draws on just two sources of information, nature and nurture, when in fact it draws on three – nature, nurture, and culture (Heyes 2018a; Shea 2012).

I use 'nature' to refer to genetic contributions to development - information about the environment obtained by natural selection and carried by DNA sequences. 'Nurture' contributes to the extent that development depends on information about the environment obtained by direct interaction between the developing system and the world in which it is developing. 'Culture' plays a role to the extent that development of a characteristic (morphological, physiological, or psychological) depends on social inheritance – information that is passed from one generation to the next, not via DNA sequences, but via social interaction (Lewens 2015). Understood in this way, as socially inherited information, culture is like nature in contributing inherited information to development, and like nurture in its dependence on learning. However, culture is different from nature in depending on social rather than genetic inheritance, and different from nurture in depending on a special kind of learning, 'cultural learning'.

Cultural learning has been characterised in a variety of ways (e.g. Tomasello, Kruger and Ratner 1993). In this article I suggest that it is helpful to think of cultural learning as learning *from* other agents. What is culturally learned from other agents can be *about* other agents or the inanimate world. Viewed in this way, cultural learning contrasts with learning about other agents or the inanimate world in a way that does not involve transfer of information from other agents. The alternative to cultural learning, which is part of nurture's contributions to development, is sometimes characterised as 'trial-and-error' learning or, as in the preceding paragraph, as learning by 'direct interaction' with the social or asocial environment.

An agent can acquire through cultural learning information she could not gain through her own efforts. For example, children in West Africa learn from adults how to 'wet' or ferment bitter manioc (cassava) so that it provides a safe and plentiful source of starch in the diet. The process, which detoxifies manioc, has several stages involving scraping, grating, washing and boiling. If novices relied not on cultural learning but instead on solitary trial-and-error learning - direct interaction with manioc – they would almost certainly die of cyanide poisoning before finding a safe method of preparation (Henrich 2015; Rozin 1988).

Notice that 'culture' is not always used, as it is here, to refer to socially inherited information. 'Culture' is often understood to refer to attributes that vary between social groups. Of course, both of these uses are legitimate, but I shall argue that problems arise when they are conflated. Socially inherited information can, but does not invariably, give rise to differences between social groups, and differences between social groups are not always due to socially inherited information.

Both nature and nurture contribute to the development of all biological characteristics in all organisms. There are no pure cases in which nature alone, or nurture alone, is responsible for mature form. Some characteristics, mostly found in humans, have a third ingredient, culture; their development also depends on socially inherited information. Since the 1980s research on 'cultural evolution' has shown through field work, mathematical modelling, and laboratory experiments that socially inherited information plays a dominant role in the development of many of our tools and technical skills, such as manioc processing, and in the development of our explicit beliefs and preferences, such as the preference for a large or small family (Campbell 1965; Cavalli-Sforza & Feldman 1981; Henrich 2015; Richerson and Boyd 2005). Recently I argued that distinctively human cognitive processes – such as language, mindreading, and imitation – are also products of socially inherited information. I suggested that, rather than being 'cognitive instincts' (Pinker 2003), distinctively human cognitive processes are 'cognitive gadgets'; they are constructed in the course of development, through social interaction, from old, genetically inherited parts, and 'designed' by selection operating on cultural rather than genetic variants (Heyes 2018a). Unlike some other distinctively human cognitive processes, morality may not involve dedicated, computationally distinctive psychological mechanisms (Greene 2015). However, morality certainly involves distinctive beliefs and preferences; judgements and intuitions about what is and is not right, about how people should and should not behave. Therefore, the recent rapid growth in understanding of cultural evolution (Youngblood & Lahti 2018) raises a fundamental question (Sterelny 2010): How important is socially inherited information in the development of morality? Or, to put it another way, is morality a gadget? To the extent that the development of morality depends on socially inherited information, there is the potential for cultural selection - the potential for morality to become adaptive, good at doing its job, through cultural rather than genetic evolution.

About a year ago I began to immerse myself in the literature on moral psychology hoping to find out whether morality is a gadget. So far, I have failed. Maybe I've been obtuse or looking in the wrong places, but I suspect there are (also) three more interesting reasons: 1) Moral psychology has a nativist bias. In spite of a recent surge of interest in 'moral learning' (Cushman, Kumar & Railton 2017; Railton 2017), most moral psychologists continue to be preoccupied by nature's contributions to the development of morality. 2) When one is trying to trace sources of information about the social environment it is difficult to distinguish nurture from culture; to keep apart cases in which the child learns *about* other people and cases in which she learns *from* other people. 3) Research on moral learning has the potential to overcome both of these problems but to do so it would need not only to overcome nativist bias and distinguish more clearly between learning *about* and learning *from* others, but also to take more interest in the opportunities for learning experienced by children in their everyday lives. Modelling and laboratory experiments can tell us how morality could *possibly* be learned, but unless these methods are combined with naturalistic observation they cannot tell us how morality is *actually* learned. I will discuss each of these three issues in turn.

1. Nativism in moral psychology

Moral psychology has been influenced greatly by what Fodor (2001) called "High Church evolutionary psychology". This is the project, most closely associated with the work of Cosmides and Tooby (1994) and Pinker (2003), that casts the human mind as a collection of "innate modules" or "cognitive instincts" - special-purpose, genetically inherited psychological mechanisms, each tailored by natural selection to do a particular adaptive task. According to Joyce (2013), a defender of moral nativism, John Stuart Mill's declaration that "moral feelings are acquired, not innate" (Mill 1861, p. 527) had rarely been challenged before Cosmides and Tooby (1992) made "cheater detection", a morally-relevant activity, into the leading example of an innate psychological module. In the wake of cheater detection, many philosophers and psychologists - increasingly likely, over the years, to identify as moral psychologists - have advanced and defended a nativist view of moral development (e.g. Bloom 2012; Dwyer 2006; Haidt 2012; Hamlin 2013; Hauser 2006; Joyce 2013; Nichols 2005).

Nativists in moral psychology acknowledge that learning and 'culture' (of some sort) are important in moral development and vary widely in what they take to be nature's contributions. For example, it has been suggested that humans genetically inherit a "moral grammar" or propensity to develop specific moral rules (e.g. prohibiting sex with a sibling; Hauser 2006); categories of moral evaluation (e.g. care/harm, fairness/cheating, purity/degradation; Haidt 2012); and a disposition to distinguish moral from conventional norms (Turiel 2002). Thus, moral nativists do not present a united, or indeed an unreasonable, front. However, following the lead of High Church evolutionary psychology, there is a strong tendency among moral psychologists to see nature as providing a critically important foundation or “first draft” of moral development (Graham et al. under review).

Surprisingly, although there is opposition to moral nativism (e.g. Greene 2017; Prinz 2014; Sterelny 2010), it rarely comes from leading figures in cultural evolutionary studies (e.g. Henrich 2015; Richerson et al. 2016). Dual-inheritance theorists, who have pioneered the application of population genetic models to cultural change, follow Darwin (1874) in assuming that morality is founded on “social instincts”, including “moral intuitions like sympathy and patriotism” (Richerson et al. 2016, p. 16). They do not explain in contemporary terms what these instincts amount to at the psychological level – for example, whether they are emotional or cognitive, rule-like or categorical - but it is clear that these cultural evolutionists see the social instincts as powerful and innate. They insist that much of the selection pressure for the evolution of social instincts came from the cultural environment, from the advantages of being able to acquire norms from group members, but assume without comment that this pressure induced genetic change, and therefore that our social instincts are part of nature’s contribution to moral development.

A preoccupation with nature’s contributions is not necessarily a bias. If there were compelling evidence that morality is built on substantial genetically inherited foundations, the preoccupation would be rational and healthy. But there are signs that nativism really is a bias in moral psychology - signs not merely that the balance of evidence is against many nativist claims, but that nativism is beginning to function as “a sacred grating behind which each novice is commanded to kneel in order that he may never see the real world, except through its interstices” (Tolman 1932, p.394)[1].

The clearest sign of nativist bias in moral psychology is a pervasive tendency to interpret the early development of a characteristic as evidence that the characteristic is ‘innate’ or genetically inherited, in spite of ample evidence that infants are prodigious learners (e.g. Aslin, Saffran & Newport 1998). For example, evidence that one-year-olds help strangers with no obvious benefit to themselves (Brownell, Ramani, & Zerwas 2006; Warneken & Tomasello 2006), and that three-year-olds share rewards for a joint task equally with their collaborator (Warneken, Lohse, Melis, &

[1] In this article I try to demonstrate nativist bias by focussing on specific, prominently published and highly-cited bodies of empirical work. In each case, I argue (here or in cited articles elsewhere) that the empirical results are understood by moral psychologists to support nativist hypotheses when they can be explained as plausibly, or more plausibly, by alternative non-nativist hypotheses. This approach is subject to the charge of ‘cherry picking’ but I doubt there is a feasible alternative. Random sampling of moral psychology would be likely to yield many empirical studies that have had little or no influence on the field. In the right journal and format, I would be open to a challenge in which moral nativists identify the empirical studies that they believe provide the strongest evidence for their position, and sceptics – including me - respond with objections and, crucially, proposals for further empirical tests that would distinguish nativist from non-nativist interpretations.

Tomasello 2011), are taken to indicate genetically inherited propensities to help and to share without consideration of where or how the behaviours could be learned (Graham et al. 2017).

There are many other examples of nativist bias, but I will consider just two in detail. The first comes from research that, until recently, played a pivotal role in sustaining moral nativism by seeming to show that infants as young as six months of age prefer "helpers" to "hinderers" in third-party interactions. In the original study (Hamlin, Wynn & Bloom 2007), published in *Nature* magazine, 6-10- month-olds were shown a sequence of events in which a red circle with googly eyes, a "climber", ascended an incline on three successive occasions. On the first two occasions it got half way up the incline and then moved back down to the bottom. On the third occasion, another shape entered the scene when the circle was half way up. In some trials the second shape was a yellow triangle, a "helper", which contacted the red circle before the triangle and circle moved together to the top of the incline. In other trials, the second shape was a blue square, a "hinderer", which contacted the red circle before the square and circle moved together to the bottom of the incline. In subsequent preference tests, the infants were reported to be more likely to reach for the helper shape than the hinderer shape when the two were presented side-by-side, and, at 10 months, to be more surprised (as indicated by looking time) when the climber moved towards the hinderer shape than when the climber moved towards the helper shape.

This helper preference effect, which has been cited more than 1200 times, was interpreted by the authors as showing that very young infants are not only capable of interpreting the movements of geometric shapes in intentional terms – that babies are mindreaders - but that their "expectations about others and their own preferences are motivated by the perceived goodness and badness of the characters" (Bloom 2012, p. 11; Hamlin 2013) – that babies are moralistic mindreaders. One might expect the evidence for such a big claim to have been scrutinised with particular care, but it was five years before anyone noticed that the original study and many of the follow-up experiments contained a major confound: The helper shape was paired with an attractive bouncing movement at the top of the hill, and the hinderer shape was paired with an aversive collision event at the bottom, allowing infants to learn a helper preference by association (Scarf, Imuta, Columbo and Hayne 2012; reply from Hamlin, Wynn & Bloom 2012). Furthermore, it was nearly 10 years before it became clear that the basic effect was not replicating reliably outside the laboratory in which it was originally observed (Holvoet, Scola, Arciszewski and Picard 2016). The helper preference has been found in only 37% of experiments conducted by other researchers (Hinten, Labuschagne, Boden and Scarf 2018). I do not doubt that those who discovered the helper preference acted in good faith and with scientific integrity (Bird 2018), but these delays suggest a lack of due diligence on the part of ‘consumers’; an excessive willingness to trust findings that are, under a rich interpretation, consistent with moral nativism.

The second example of nativist bias relates to a phenomenon with yet broader implications for moral psychology – emotional contagion or affective empathy. Emotional contagion is what happens in the first few hundred milliseconds after one looks at Figure 1. Viewing this photograph, one feels the child’s distress immediately and viscerally - in the limbs, gut and respiratory system. Recent research in cognitive neuroscience suggests that emotional contagion – rapid, matching, visceral reactions to positive and negative emotions - plays a role in most, if not all, moral responses; in generating pro- and anti-social behaviour, and in formulating moral judgements. Brain imaging and electrophysiological measures indicate that emotional contagion is ubiquitous. It occurs whenever we see or are told about emotionally charged situations, even when we are actively encouraged to keep a cool head, or distracted by another task (e.g. Fan et al. 2011; Gonzalez-Liencres et al. 2013; Lamm et al. 2011).

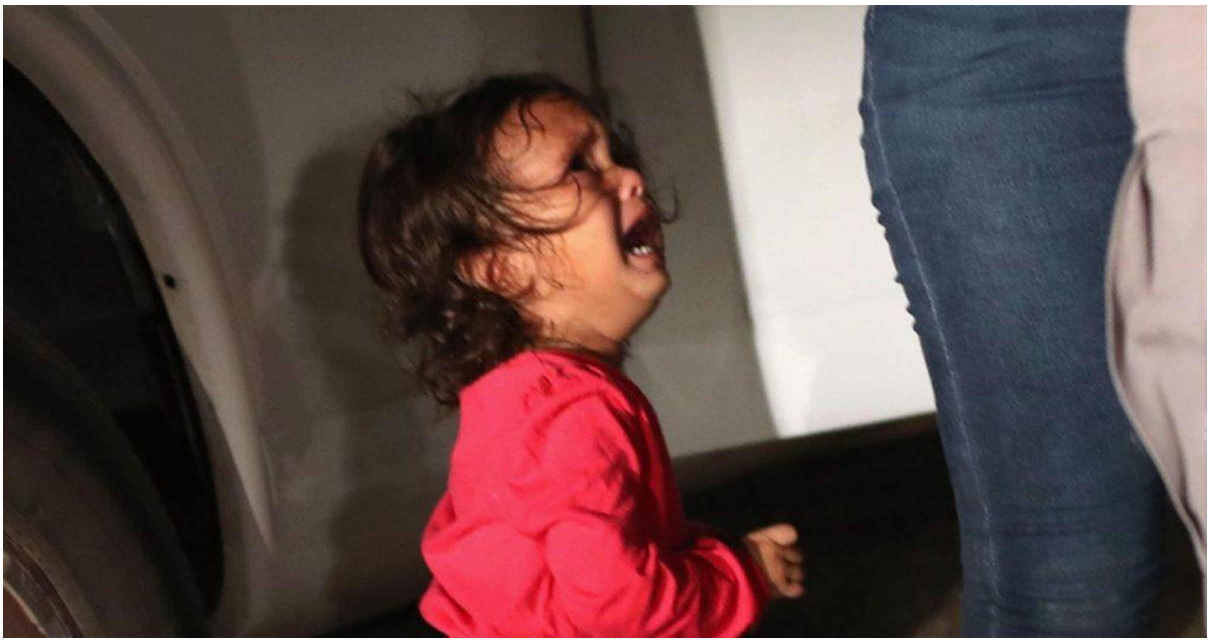


Figure 1. A photograph taken in June 2018 when, on the authority of President Donald Trump, United States immigration officials were separating children from their families at the Mexican border (Getty Images).

It is widely believed that emotional contagion is ‘innate’, that humans and other mammals genetically inherit a propensity to feel the emotions they witness in others, and that experience plays a minimal role in the development of this propensity. This is due in large measure to an article published by Preston and de Waal in 2002, around the time that moral psychology was getting started. Focussing on nonhuman animals, Preston and de Waal proposed that emotional contagion depends on a “perception-action mechanism” (PAM). They did not say how the mechanism works but suggested that PAM was favoured by natural selection operating on genetic variants during the early evolution of mammals, when parental care was becoming important, and during primate evolution, when cooperation among group members was increasingly at a premium. The 2002 article has been highly influential despite exhibiting clear confirmation bias. It argued in some detail that PAM could explain empirical effects reported in the literature on emotional contagion (effects of similarity, familiarity, past experience, explicit teaching, and salience) but failed to acknowledge that alternative theories, assigning a greater role to experience in the development of emotional contagion, could explain these effects at least as well (Heyes 2018b).

More seriously, neither the 2002 article nor a sequel on PAM (de Waal & Preston 2017), acknowledged findings that, although predicted by learning models, are very difficult for PAM to explain. For example, if contagious crying in human newborns was due to a genetic adaptation for empathy – an other-directed emotion - one would expect it to be activated more by the cries of other infants than by playback of the infant’s own cries, and to be activated as much by the cries of older infants as by those of fellow newborns. But these are not the patterns observed. Newborns cry more when they hear their own pre-recorded cries than when they hear another newborn crying, and less when they hear the cries of a 6-month-old than those of a newborn infant (Simner 1971). These findings suggest that, rather than having a genetically inherited PAM, infants learn to cry contagiously through hearing their own cries when feeling distressed (Heyes 2018b; Ruffman et al. 2017).

Other unacknowledged anomalies have emerged in studies of emotional contagion in adult humans. For example, automatic empathic responses can be converted into automatic “envious” or “sadistic” responses by a brief period of counter-conditioning (Englis et al. 1982). Electrophysiological recordings from face muscles indicate that people who have observed, in the context of a game, smiling while grimacing, and vice versa, are more likely than other people to grimace when they see smiling, and to smile when they see grimacing (Englis et al. 1982). Of course, intensive training might overcome the influence of a genetically inherited PAM, but the people in this experiment received very little training. They were given only 16 learning trials – they saw grimacing while smiling, or vice versa, on just 16 occasions – and envious or sadistic responses were not rewarded in any way.

The PAM hypothesis has been published in high profile journals, cited thousands of times, and used to support Moral Foundations Theory, a prominent framework suggesting that all moral intuitions and judgements depend on a set of genetically inherited categories of evaluation (Graham et al. under review, Haidt 2012). More generally, PAM has fortified the view that empathy is antagonistic to morality because empathy inevitably favours members of one's own group over others (Bloom 2017, Graham et al. 2017; Singer 2015). Given the obvious weakness of PAM's empirical base, it is hard to explain the extent of this model's influence without inferring nativist bias. Like the helper-hinder case, the PAM example suggests that moral psychology has a soft spot for theories and data underlining nature's contributions to moral development.

2. Learning *about* and learning *from* other people

Nativist bias makes it hard to get a clear picture of moral development because it draws attention away from the contributions of experience. It focusses on nature at the expense of both nurture and culture. A second problem is that nurture and culture are readily confused in the social domain, including the moral domain. Even when attention is focussed on the roles of experience in moral development, the influence of culture - socially inherited information - is easy to conflate with that of nurture, i.e. information acquired by trial-and-error learning or varieties of 'social learning' that do not support inheritance.

It is empirically but not conceptually demanding to parse the contributions of culture and nurture to the development of asocial capacities – skills in dealing with the world of things rather than the world of people. My ability to make pastry – an asocial capacity in that it requires me to know about fat, flour and water rather than about other people - arose from two kinds of learning that are obviously different from one another. On the one hand, there was cultural learning in which my mother, passing on a skill she inherited from her mother, taught me the ingredients and utensils to use, and demonstrated a rubbing technique, involving repetitive movement of the thumbs over the fingertips, which combines the fat with the flour. On the other hand, there was practice-based, trial-and-error learning which I did by myself. Alone in the kitchen, I repeated the thumb and finger movements until they were fluent, and tried out different proportions of fat and flour, using my own palate to test the results. So, if we identify the contributions of culture with information inherited from others through social interaction, and the contributions of nurture with things learned by trial-and-error, it is pretty clear which aspects of my pastry skill are due to culture (ingredients, utensils, topography of the rubbing technique), which are due to nurture (exactly proportions of the ingredients, fluent execution of the rubbing technique), and how these contributions were combined.

Parsing the development of social capacities, competence in dealing with other people, presents more of a conceptual challenge. Consider the case of configural face processing – the ability to recognise faces, not using distinctive elements (e.g. a bulbous nose, violet eyes), but according to the overall spatial arrangement of the features (e.g. the location of the nose relative to the locations of the eyes and mouth; see Figure 2; Murphy et al 2017). Configural face processing is a social competence – it enables us to distinguish one person from another – that depends on social experience for its development. To become capable of processing faces configurally, rather than elementally, a child must see faces (Le Grand et al. 2004; Michel et al. 2006; Susilo et al. 2009), and most of the faces seen by a typical child are faces of people who belong to the same 'culture' (*sensu* social group) as the child. However, there is no reason to suppose that configural face processing is culturally inherited.

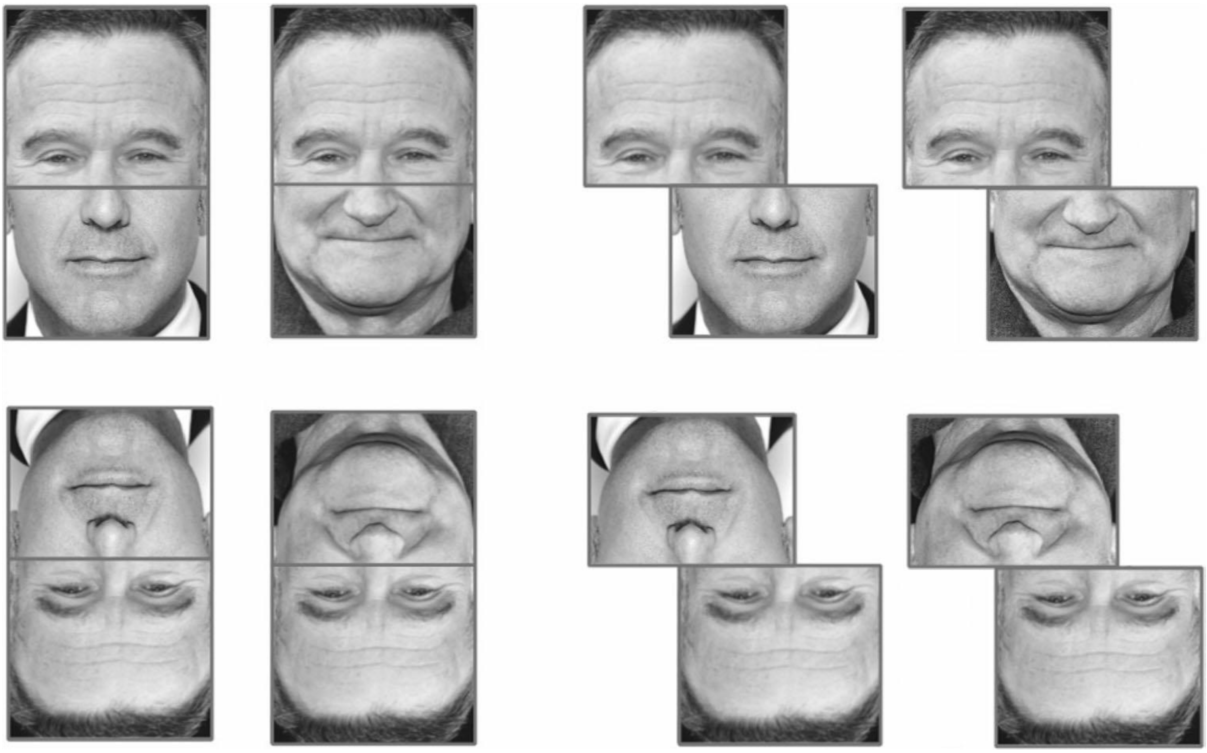
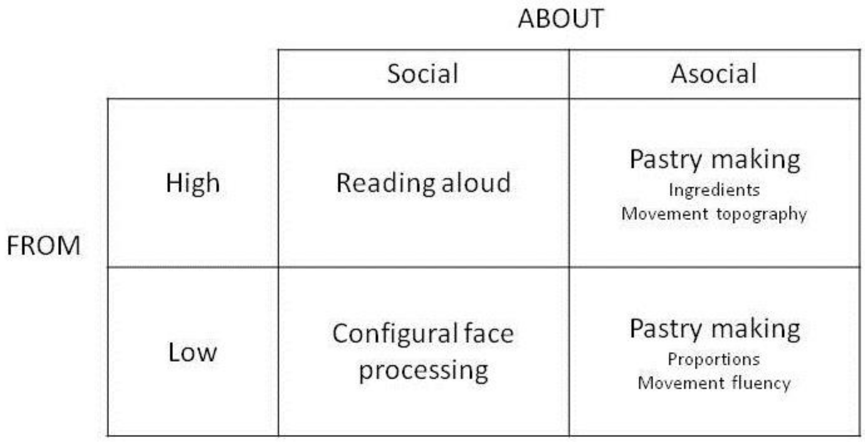


Figure 2. The 'composite face illusion'. The images in the top row are identical; each shows the upper part of Robin Williams' face. But the image on the left does not look like Robin Williams because it is spatially aligned with the lower half of someone else's face. When the upper and lower images are aligned, they are processed 'as a whole', configurally, rather than in an elemental way (reproduced with permission from Murphy, Gray and Cook 2017).

During the development of configural processing children learn *about* faces. Rather than facilitating or assisting learning, the faces of other people are what-is-learned. There is no transfer of expertise from 'face owners' to the children who see their faces. A child exposed only to the faces of people who, due to brain injury, were not capable of configural processing would develop the competence just as successfully as children exposed to the faces of neurotypical people. As it happens, most human faces are ‘backed’ by minds capable of configural processing, but the competence of face owners does not play a causal role in the acquisition of configural processing by face viewers. The role of faces in the development of configural face processing is like the role of fat and flour in the development of pastry making; the novice learns *about* them, not *from* them. Thus, while nurture contributes via social experience to the development of configural face processing, as far as we know, culture (*sensu* socially inherited information) does not.

Compare the development of configural face processing with learning to read aloud. Reading is also a social competence that depends heavily on social experience for its development. But in the case of reading, children learn not only *about* but *from* other people. Printed words are like faces in being what-is-learned. To develop the cognitive mechanisms of reading (Coltheart et al. 2001) a child must be exposed to printed words, and printed words are social stimuli in that they are products of human action. However, exposure to printed words is radically insufficient for learning to read aloud. At minimum, the child must also hear words while looking at the printed versions – they must experience correlations between graphemes and phonemes – and this experience comes from other people who are able to provide it because they already have the cognitive mechanisms responsible for reading aloud. A child could acquire configural face processing by looking at the faces of people who cannot do configural processing themselves, but a child could not learn to read by listening to the speech of illiterate people. Unless the speaker is also a reader, and looking at the same text as the child, the word sounds produced by the speaker will not be correlated with the word forms on the page. The reading competence of the voice owner plays a causal role in acquisition of reading competence by the voice listener. Both nurture and culture – socially inherited information – contribute to the development of reading.



		ABOUT	
		Social	Asocial
FROM	High	Reading aloud	Pastry making Ingredients Movement topography
	Low	Configural face processing	Pastry making Proportions Movement fluency

Table 1. Examples of learning *about* (nurture) and learning *from* (culture). Reading aloud is a competence where we learn *about* and *from* other people. Configural face processing is a

competence where we learn *about* but not *from* other people. Pastry making is an asocial skill - the novice learns about fat, flour and water, not about other people – and some parts of the skill come from other people (e.g. information about ingredients, utensils, finger movement topography), whereas other parts are mastered through the learner's own efforts (e.g. proportions of ingredients, finger movement fluency).

So, we have an example in each cell of a 2 x 2 table (see Table 1). Reading aloud is a competence where we learn *about* and *from* other people. Both nurture and culture contribute to development. Configural face processing is a competence where we learn *about* but not *from* other people. Nurture contributes but, as far as we know, culture does not. Pastry making appears in two cells. It is an asocial skill - the novice learns about fat, flour and water, not about other people – and, at least in my case, some parts of the skill came from other people (I inherited information about ingredients, utensils, and finger movement topography from my mother), whereas other parts I learned through my own efforts (proportions of ingredients, and finger movement fluency).

In discussing these examples, I have tried to make the social/asocial, and nurture/culture distinctions as sharp as possible because - like Sterelny (2009), who distinguished social contents from social channels of learning - I think they are important and often elided. However, as signalled by the pastry example, I do not imagine that every 'whole skill' – defined as such by folk psychology or cognitive science - can be neatly assigned to one of the four cells in Table 1. Component skills, such as movement topography in the case of pastry making, are typically classifiable in this way, but it is often more helpful to represent whole skills in a two-dimensional space. In such a space, reading aloud is high on both dimensions – the social character of what-is-learned (learning *about* other agents), and the extent of social inheritance (learning *from* other agents). Configural face processing is high on the first dimension and low on the second, and pastry making, viewed as a whole, is low on the first and intermediate on the second.

Like reading aloud and configural face processing, moralising is, among other things, a social skill. To be a competent moraliser an agent needs lots of information about her social world. At minimum, she needs information about the typical effects of various actions on the well-being of others. On a richer cognitive view of morality, she may also need explicit beliefs about what the members of her social group consider to be right and wrong, the reasons for these moral norms, and whether the reasons are justified. In either case, whether one takes a lean or rich view of what it is to be moral, it is clear that much of the necessary information is about agents - how they respond to, and what they believe about, various actions. Consequently, if one understands 'culture' to refer only to features of people that vary across social groups, a bland conclusion will follow: regardless of how it develops - whatever the contributions of nature, nurture and social inheritance - morality is always 'cultural' because it is *about* other agents.

I have not found this view stated explicitly in moral psychology, but sometimes I think I get a glimpse of it hovering in the background. In a recent summary of Moral Foundations Theory (MFT), Graham et al. (under review, p. 3) describe MFT as not only a nativist theory, but

> "a cultural theory that describes the 'editing process' by which the [genetically inherited] universal first draft of the moral mind becomes a culturally specific and culturally competent adult morality. For example, Hindu cultures emphasize respect for elders and other authorities, as can be seen in the common practice of children bowing to elders and often touching elders' feet. By the time these children reach adulthood, they have gained culturally-specific knowledge that may lead them to automatically initiate bowing movements when encountering elders or other revered people. In more individualistic and secular cultures that do not emphasize respect for authority, children are not taught to bow to elders. This might make it easier for them to address authority figures by first name or question their authority later in life. These different social practices in different cultures help explain cultural differences in moral values."

This Hindu example is enough to demonstrate that MFT is a "cultural theory" if we take culture to be features of people that vary across social groups, and learning *about* those features to be sufficient to make the learning cultural. In that case, the content of what-is-learned - bowing to elders, high levels of respect for authority – is sufficient to make the learning cultural because the content relates to practices and dispositions that vary across social groups. As long as what-is-learned is "culture-specific knowledge", the learning is cultural.

When culture is instead understood to be socially inherited information, the Hindu case may or may not be a good example of cultural learning. To find out, we would need to know how children learn to bow to their elders. In one plausible scenario, children are explicitly taught. They are told to perform the action, perhaps gently pushed into a bowing position, and it is explained to them that this action expresses respect for authority. If this is what happens, the learning is certainly cultural. As in the case of reading aloud, children learn *from* other people who already have the competence, and the competence of the experts plays a causal role in learning by the novices.

In another plausible scenario, children learn by observation without commentary. They see adults bowing to their elders and copy the action. At first the children are not clear about who they should bow to, and have no idea what the action signifies, but they make some guesses and test them against what they observe. 'Perhaps I should bow to anyone who is bigger than me? No, that can’t be right, he just bowed to someone smaller...’ Eventually, through hypothesis testing and without instruction, children bow to the right people, and come to see the bowing as an expression of respect for authority. If this is what happens, the learning is minimally cultural. The development of a mature concept of respect may require linguistic input, but the idea that bowing expresses respect for authority does not. The learning is predominantly *about* not *from* other people. Just as configural face processing could be learned by looking at the faces of people who are not themselves capable of configural processing, bowing to elders could be learned by observing the bowing behaviour of people who bow just because it gets them what they want, or for whom bowing is a purely conventional rather than a moral norm.

In the quoted passage, Graham and colleagues mention teaching, saying that children in more individualistic cultures “are not taught to bow to elders”. Teaching is certainly a form of cultural learning, but social inheritance does not require that experts give explicit verbal instructions, or even that they act with the intention of influencing the knowledge or behaviour of novices. Suppose that Hindu children test their hypotheses, not only against observed bowing, as in the second scenario, but also by monitoring the effects of their actions on other people. If others look happy when I bow to a large peer, it strengthens my hunch that I should bow to people who are bigger than me; if they look unhappy or withdraw their attention, it weakens that hypothesis. Use of social feedback in this way effects transfer of the norm – enables me to learn the norm *from* other people - provided that the monitored behaviour is caused by acceptance of the norm, that the people giving feedback are happy or unhappy *because* they take my behaviour to be appropriate or inappropriate in relation to that norm. It is not necessary for me to speculate or know about the causes of their behaviour, or, crucially, for them to intend to influence me with their contented or discontented responses. It is enough for norms to leak out of norm-holders in the form of emotionally charged behaviour. Norm-dependent automatic reactions to observed behaviour are sufficient for cultural learning. Pedagogy, deliberately telling and showing, although powerful for some types of knowledge (Buckwalter & Turri 2014; Cath 2019), especially when it is based on exemplars (Zagzebski 2013), is not required for cultural learning. Therefore, any tendency to equate cultural learning with teaching – by, for example, calling cultural learning “testimony” (Heiphetz & Young 2017) – risks under-estimation of the contribution of socially inherited information to moral development.

I have argued in this section that it is important to distinguish learning *about* other people and learning *from* other people. In the latter case, but not the former, there is social inheritance - transfer of competence from one agent to another. What I get depends on what you have. Only when agents learn *from* others is there the potential for cultural evolution - adaptation of moral beliefs and intuitions via a Darwinian selection process operating on socially inherited, rather than genetically inherited, traits. Failure to distinguish learning *about* (nurture) and learning *from* (culture) can result in both over-estimation and under-estimation of the role of culture in moral development. When the term culture is applied indiscriminately to learning *about* and learning *from*, it can give the impression that the influence of culture on moral development is pervasive. Even if we genetically inherited a tightly specified moral grammar (Dwyer 2006; Hauser 2006), we would need to learn a good deal about other agents in order to apply the grammar in action. On the other hand, if culture is equated with teaching or testimony, rather than with 'learning *from*', it gives the misleading impression that social inheritance contributes to moral development only when adults deliver explicit verbal instruction, or otherwise act with the intention of changing the knowledge or behaviour of children. This equation fails to recognise that morality can also leak out of one agent and into another. It can be learned from social cues as well as social signals. The unintentional emotional reactions of adults to a child's behaviour can act as rewards and punishments that shape the behaviour of novice moralisers both directly and by inspiring new hypotheses about what is right.

3. Moral learning

In the last few years there has been a surge of interest in 'moral learning'. In contrast with other research on moral development, the field known as 'moral learning' is peopled by a lively interactive population of philosophers and psychologists, and is concerned primarily with how domain-general processes of learning contribute to the development of moral rules and values. These processes - which include Bayesian inference, reinforcement learning, and other machine learning techniques – are domain-general in that they operate in the same way, via the same computations, when processing morally-relevant information (e.g. harmful behaviour), social information with minimal moral relevance (e.g. other people's technical skills), and asocial information (e.g. the spatial layout of a forest).

In 2017, *Cognition*, the journal of choice for many moral psychologists, published a special issue on moral learning consisting of 20 articles by major figures in the field. As noted by the editors, these articles communicate a "palpable shared sense of excitement and progress" (Cushman, Kumar & Railton 2017, p.8). Morale is high in the field of moral learning, and with good reason. Many contributors write with penetrating insight, clarity and imagination (e.g. Cushman 2013; Greene 2017; Ho et al. 2017; Railton 2017). Furthermore, while other contemporary research on moral development tends merely to document change over time – for example, how sharing behaviour, or the capacity to distinguish moral from conventional norms, changes between 5 years and 9 years of age - moral learning enthusiasts are doing "process-oriented research" (Rhodes & Wellman 2017). Using a range of the most sophisticated models and methods from cognitive science, they are tackling deep and interesting questions about the processes that drive moral development. And, at the broadest level, research on moral learning is a fine example of productive and apparently frictionless interdisciplinary collaboration. In a way that Hume (1751) could only dream of, moral learning is bringing empirical methods to bear on age-old questions in moral philosophy (Greene 2017; Railton 2017).

Given these strengths, and its focus on domain-general processes, research on moral learning is well-placed to overcome nativist bias (section 1), and to distinguish learning *about* and learning *from* (section 2), so that we can get a clearer picture of the contributions of nature, nurture and culture to moral development. But that does not seem to be the current direction of travel. Nativist bias, and failure to distinguish nurture and culture, are evident even among 'moral learners' – researchers investigating moral learning.

A few moral learners explicitly reject nativism. For example, Greene (2017) recently came out as a moral empiricist, and Rhodes and Wellman (2017) suggest that moral competence is founded on implicit theories - 'intuitive psychology' and 'intuitive sociology' - that are learned in infancy via domain-general mechanisms. But even these moral learners endorse problematic evidence of early mindreading and moral sentiments (see section 1), and most others not only cite that evidence but take it to indicate that nature makes a major contribution to moral development. Kleiman-Weiner, Saxe and Tenenbaum (2017), for example, align their views with Chomsky's early (maximally nativist) theory of language development: "in our framework for moral learning, the challenge of explaining how children learn culturally appropriate weights for different groups of people may be analogous to the challenge of explaining linguistic diversity, and may yield to similar solutions, such as the frameworks of 'principles and parameters' (Baker 2002; Chomsky 1981)".

There is no contradiction here. In principle, domain-general mechanisms of learning, of the kind that interest moral learners, may do nothing more than fill in the culture-specific details of a species-wide, genetically inherited template of moral rules and values. They may proceed from strong, genetically inherited 'priors'; or, to use an older and less Bayesian term, domain-general mechanisms could work within tight genetic 'constraints' (Shettleworth 1972). There is no contradiction, but it is surprising that researchers who understand and emphasise the power of domain-general mechanisms are, on the whole, willing to embrace putative evidence of genetic constraints apparently without criticism or curiosity. It is surprising because much of this evidence rests on the unexamined assumption that moral competence emerges too early in development to be built on genetically unconstrained learning. Moral learners are ideally placed to test this assumption – to come up with alternative, learning-based hypotheses and to test them against the nativist alternatives using behavioural and neurophysiological measures with infants and children. But there appears to be little appetite for such an attempt to measure, rather than presuppose, the contributions of nature and nurture/culture to moral development.

As for the distinction between nature and culture, it is not among those that particularly interest moral learners. They are very interested in the difference between model-free and model-based learning (Glascher, Daw, Dayan and O'Doherty 2010), and the possibility that these two kinds of learning respectively underpin emotional and cognitive, or deontological and utilitarian, aspects of morality (Crockett 2013; Cushman 2013; Greene 2017), but the distinction between learning *about* and learning *from* has a much lower profile.

Cushman (2013) describes the kind of learning involved in the development of moral values as "social learning" and says that "social learning depends on 'observational' and 'instructed' knowledge—that is, on information about rewards and punishment derived from watching other people or listening to their advice" (p 281). This characterisation of social learning lumps together cases in which an agent learns *about* and *from* other agents. Anything described as instruction or advice is likely to involve learning *from* – social inheritance – but much "observational" learning, based on "watching other people" is learning *about*. What the learner learns does not depend on what the model knows. For example, people can learn that an object is dangerous by observing another agent, a model, wincing when she touches the object, but if the model knew the object was dangerous, she probably wouldn't touch it (Debiec and Olsson 2017). Returning to the Hindu example (section 2), watching other people bowing and not bowing to their elders, and the outcomes of these actions, could help me learn through my own efforts that I should bow to elders as a mark of respect. I might come up with this hypothesis all by myself, and have it confirmed by the action-outcome relationships I observe, even if the people I see bowing, and responding to bowing, are acting on habit or in accordance with what they regard as a purely conventional norm.

Rhodes and Wellman (2017) also elide the distinction between nurture and culture, learning *about* and learning *from*:

> "although we suggest that children revise their intuitive theories based on experience, we take a broad view on what these experiences might entail. The 'evidence' that could prompt theory-revision might include children's own observations and direct experiences of morally relevant action, but can also involve input from other sources, including rather subtle features of language" (p. 197).

Although Wellman’s theory-theory suggests that children are like scientists, the “broad view” outlined in this passage implies that explanations of moral development need not distinguish the roles of data from primary sources (learning *about*) and secondary sources (learning *from*), in changing children’s theories. It doesn’t matter whether the child’s hypotheses are self-generated or supplied by others. The “broad view” consigns nurture and culture to one pot marked “evidence”, and, when there is uncertainty about their contributions, Rhodes and Wellman (2017) are apt to assume that nurture is dominant.

For example, Rhodes and Wellman (2017) refer to an excellent “microgenetic study” in which 3-4 year old children were first tested for representational mindreading, and then required to complete four false belief tasks each week for six weeks, and asked to explain the behaviour they observed in the tasks (Rhodes & Wellman 2013). The results showed that it was only the children who were close to understanding false belief at pre-test who benefitted from this experience – who had a better grasp of false belief at the end of the study than at the beginning. As the authors suggested, these findings are consistent with the idea that the development of mindreading involves conceptual change. However, these findings do not show that the primary engine of conceptual change is inside the child’s head rather than in the social environment. In their commentary on the study, Rhodes & Wellman (2017) imply that the conceptual change occurred because the children were “prompted to puzzle over” the anomalies they observed in the false belief tests, e.g. to try to work out for themselves why an agent might go to a place where a desired object cannot be found. That is certainly possible, but it is also possible that the puzzling had its effect via a social loop. In their everyday lives, between microgenetic training sessions, children who had been puzzled may have been more likely to engage adults in conversation about the mind, and to have received in those conversations information that pushed them over the edge into representational mindreading.

It is surprising that Rhodes and Wellman (2017) did not raise the possibility that conversation played a role in their microgenetic effect because, in a beautiful study with Chalik, Rhodes has found evidence that children learn, not just about the mind, but about moral obligations through conversation with their parents (Chalik and Rhodes 2015). Attempting to explain why 4-year-olds express a stronger obligation to help and not to harm in-group than out-group members, Chalik and Rhodes (2015) discovered that, in conversation with their children over a picture book, parents refer more often to fairness when explaining why in-group members should be helped and not harmed. Thus, although the parents were equally likely to say that in-group and out-group members should be helped and not harmed, subtle linguistic cues, of which they were probably unaware, hinted that moral obligations apply only to in-group members. So, it seems that, even in a pedagogical context, morality can leak out of one agent and into another (see section 2 above).

Rhodes and Wellman's tendency to assume that moral development is powered by the hard labour of the child, rather than by information supplied by knowledgeable adults, is evident in the work of other moral learners. Kleiman-Weiner and colleagues (2017) propose that moral development is driven by 'internal alignment' and 'external alignment'. Internal alignment is a ruminative, inward-looking process in which the child tries to identify and iron out inconsistencies between her moral theory and her attitudes towards specific individuals. External alignment is more outward-looking, a process in which children "internalize the values of the people they value, aligning their moral theory to those that they care about" (p. 109). External alignment sounds like learning *from* others, social inheritance, but closer reading suggests that external alignment includes some learning *about,* and excludes unambiguous cases of learning *from*; cases in which children are given explicit moral instruction. Learning *about* creeps in because external alignment is possible whenever a child is rewarded or punished, regardless of whether the rewarding or punishing agent has the knowledge the child acquires from the experience. Explicit instruction in moral principles – a process that allows a child to learn from others without hard labour – is dismissed as rare on the strength of a single study involving two children (Wright and Bartsch 2008).

Ho, MacGlashan, Littman and Cushman (2017) are different. In contrast with other moral learners, Ho and colleagues are clear and firm in distinguishing learning *about* and learning *from*. They encapsulate this distinction snappily as the difference between "adapt" and "adopt" strategies, and focus squarely on the latter, on what cultural learning contributes to moral development. Specifically, Ho and colleagues suggest that "evaluative feedback", rewards and punishments delivered by other agents, is a powerful driver of moral development, and that "social is special" – evaluative feedback is not processed in the same way as other rewards and punishments. Unlike asocial reinforcers – for example, satisfaction when a heavy door opens, or pain when it hits you in the face – social reinforcers - rewards and punishments delivered by other agents, verbally or nonverbally - are interpreted by the child as communicative signals about the social or moral value of an action. Consequently, evaluative learning – the cultural learning that, according to Ho and colleagues, delivers moral competence - depends on inferences about communicative intent, goals and other mental states.

The work of Ho and colleagues on evaluative learning is impressive in at least three respects: It does not confuse learning *about* and learning *from*; it casts a spotlight on the role of learning *from*, cultural learning, in moral development; and it makes a compelling normative case for the value of a particular kind of cultural learning. Specifically, Ho and colleagues argue persuasively that it would be very helpful for moral development if infants and young children could interpret social rewards and punishments as communicative signals. The one problem, as I see it, is that nativist bias leads Ho and colleagues to believe that this powerful kind of moral learning comes online much earlier in development than is likely to be the case. Their evidence that evaluative learning is a primary motor of moral development, rather than a sophisticated form of moral learning which kicks-in after language, comes from empirical studies with marked mentalistic and nativist biases (see section 4 of Ho et al. 2017 for review). These studies, many of them relating to "natural pedagogy", typically find a competence in infancy, describe it in richly intentional terms – in ways implying that infants can interpret communicative signals - and assume the competence, thus described, is genetically inherited. They do not test the rich, mentalistic characterisation of the behaviour against plausibly lean alternatives, or ask whether the competence could be learned (Heyes 2016).

Of course, further testing may reveal that some of the mentalistic and nativist assumptions were justified, but the signs are not good. Where there has been closer examination, it has tended to support subpersonal explanations for phenomena that were assumed to reflect mindreading in infancy (Heyes 2014a; 2014b; Holvoet 2016; Sabbagh & Paulus 2018). For example, evidence that infants are more likely to copy full actions (e.g. an adult model puts a loop of string on a hook) than 'failed attempts' (the adult drops the loop before it reaches the hook) was taken by the original author (Meltzoff 1995) and by Ho et al. (2017) to indicate that infants conceptualise some observed actions as accidental and others as "intended" or "aimed at" particular outcomes. However, subsequent experiments, in which "failed attempts" more fully demonstrated object affordances (e.g. the loop contacted the hook before falling), supported a leaner interpretation (Huang, Heyes and Charman 2006): Infants are more likely to copy full actions, not because they understand models as agents with intentions, but because full actions provide more information about what objects (rather than minds) can do.

If the development of mindreading, and therefore of evaluative learning, follows rather than precedes the development of language – if the cognition required for language learning is less Gricean than we thought (Moore 2016, 2017; Heyes & Frith 2014) – evaluative learning could not contribute to the early development of morality. Would it follow from this that cultural learning plays little or no role in the development of morality before the age of four or five years? Ho and colleagues (2017) might say yes because they doubt that social rewards and punishments can produce enduring change unless they are interpreted by the child as communicative signals. However, a long, cumulative tradition of research on animal learning (Pearce 2008) suggests that this answer would be unduly pessimistic. To see why, let us consider an example used by Ho and colleagues:

> "a father who wishes to teach his daughter to share her toys with her playmates. Thus, he punishes her when she hoards her toys but rewards her for sharing. What is the goal of his behavior? One possibility is to assume that she will treat his evaluative feedback identically to a non-social reward. If so, then his goal must be to shape her behavior by providing an external incentive for the behavior he desires her to perform (sharing). In other words, he hopes that she will fear his continued punishment and seek his continued praise, and so she will share. Intuitively, however, this explanation seems incomplete. At the very least, an obvious problem is that the daughter would no longer be motivated to share once the father is no longer around to shape her behavior." (Ho et al. 2017, p.94)

The first thing to note about this example is that it focusses on the father's intentions – on the "goal" of his behaviour – but neither here nor elsewhere do Ho and colleagues explain why they see the intentions of a moral expert as crucial in determining what can be learned from the expert by a moral novice. As I argued in Section 2, it is possible that experts who leak moral values – who deliver social rewards and punishments inadvertently, as side-effects of their emotional responses to observed behaviour – are just as effective in educating novices as experts who intend to teach. Dad may be just as effective when he's genuinely upset about his daughter's failure to share as when he pretends to be upset in order to change her behaviour.

Second, and more specifically, research on animal learning suggests we need not worry that, if the daughter treats her father's approval and disapproval just like any other rewards and punishments – if she does not 'interpret' them - she will cease to share toys when her father is no longer around. Research in which asocial rewards and punishments are given to rats and pigeons - animals that we have no reason to believe are capable of mindreading – often reveals remarkable 'resistance to extinction'; the animals go on doing what they have been rewarded for doing, and continue to omit previously punished responses, long after the actions in question have ceased to be rewarded or punished by the experimenter (e.g. Delamater 1996; Mackintosh 1974; Pearce 2008). Furthermore, this research with humble animals shows that two conditions, which are likely to be met in the father-daughter example and many other human cases, promote resistance to extinction – 'partial reinforcement' and 'secondary reinforcement'. Behaviours that have been rewarded partially – now and again, rather than every time they occurred – persist for longer in extinction than behaviours that were rewarded continuously before they ceased to be rewarded at all. So, if a father is not always around when his daughter has access to toys, as is likely, her sharing behaviour will be only partially reinforced, and therefore continue for longer when dad withdraws altogether. Similarly, through learning of associations among stimuli, cues that are correlated with external reward become secondary reinforcers, i.e. rewarding in themselves (Carder & Berkowitz 1970). Thus, just as the daughter need not interpret dad's approval *as* approval in order for it to influence her behaviour, she can begin to be influenced by events that were correlated with her father's approval – such as the eager responses of her playmates as she gives them her toys – without decoding the meaning of these secondary reinforcers. Finally, and perhaps most important of all in the case of moral learning, other agents are likely to supplement and extend the father's training regime, inadvertently or deliberately. In a social group where sharing is valued, many people, not just parents, will nod, smile, pat and praise when they see a child 'playing nicely'.

Of course, learning a propensity to share toys with other children is not the same thing as learning a generalised norm of sharing, and it is possible that evaluative learning, or something comparably sophisticated, is necessary for the learning of generalised norms. However, research with humble animals, rats and pigeons, suggests that the foundations of morality could be culturally inherited via unsophisticated, model-free reinforcement learning long before children can interpret the behaviour of others as communicative cues. If we set the bar too high for cultural learning – assume that it necessarily involves hard cognitive labour on the part of the child – we are likely to under-estimate the contributions of culture to moral development.

4. What to do?

I have argued that moral psychology should seek to discover the contributions of nature, nurture and culture to moral development, and that there are two major obstacles to this project: nativist bias and failure to distinguish nurture from culture. When culture is understood to be socially inherited information, rather than attributes that vary between social groups, culture contributes to moral development to the extent that novices learn *from*, rather than *about*, other people. In the previous section, I suggested that research on moral learning is well-placed to overcome these obstacles. Alert to the power of domain-general mechanisms of learning, and equipped with both the analytic resources of philosophy and the empirical methods of psychology, moral learners are more than capable of escaping nativist bias, thinking carefully about the difference between nurture and culture, and mounting an empirical programme that would ultimately give us a much clearer picture of how humans become moral animals. But what would this programme look like? What do I, an armchair psychologist, have the audacity to want the workers to do?

1) Less modelling and more data collection. It's good to know what domain-general learning could achieve in the moral domain, given certain kinds of input, but my sense is that the balance between modelling work, asking how morality could *possibly* develop, and empirical work, asking

how morality *actually* develops, is currently weighted towards the former. We need to know more about the inputs that children actually receive – the genetic predispositions and experiences that contribute to moral development – and about how they are processed, at the psychological level, in different contexts and phases of ontogeny.

2) Test model-based against model-free explanations. There is no reason to doubt that, at some point in development, children become capable of model-based thinking – for example, of modelling the behaviour of other agents with reference to mental states – or that model-based thinking could be a boon for moral development. But we also know that humans, at all stages of development, have other cognitive resources – model-free learning and other sub-personal processes – which we share with a broad range of other animals (e.g. Behrens et al. 2008). Consequently, more caution is needed in making generalisations about model-based learning. Unless 'model-based' is taken to be merely a style of explanation, applied as a matter of taste, rather than a type of explanation with real empirical bite, evidence that, for example, six-year-olds use model-based processes to find out about causality does not justify the assumption that 18-month-olds use model-based processes for imitation. Model-based explanations need to be tested against alternatives, inspired by research on model-free reinforcement learning and other cumulative work in experimental psychology, for each new period of development and functional context in which they are applied.

3) Test nativist hypotheses with an eye on both nurture and culture. Clearly, the early emergence of a competence is not sufficient reason to suppose that the competence depends on specific, genetically inherited information. Even half an hour after birth, a baby might cry in response to the sound of another infant crying because the baby has associated the sound of its own cries with feelings of distress. Before endorsing a nativist hypothesis we should look for evidence that the competence co-varies with specific genetic factors, and does not co-vary with experiential factors. In other words, we should look for "poverty of the stimulus" (Chomsky 1975) – signs that the environment does not contain enough information, attainable directly (nurture) or from other agents (culture), to support development of the competence. I'm not a fan of parsimony arguments (Heyes 1998; 2012), so I am *not* suggesting that empiricist hypotheses should be preferred because they are simpler or more tractable. Whether or not they have these virtues, an empiricist bias of this kind – or of any other kind - would be just as counter-productive as the nativist bias that currently afflicts moral psychology. All hypotheses – whether they emphasise the roles of nature, nurture or culture in moral development - need to be tested against plausible alternatives.

4) Study moral development on the ground. Formidable challenges face any attempt to document the kinds of experience that contribute to moral development under naturalistic

conditions. Observing children in their natural environments is expensive, demanding, and less-than-glamourous work; variation in child rearing practices over time, across cultures, and between social classes means that the work certainly cannot be done once and for all; and moral education is a focus of intensive political debate, making it difficult to secure funding and report the research without bias. But surely any research programme on "moral learning" must feed its models with reliable information about the input children actually receive. Fortunately, not all of this information has to come from observational studies of children in their home and school environments. For example, Grusec and her colleagues have developed a promising retrospective method, in which they ask adults about the contexts in which they learned important moral lessons (Grusec 2014; Grusec et al. 2006; Vinik et al. 2013). Memory is unreliable, but retrospective methods can be used alongside naturalistic experiments, allowing parent-child interactions to be analysed in the laboratory (e.g. Chalik and Rhodes 2015; Walker and Lombroso 2017), to triangulate on typical inputs to moral development.

Moral psychology, and especially "moral learning", has huge potential. I hope it will use some of its strategic momentum and intellectual resources to tell us more about the contributions of nature, nurture and culture to moral development.

Acknowledgements
I am grateful to Kim Sterelny, Caterina Dulles and two anonymous referees at *Synthese*, to Rich Cook for advice on configural face processing, John Pearce for guidance on partial reinforcement, and, especially, to Peter Railton for enriching conversation and detailed comments – many of them dissenting – on an earlier version of this manuscript.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. Psychological Science, 9, 321-324.

Baker, M. C. (2002). The atoms of language: The mind's hidden rules of grammar. Basic Books.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature*, *456*(7219), 245-249.

Bird, A. (2018). Understanding the replication crisis as a base rate fallacy. *The British Journal for the Philosophy of Science*.

Bloom, P. (2012). Moral nativism and moral psychology. The social psychology of morality: Exploring the causes of good and evil, 71-89.

Bloom, P. (2017). Against empathy. Bodley Head Limited.

Brownell, C. A., Ramani, G. B., & Zerwas, S. (2006). Becoming a social partner with peers: Cooperation and social understanding in one-and two-year-olds. Child Development, 77, 803-821.

Buckwalter, W., & Turri, J. (2014). Telling, showing and knowing: A unified theory of pedagogical norms. Analysis, *74*, 16-20.

Campbell, D. T. (1965) Variation and selective retention in socio-cultural evolution. In: Social change in developing areas: A reinterpretation of evolutionary theory, ed. H. R. Barringer, G. I. Glanksten & R. W. Mack, pp. 19–49. Schenkman.

Carder, B., & Berkowitz, K. (1970). Rats' preference for earned in comparison with free food. Science, 167, 1273-1274.
Cath, Y. (2019). Knowing what it is like and testimony. *Australasian Journal of Philosophy*, *97*, 105-120.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). Cultural Transmission and Evolution: a Quantitative Approach. Princeton: Princeton University Press.

Chalik, L., & Rhodes, M. (2015). The communication of naïve theories of the social world in parent–child conversation. Journal of Cognition and Development, 16, 719-741.

Chomsky, N. (1975). Reflections on Language. New York: Pantheon Books.

Chomsky, N. (1981). Lectures on government and binding: The Pisa lectures. Walter de Gruyter.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychological Review, 108, 204-256.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. The Adapted Mind: Evolutionary Psychology and the Generation of Culture, 163, 163-228.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. Cognition, 50, 41-77.

Crockett, M. J. (2013). Models of morality. Trends in Cognitive Sciences, 17, 363–366.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. Personality and Social Psychology Review, 17, 273-292.

Cushman, F., Kumar, V., & Railton, P. (2017). Moral learning: Psychological and philosophical perspectives. Cognition, 167, 1-10.

Darwin, C. (1874). The Descent of Man, and Selection in Relation to Sex: Reprinted from the Second English Edition, Revised and Augmented. Burt.

Debiec, J., & Olsson, A. (2017). Social fear learning: from animal models to human function. Trends in Cognitive Sciences, 21, 546-555.

Delamater, A. R. (1996). Effects of several extinction treatments upon the integrity of Pavlovian stimulus-outcome associations. Animal Learning & Behavior, 24, 437-449.

de Waal, F.B., Preston, S.D. (2017). Mammalian empathy: behavioural manifestations and neural basis. Nature Reviews Neuroscience, 18, 498–509.

Dwyer, S. (2006). How good is the linguistic analogy? In P. Carruthers, S. Laurence and S. Stich (eds), *The Innate Mind: Volume 2 Culture and Cognition*. Oxford: Oxford University Press, 237–255.

Englis, B.G., et al., 1982. Conditioning of counter-empathetic emotional responses. Journal of Experimental Social Psychology, 18, 375–391.

Fan, Y., et al. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. Neuroscience & Biobehavioral Reviews, 35, 903–911.

Fodor, J. A. (2001). The mind doesn't work that way: The scope and limits of computational psychology. Boston: MIT Press.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron, 66, 585-595.

Gonzalez-Liencres, C., et al. (2013). Towards a neuroscience of empathy: ontogeny, phylogeny, brain mechanisms, context and psychopathology. Neuroscience & Biobehavioral Reviews, 37, 1537–1548.

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (under review). Moral Foundations Theory: On the Advantages of Moral Pluralism Over Moral Monism. Retrieved June 2017 from: http://wwwbcf.usc.edu/~jessegra/papers/GHMMIM.MFT%20Atlas%20chapter.pdf

Graham, J., Waytz, A., Meindl, P., Iyer, R., & Young, L. (2017). Centripetal and centrifugal forces in the moral circle: Competing constraints on moral learning. Cognition, 167, 58-65.

Greene, J. D. (2015). The rise of moral cognition. Cognition, 135, 39-42.

Greene, J. D. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. Cognition, 167, 66-77.

Grusec, J. E., Chaparro, M. P., Johnston, M., & Sherman, A. (2006). The development of moral behavior and conscience from a socialization perspective. Handbook of Moral Development, 243-265.

Grusec, J.E. (2014). Parent-child conversations from the perspective of socialization theory. In C. Wainryb & H.E. Recchia (Eds.), Talking about right and wrong: Parent-child conversations as contexts for moral development (pp. 334–366). New York, NY: Cambridge University Press.

Haidt, J. (2012). The Righteous Mind: Why good people are divided by politics and religion. New York: Pantheon.

Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. Current Directions in Psychological Science, 22, 186-193.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. Nature, 450, 557.

Hamlin, J. K., Wynn, K., & Bloom, P. (2012). Reply to Scarf et al.: nuanced social evaluation: association doesn't compute. *Proceedings of the National Academy of Sciences*, *109*, E1427-E1427.

Hauser, M. (2006). Moral Minds: How Nature Designed our Universal Sense of Right and Wrong. Ecco/HarperCollins Publishers.

Heiphetz, L., & Young, L. L. (2017). Can only one person be right? The development of objectivism and social preferences regarding widely shared and controversial moral beliefs. Cognition, 167, 78-90.

Henrich, J. (2015). The Secret of Our Success. Princeton: Princeton University Press.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, *21*, 101-114.

Heyes, C. (2012). Simple minds: a qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 2695-2703.

Heyes, C. (2014a). False belief in infancy: a fresh look. Developmental Science, 17, 647-659.

Heyes, C. (2014b). Submentalizing: I am not really reading your mind. Perspectives on Psychological Science, 9, 131-143.

Heyes, C. M. & Frith, C. (2014) The cultural evolution of mind reading. Science, 344, 1243091.

Heyes, C. (2016). Born pupils? Natural pedagogy and cultural pedagogy. *Perspectives on Psychological Science*, *11*, 280-295.

Heyes, C. M. (2018a). Cognitive Gadgets: The Cultural Evolution of Thinking. Harvard University Press.

Heyes, C. M. (2018b). Empathy is not in our genes. Neuroscience & Biobehavioral Reviews, 95, 499-507.

Hinten, A. E., Labuschagne, L. G., Boden, H., & Scarf, D. (2018). Preschool children and young adults' preferences and expectations for helpers and hinderers. Infant and Child Development, e2093.

Ho, M. K., MacGlashan, J., Littman, M. L., & Cushman, F. (2017). Social is special: A normative framework for teaching with and learning from evaluative feedback. *Cognition*, *167*, 91-106.

Holvoet, C., Scola, C., Arciszewski, T., & Picard, D. (2016). Infants’ preference for prosocial behaviors: a literature review. Infant Behavior and Development, 45, 125-139.

Huang, C-T., Heyes, C. M. & Charman, T. (2006) Preschoolers' behavioural re-enactment of 'failed attempts': the roles of intention-reading, emulation and mimicry. Cognitive Development, 21, 36-45

Hume, D. (1751). An Enquiry Concerning the Principles of Morals. London: A. Millar*.*

Joyce, R. (2013). The many moral nativisms. In Cooperation and Its Evolution, eds. Sterelny, K., Joyce, R., Calcott, B., & Fraser, B. Boston: MIT Press. pp. 549-572.

Kleiman-Weiner, M., Saxe, R., & Tenenbaum, J. B. (2017). Learning a commonsense moral theory. Cognition, 167, 107-123.

Lamm, C. et al. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. NeuroImage 54, 2492–2502.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. Psychological Science, 15*,* 762-768.

Lewens, T. (2015). Cultural Evolution: Conceptual Challenges. Oxford: Oxford University Press.

Mill, J. S. (1861). Utilitarianism. Fraser's Magazine, November, pp. 525-534.

Mackintosh, N. J. (1974). The psychology of animal learning. Academic Press.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. Developmental Psychology, 31, 838.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. Psychological Science, 17, 608-615.

Murphy, J., Gray, K. L., & Cook, R. (2017). The composite face illusion. Psychonomic Bulletin & Review, 24, 245-261.

Moore, R. (2016). Gricean communication and cognitive development. *The Philosophical Quarterly*, *67*, 303-326.

Moore, R. (2017). Social cognition, Stag Hunts, and the evolution of language. *Biology & Philosophy*, *32*, 797-818.

Nichols, S. (2005). Innateness and moral psychology. In P. Carruthers and S. Laurence (eds), The Innate Mind: Structure and Content. New York: Oxford University Press.

Pinker, S. (2003). The language instinct: How the mind creates language. Penguin UK.

Railton, P. (2017). Moral Learning: Conceptual foundations and normative relevance. Cognition, 167, 172-190.

Railton, P. (2017). Moral Learning: Conceptual foundations and normative relevance. Cognition, 167, 172-190.

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., ... & Ross, C. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. Behavioral and Brain Sciences, 39, 1-68.

Rhodes, M., & Wellman, H. (2013). Constructing a new theory from old ideas and new evidence. Cognitive Science, 37, 592-604.

Rhodes, M., & Wellman, H. (2017). Moral learning as intuitive theory revision. Cognition, 167, 191-200.

Rozin P. (1988). Cultural approaches to human food preferences. In Nutritional Modulation of Neural Function, ed. JE Morley, MB Sterman, JH Walsh, pp. 137-153. San Diego: Academic Press.

Pearce, J. M. (2008). Animal learning and cognition: An introduction (3rd ed.). New York: Psychology Press.

Preston, S.D., De Waal, F.B. (2002). Empathy: its ultimate and proximate bases. Behavioral & Brain Sciences, 25, 1–20.

Prinz, J. J. (2014). Where Do Morals Come From?–A Plea for a Cultural Approach. In *Empirically Informed Ethics: Morality between Facts and Norms* (pp. 99-116). Springer, Cham.

Priva, U. C., & Austerweil, J. L. (2015). Analyzing the history of Cognition using topic models. Cognition, 135, 4-9.

Richerson, P. J., & Boyd, R. (2005). Not by genes alone. Chicago: University of Chicago Press.

Ruffman, T., et al., 2017. Do infants really experience emotional contagion? Child Development Perspectives, 11, 270–274.

Sabbagh, M. A., & Paulus, M. (2018). Replication studies of implicit false belief with infants and toddlers. Cognitive Development, 46, 1-3.

Scarf, D., Imuta, K., Colombo, M., & Hayne, H. (2012). Social evaluation or simple association? Simple associations may explain moral reasoning in infants. PloS One, 7, e42698.

Shea, N. (2012). Inherited representations are read in development. The British Journal for the Philosophy of Science, 64, 1-31.

Shettleworth, S. J. (1972). Constraints on learning. In *Advances in the Study of Behavior* (Vol. 4, pp. 1-68). Academic Press.

Simner, M.L., 1971. Newborn's response to the cry of another infant. Developmental Psychology, 5, 136–150.

Singer, P. (2015). The most good you can do: How effective altruism is changing ideas about living ethically. New Haven, CT: Yale University Press.

Sterelny, K. (2009). Peacekeeping in the culture wars. In Kevin Laland and Bennett Galef (eds) The Question of Animal Culture, Harvard University Press, pp 288-304.

Sterelny, K. (2010). Moral nativism: A sceptical response. Mind & Language, 25, 279-297.

Susilo, T., Crookes, K.,McKone, E.,& Turner, H. (2009). The composite task reveals stronger holistic processing in children than adults for child faces. PLoS One, 4, e6460.

Tolman, E. C. (1932 Purposive Behavior in Animals and Men. New York, Century Co. 463 pp.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural Learning. Behavioral and Brain Sciences, 16, 495-511.

Turiel, E. (2002). The culture of morality: Social development, context, and conflict.
Cambridge: Cambridge University Press.

Vinik, J., Johnston, M., Grusec, J. E., & Farrell, R. (2013). Understanding the learning of values using a domains-of-socialization framework. Journal of Moral Education, 42, 475-493.

Walker, C. M., & Lombrozo, T. (2017). Explaining the moral of the story. Cognition, 167, 266-281.

Warneken, F., Lohse, K., Melis, A. P., & Tomasello, M. (2011). Young children share the spoils after collaboration. Psychological Science, 22, 267-273.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. Science, 311, 1301-1303.

Wright, J. C., & Bartsch, K. (2008). Portraits of early moral sensibility in two children's everyday conversations. Merrill-Palmer Quarterly, 56–85 (1982-).

Youngblood, M., & Lahti, D. (2018). A bibliometric analysis of the interdisciplinary field of cultural evolution. *Palgrave Communications*, *4*, 120.

Zagzebski, L. (2013). Moral exemplars in theory and practice. *School Field*, *11*, 193-206.