How Children Learn the Meanings of Moral Words: Expressivist Semantics for Children[1]

Moral philosophy has long neglected children. This neglect has perhaps become most embarrassingly obvious in certain areas of first-order moral theorizing,[2] but has also left its mark on discussions of moral psychology and meta-ethics. One facet of this neglect, admirably highlighted by Joshua Gert,[3] is that debates over expressivism have ignored the question of how children could learn the meanings of moral words if expressivism were true. Moreover, this neglect has not been limited to philosophers. If experts in the field are correct then the question of how children learn moral words has been neglected by developmental psychologists too:

> A particularly interesting domain is the acquisition of moral terms like *fair* and *wrong*. An examination of how children learn such words would enlighten us not only about language development but also about the nature and development of moral thought. But with the exception of a fascinating discussion by Macnamara (1991), this domain of word learning has been ignored in the developmental literature.[4]

Gert argues that the prospects for an account of how children could learn an expressivist semantics for moral words are not encouraging. Very roughly, Gert argues that the fact that expressivism is not widely believed means that parents are not disposed to correct their children's speech when that speech is infelicitous according to expressivism. Without such correction, Gert argues that children could not learn expressivist meanings for moral words. Since children obviously do learn the meanings of moral words, Gert concludes that this is a reductio of expressivism. Moreover, Gert suggests that this problem is independent of the other problems facing expressivism and would remain even if those other problems were resolved. Gert extends this challenge to speaker

---

[1] Note removed to preserve anonymity.

[2] For example, Susan Moller Okin's argues that a Lockean theory of property rights like Nozick's seems to imply that women (or at least parents) own their children, a result that Okin argues is inconsistent with the Lockean theory itself. See chapter four of Okin's *Justice, Gender and the Family* (New York: Basic Books, 1989).

[3] Joshua Gert, "Expressivism and Language Learning," *Ethics* 112 (2002): 292-314. Further unmarked page references in the text are to this paper.

[4] Paul Bloom, *How Children Learn the Meanings of Words* (Cambridge, MA: MIT Press, 2000), pp. 262-263.

relativist theories, but I focus here on expressivism. The issues raised by this challenge are empirical ones. Although philosophers sometimes find this odd, it should not really be surprising. Expressivism is an account of the actual meanings of moral words, and the actual meanings of words in natural languages is of course an empirical matter.[5] Gert's challenge is important and promises to deepen our understanding of expressivism. Nonetheless, expressivists have ample resources with which to meet it.

I. WHAT ABOUT THE CHILDREN?

Expressivism is best understood as the thesis that moral utterances do not express beliefs guaranteed to provide the truth-conditions for those utterances, but rather serve to express certain non-cognitive attitudes. Some early expressivists maintained that the predicates 'is true' and 'is false' simply do not apply to moral utterances, ordinary usage notwithstanding. However, most expressivists now defend the view that moral utterances are truth-apt on a suitably minimalist account of the truth predicate. Very roughly, minimalist accounts of the truth predicate maintain that there is nothing more to truth than disquotation, so that to say that 'p' is true is just to say that p.[6]

---

[5] Expressivists could insulate their position from these empirical questions by instead defending the position that even if expressivism does not capture the actual meanings of moral words it provides an account of the meaning we should adopt for those words. However, this is not how expressivism is normally understood. Furthermore, such a revisionary theory would in some ways be less interesting and run the risk of simply changing the subject.

[6] Gert's characterization of expressivism differs from mine in one respect, though this will not matter to the arguments developed in the text. On Gert's account, expressivism also maintains that moral utterances do not express beliefs at all. Gert provides a (mostly implicit) argument for of this aspect of his characterization when he remarks that, "if normative claims expressed the attitude of belief, then such claims could themselves be true or false also." (292) Gert explicitly notes that some expressivists are willing to allow that moral utterances are truth-apt in a minimalist sense, so his point here must really be the more subtle one that if normative claims expressed beliefs then they could be true in just the same way as factual claims, and the distinction between expressivism and cognitivism would then seem to evaporate. However, this line of argument misfires for two reasons. First, the mere fact an utterance expresses a belief does not even entail that the utterance is truth-apt. Questions and commands illustrate the point; neither are plausibly truth-apt but both can express beliefs. "Have you stopped smoking?" and "Shut the door" express the beliefs that one's interlocutor has smoked in the past and that the contextually salient door is open but neither utterance is plausibly truth-apt. Second, expressivists who allow that moral utterances are truth-apt in a minimalist sense could allow that those utterances conventionally express certain beliefs but deny that those beliefs provide the truth-conditions for moral utterances. For further discussion see Michael Ridge, "A Solution For Expressivists," Brown Electronic Article Review Service (BEARS), J. Dreier and D. Estlund, eds., 10/5/99, http://www.brown.edu/Departments/Philosophy/bears/9910ridg.htm.

On any account, expressivism is a theory about the meanings of moral words, so the challenge to explain how children could learn an expressivist moral semantics is a fair one. Gert's argument begins by assuming for reductio that expressivism is true. The first premise of the argument is that even if expressivism is true then it is not believed to be true by most people. In itself, this is not a serious objection to expressivism (nor does Gert suggest it is), for understanding the meaning of a word is better explicated in terms of knowing how to use the word properly rather than knowing that it means such-and-such. Expressivism might be a good theory of the sort of know-how associated with moral vocabulary even ordinary speakers do not have knowledge-that it is. Gert argues that because most people do not believe expressivism most parents do not teach their children to call something 'morally wrong' (or its equivalent) in all and only circumstances in which the child disapproves of the action in the relevant way.[7] Rather, most parents correct their children's use of 'morally wrong' only when the child says something is morally wrong the parent takes not to be morally wrong.[8] This suggests that the child will internalize a rule along the lines of, 'apply the predicate 'morally wrong' to actions only if the action satisfies such-and-such descriptive criteria' where the relevant descriptive criteria are a reflection of his parents' moral outlook. This suggests that children learn to use 'morally wrong' not to express disapproval but instead to express beliefs as to which actions satisfy the relevant descriptive criteria.

The final premise of the argument is that ostension and correction are essential to language learning (296; 307-309). Gert realizes that this premise is both controversial and essential to his argument. On a more Augustinian conception of language learning,

[7] For expository reasons I shall henceforth drop the 'in the relevant way' caveat and simply use 'disapproves' to refer to whatever sort of non-cognitive attitudes turn out to be the most plausible candidates for being the moral attitudes expressed when someone says something is wrong.
[8] Moreover, Gert contends, depending on the complexity of the expressivist account, it may be very difficult to determine when a child has the relevant attitude(s) of disapproval. So even those few parents who do accept expressivism might not be able to correct their children when they say something is wrong without disapproving of it reliably enough to help the child learn the meanings of moral words. Even A.J. Ayer might be unable to convey an expressivist semantics to his children.

children can sometimes learn the meaning of a word by observing the linguistic behavior of those around them without being corrected by them.[9] Wittgenstein famously ridiculed St. Augustine's account of how children learn the meanings of words, and Gert favors what he calls the 'Wittgensteinian teaching and testing based on publicly observable criteria' (308) of language learning. For expository simplicity I shall stick with Gert's terminology but I do not mean to suggest that Wittgenstein actually held the view sort of view Gert defends. Gert gives three reasons for the more Wittgentsteinian view. First, he argues that it can better explain why expressivism is uncontroversial for words like 'funny' and 'disgusting' but controversial for moral vocabulary. In the case of words like 'funny' and 'disgusting' Gert contends that it is obvious that these words have an expressive meaning precisely because in those cases parents *do* correct their children when they say something is funny but do not find it amusing or call something disgusting when it does not elicit their disgust. Gert suggests that a more Augustinian conception of language learning makes this asymmetry obscure. Second, Gert argues that the complex sorts of moral attitudes in which expressivists are interested are too subtle for young children reliably to detect (308 and 312). Third, Gert argues that the Augustinian conception does not fit well with the empirical evidence. Some studies indicate that children may need to be corrected in their use of personal pronouns, as otherwise they sometimes come to understand these pronouns as proper names, using 'you' to refer to themselves and 'I' to refer to others. Other research suggests that, "the kind of perspective-taking that is presupposed by the speech-role hypothesis comes with, rather

[9] In a often quoted passage, St. Augustine writes (translated from the Latin),

> When they (my elders) named some object, and accordingly moved towards something, I saw this and grasped that the thing was called by the sound they uttered when they meant to point it out. Their intention was shewn by their bodily movements, as it were the natural language of all peoples: the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of voice which expresses our state of mind in seeking, having, rejecting or avoiding something. Thus, as I heard words repeatedly used in their proper places in various sentences, I gradually learned to understand what objects they signified; and after I trained my mouth to form these signs, I used them to express my own desires.

The Confessions of St. Augustine (New York, Random House: 398/1961), p.11.

than precedes the acquisition of personal pronouns." (308, n. 19) The speech-role hypothesis Gert attacks here is an Augustinian one according to which children learn the meanings of pronouns by observing others to see to what those pronouns refer from one context to the next and then working out the function that takes one from a context to a referent for each pronoun. Because these points concern indexicals they are more directly relevant to Gert's extension of his argument to speaker relativist theories. Nonetheless, this evidence does cast some doubt on Augustinian conceptions more generally. More directly relevant here is Gert's observation that in one of the very few papers devoted to the question of how children learn the meanings of moral vocabulary, John MacNamara suggests that "there are descriptive predicates in the language of thought that get labelled 'good' and 'bad' in English through the normal processes of language learning." (308, n. 19) Gert concedes that MacNamara's argument begs important question in his favor but insists that we should take seriously the views of those familiar with the relevant empirical data.

To summarize, Gert's argument goes roughly as follows:

(1) Expressivism is true [assumed for reductio]
(2) Children somehow learn an expressivist semantics for moral words [from (1) plus the very plausible assumption that children do learn the meanings of moral words]
(3) Children could learn an expressivist semantics for moral words in only one of two ways, either in part by being corrected when their speech violates the rules of expressivist semantics (the Wittgensteinian account) or without such correction, simply by observing the speech of those around them (the Augustinian account). [plausibly an exhaustive survey of all of the prima facie plausible available explanations]
(4) Children are not corrected when their speech violates the rules of expressivist semantics. [for their instructors, their parents, do not believe expressivism even if it is the correct theory of what their words mean]
(5) Simply observing the speech of those around them would not allow children to learn an expressivist semantics [for (a) this theory could not explain the asymmetry between expressivism about 'funny', 'disgusting' on the one hand and moral predicates on the other, (b) the moral attitudes are too subtle for young children to detect, and (c) the available empirical evidence suggests that such Augustinian conceptions are mistaken.
(6) So, children do not learn an expressivist semantics. [from (3), (4), and (5)]
(7) So expressivism is false [from (1), (2) and (6) by reductio]

My defense of expressivism rejects premise (5) of this argument, although this is not the only promising line of reply. One might instead reject (3) if the independent case for expressivism is strong enough apart from the issue of language learning. For in that case one might insist that since we know this is what people mean and have meant by 'morally wrong' through the years there must be *some* way in which children learn an expressivist semantics for moral terms. The fact (if it were a fact) that the two mechanisms enumerated in premise (3) cannot explain how children do so might just be taken as evidence that there must be some other way even if we at this point lack the imagination to see what it might be.[10] Alternatively, one might challenge (4) on the grounds that parents take moral utterances to be insincere (as opposed to false) when children fail to have the relevant attitudes. Rather than correcting the child by saying that what they have said is false, in such cases parents might well tell the child that they do not mean what they have said. Nonetheless, Gert is right that a more Augustinian model of language learning is a very tempting strategy for the expressivist and it is worth seeing whether it could do the job.

II. AUGUSTINIAN THEORIES, DISCIPLINE AND THE GOLDEN RULE

The empirical evidence that children can learn language through observation of adult conversation without much or anything in the way of linguistic correction is impressive. In some cultures, parents do not try to teach their children how to talk for a variety of reasons, but these children still manage to learn the meanings of words. For example, in some cultures and subcultures (like Trackton, a rural African-American community in the Piedmont Carolinas) people think that children cannot be taught language but instead must learn for themselves. Further, empirical evidence indicates that adults in these cultures really do not speak to children very much until they are producing multi-word

utterances. Nonetheless, these children seem to have no trouble becoming competent speakers.[11] Further, severely neglected children manage to become fluid in their native language, which strongly suggests that childhood correction is not essential.[12] Paul Bloom summarizes this trend, "Augustine's proposal is no longer seen as the goofy idea that it once was. Increasing evidence shows some capacity to understand the minds of others may be present in babies before they begin to speak."[13]

One might worry about Augustinian accounts on more philosophical grounds, empirical evidence notwithstanding. Wittgenstein's famous critique of the Augustinian model simply presupposes that there is no innate "language of thought" in which a child could use to learn a natural language. However, that presupposition has since come under considerable philosophical fire, most notably in the work of Jerry Fodor and his followers.[14] More importantly, one does not need to accept a Fodorian computational account of a language of thought to welcome the more general and intuitive idea that non-linguistic children and nonhuman animals nonetheless are capable of thought in an important sense. More anti-realist accounts of folk psychology, like Daniel Dennett's "intentional stance" account, also make room for the possibility of beliefs without natural language.[15] Moreover, even if one's initial vocabulary must be learned in a non-Augustinian way, that vocabulary could allow a child to build a further vocabulary in a more Augustinian way on that initial Wittgensteinian scaffolding.

In any event, the empirical evidence shows that children somehow manage to learn a language even when they must rely on observation of the overheard conversations between adults to do so. Since what is actual must be possible, any

---

[10] Note removed to preserve anonymity.
[11] See Elena V.M. Lieven, "Crosslinguistic and Crosscultural Aspects of Language Addressed to Children," in *Input and Interaction in Language Acquisition*, Claire Gallaway and Brian Richards (eds.) (Cambridge: Cambridge University Press, 1994), p. 60. Lieven discusses a number of examples of this phenomenon.
[12] See Bloom, p. 83.
[13] Bloom, p. 61.
[14] See especially Fodor's classic *The Language of Thought* (New York: Thomas Crowell Company, 1975).
[15] See, e.g., Daniel Dennett, *The Intentional Stance* (MIT Press: Cambridge, 1993).

philosophical argument that this is impossible must go wrong somehow. So children could learn by observing such conversations that people typically call something morally wrong only if they disapprove of it even if their parents did not correct them when they said something was wrong but did not disapprove of it. However, in the case of moral vocabulary, children are not nearly so reliant on passive observation. Most notably, children are regularly told by their parents when they do something morally wrong, are thereby discouraged from doing it again and sometimes punished for their transgression. If expressivism is correct (as Gert assumes for reductio) then when parents scold their children for acting morally wrongly they typically are expressing their disapproval. Moreover, common sense supports the expressivist on this score. It is a familiar point that parents are often exasperated by the exhausting and sometimes frustrating work of raising children, and hence are prone to transparent displays of emotion when their children act wrongly. Children typically find expressions of parental anger or disappointment unpleasant, so this is a sensible strategy for modifying their behavior.[16]

So unlike the more general case of Augustinian language learning, in which the young child must passively observe conversations between adults, the case of moral language should be much easier. In the moral case, the parents often are talking directly to the child, telling that her that she has done something morally wrong (or 'bad' or 'naughty') and displaying a suitable attitude of disapproval. Moreover, in many cases this directed expression of emotion is intense; even the most self-controlled parents sometimes become exasperated enough to display not merely stern firmness but anger, disappointment and frustration. If the passive observation of conversations between adults is enough for children to learn language (as cross-cultural evidence and severely

[16] One reason this is so effective may be that even very young children involuntarily mimic the facial expressions of those around them, and empirical evidence suggests that forming the facial expression associated with an unpleasant emotion can trigger a feedback loop so that the child feels that unpleasant emotion too. See Martin L. Hoffman, *Empathy and Moral Development* (Cambridge: Cambridge University Press, 2000), esp. pp. 36-44. Older children presumably also fear punishment associated with parental

neglected children suggest) then such intense and regular expressions of emotions by parents at children should be more than enough for the child to learn that one calls something 'morally wrong' (or 'bad' or 'naughty') only if one disapproves of it. The case is bolstered even more by the enormous frequency of these kinds of exchanges. By the end of the second year, discipline encounters involving the child's' harming another are very common though there is no clear data on how common.[17] In any event, by the end of the second year, fully two-thirds of parent-child interactions are discipline encounters in which parents attempt to change children's behavior against their will.[18] Furthermore, children are able to seek out and detect their parents' emotions as early as four months and modify their behavior in ways appropriate to those emotions. For example, young infants look to their mothers before crossing the deep side of a visual cliff tend to cross only if their mother smiles and almost never cross if the mother's face shows fear. The evidence also suggests that other emotions, including anger, can be detected by young children.[19] So it would not be too hard for children to determine one calls something 'morally wrong' only if one disapproves of it.

If there ever was a good case for Augustinian learning then this would seem to be it. Reflection on discipline encounters reveals that the entirely passive classical model of Augustinian learning attacked by Wittgenstein and others on the one hand, and a model that relies on active parental correction of linguistic mistakes on the other is a false dichotomy. The actual case is one in which an enormous amount of emotionally charged moral discourse is aimed directly at the child to correct moral mistakes, as opposed to linguistic mistakes. Linguistic mistakes are generally corrected too, but quite apart from

---

disapproval and also fear potential love withdrawal given their dependence on their parents. See Hoffman, 148-150 for a discussion of how fear of love withdrawal might play a role in a child's moral development.

[17] See Hoffman, p. 141.

[18] See Hoffman, p. 141.

[19] See Klinnert, et. Al., "Emotions as Behavior Regulators: Social Referencing in Infancy," in *Emotion: Theory, Research and Experience,* vol. 2, Robert Plutchik and Henry Kellerman (eds.) (Academic Press, New York: 1983) 57-86, p. 73.

that linguistic correction, a lot more is going on here than in the standard paradigm of Augustinian learning through passive observation of conversations between adults.

Of course, not all discipline encounters concern distinctively moral issues, so this account needs to be supplemented to explain how children come to distinguish moral infractions from violations of norms of etiquette, custom and the like. As children mature, parents usually start to tell them not merely that they have done something morally wrong, but explain why it was morally wrong. In the case of distinctively moral infractions, these explanations very often take the form of appeals to something like the Golden Rule or other attempts to get the child to imagine how she would feel if someone treated her in the way she has treated someone else.[20] Quite often, the point of this exercise is to get the child to see that if someone treated her in the way that she has just treated another then the child would be angry at that person. Because human beings have some natural tendency to empathize with others, this can function to get children to feel what Martin Hoffman has called "empathic anger" on behalf of their victim directed at themselves.[21] I cannot do justice to Hoffman's account here, but he makes a strong case that empathic anger is an important source of moral motivation. So this sort of instruction could help children learn to associate the use of 'morally wrong' not just with anger in general but specifically with empathic anger on behalf of a victim. In this way, children's initially coarse-grained appreciation of the ways in which 'morally wrong' is used to express disapproval could become more refined as they age, and they might thereby distinguish moral infractions from transgressions of non-moral norms.

None of this undermines Gert's argument that children will also learn to associate a particular descriptive meaning with 'morally wrong' on the basis of their parents' correcting them when they call something morally wrong that the parents take

[20] Hoffman refers to these attempts as inductions, and some evidence suggests that middle-class parents in the US rely heavily on such inductions, though there are socio-economic differences here. See Hoffman, p. 161, n. 4.

not to be morally wrong. Indeed, this seems almost certainly to be true in the majority of cases. However, this is not in itself a problem for expressivism. Children may well come to associate the descriptive criteria for wrongness implicit in their parents' moral outlook with uses of 'morally wrong' but it does not follow from this that those descriptive criteria fix the meaning of 'morally wrong'. Perhaps children come to internalize two sorts of rules to govern their use of 'is morally wrong':

(1) Do not call something 'morally wrong' unless you disapprove of it.

(2) Do not call something 'morally wrong' unless it has such-and-such descriptive features.

This much is compatible with expressivism so long as the first of these two rules in some way is primary, and so long as the widespread acceptance of the first rule does not entail that 'morally wrong' simply means 'is disapproved of by me'. On an expressivist account the first rule should be understood as capturing the sincerity conditions for moral utterances rather than their truth-conditions. Just as a speaker should not assert a factual proposition p unless she believes that she believes that p (on pain of insincerity), a speaker should not assert an evaluative proposition q unless she believes she has the appropriate attitude(s) (on pain of insincerity). In neither case is it plausible to suppose that the sincerity conditions provide truth-conditions (the truth-condition for a factual proposition 'p' is not that the speaker believes that p). However, a detailed discussion of the distinction between truth-conditions and sincerity conditions raises much more general issues about how to distinguish expressivism from subjectivism, and those issues remain even if we restrict our attention to the use of moral predicates by adults. The aim here is to see whether language learning poses any new problems,[22] so the issue is in what sense the first of these rules might be primary to the meaning of 'morally wrong'.

---

[21] See Hoffman, pp. 96-102.

[22] For useful discussion, see Ridge, previously cited as well as Michael Smith and Daniel Stoljar, "Is There a Lockean Argument Against Expressivism," forthcoming in *Analysis*, both of which are replies to Frank Jackson and Philip Pettit's "A Problem for Expressivists," *Analysis* 58 (1998) 239-251.

Historically expressivists have allowed that particular speakers may associate a descriptive and expressive meaning with moral words and argued that the latter is primary.[23]

There are two senses in which it is plausible that the expressive criteria a child learns to associate with moral words are central to the meaning of those words, and both rely on the expressivist idea of moral disagreement understood as disagreement in attitude.[24] First, as a child matures she may come to realize that many people do not share her parents' moral values but nonetheless use moral vocabulary without being accused of linguistic confusion. These disagreements might in some cases rely on agreement at a deeper level, as when two utilitarians agree about the criterion of right action but disagree as to whether a particular action satisfies that criterion. However, the child will come realize that sometimes moral disagreement goes much deeper than this, and reflects fundamentally different values. Nonetheless, moral conversations between such people do not degenerate into accusations of linguistic confusion. Some might argue that this sort of deep moral disagreement is actually very rare,[25] but if that were true then it would be a problem for expressivists quite apart from the issue of language learning. Again, the aim here is to see whether language learning poses any new difficulties. Moreover, documenting the extent of actual moral disagreement that transcends factual disagreement would go well beyond the scope of this paper. In any event, as the child comes to appreciate that competent speakers call actions 'morally wrong' without linguistic rebuke even though the speaker knows those actions do not have the descriptive features the child has internalized from her parents, she might realize (implicitly) that those descriptive criteria are not required for linguistic

[23] R.M. Hare is a nice case. See R.M. Hare's discussion of descriptive and evaluative meanings in his *The Language of Morals* (New York: Oxford University Press, 1964).
[24] An idea pioneered in philosophy by Charles Stevenson. See his *Ethics and Language.* (New Haven, Yale University Press: 1944).
[25] This is suggested by Gert's discussion of vagueness as explaining how people could agree on all the facts but disagree about morality; see Gert, 298, fn. 9. The general issue of whether deep moral disagreement is best understood in terms of vagueness raises difficult issues that would require a paper of their own, so it is

competence with 'morally wrong'. Indeed, the child's own moral values might dramatically change if she rebels against her parents. She might well persist in calling actions morally wrong just in case she disapproves of them, but she will no longer rely on her parents' descriptive criteria. Even if the child does not reject her parents' values, the tacit realization that she could do so without linguistic confusion means that from the child's point of view the expressive meaning of 'morally wrong' is primary. The fact that so many people share the Moorean intuition that for any descriptive characterization of an action, it always remains at least a conceptually open question whether it is morally wrong provides further support for thinking people implicitly come to see the expressive meaning of moral words as primary.[26]

There is a second sense in which the expressive meaning of moral words is primary. In spite of the preceding line of argument, perhaps some children do not come, with maturity, to realize (implicitly) that they could shift the descriptive criteria they associate with 'morally wrong' without linguistic confusion. Nonetheless, the very fact that different people associate different descriptive criteria with the use of 'morally wrong' and its cognates without thereby being thought to be guilty of conceptual confusion suggests that no such descriptive criteria is central to the meaning of 'morally wrong'. For any given speaker, a particular descriptive criteria might reflect what the speaker means to communicate in calling something 'morally wrong' but we must distinguish between what particular speakers may mean to convey with their words on a particular occasion from what those words mean. The former is a highly pragmatic issue that depends on peculiarities of each context of utterance; the latter is a function of more firmly entrenched and widely shared conventional practices. By contrast, if expressivism

---

dialectically convenient that this is an issue expressivists must eventually face quite apart from language learning.

[26] See G.E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903). Of course, Moore himself inferred that moral words refer to irreducible non-natural properties but there are further

faces no other problems then what all competent speakers use moral discourse (when being sincere) to express their attitudes of approval and disapproval. So the ubiquity of moral disagreement means that any particular descriptive criteria looks like a poor candidate for fixing the meaning of 'morally wrong'. By contrast, if expressivism is correct then the rule that a speaker calls something 'morally wrong' only if she disapproves of it is widely shared in spite of the ubiquity of moral disagreement.[27] On the Augustinian account developed here, facts about language learning do nothing to overturn the expressivist position on this score.

Here it is important to remember that the sense of 'express' in 'expressivism' is such that an utterance can express an attitude not actually had by its speaker. The analogy with belief is helpful here; an utterance of 'the duck-billed platypus is a mammal' in this sense expresses the belief that the duck-billed platypus is a mammal even if the speaker lacks that belief. This utterance serves to express this belief only because of larger conventional practices that rationally lead competent speakers to infer from the fact that someone makes such an utterance that if she is being sincere then she does believe the duck-billed platypus is a mammal. The analogous expressivist claim is that background practices lead competent speakers rationally to infer from the fact that a speaker says something is morally wrong that if she is being sincere then she disapproves of it. The point here is that on the Augustinian account all competent speakers accept the rule, 'do not call something morally wrong unless you disapprove of it' and presumably they also come to realize (again, implicitly) that others accept this rule as well.

---

independent reasons for supposing that ordinary speakers and their practices do not support this interpretation.

[27] Indeed, someone who called something 'wrong' but did not take themselves to disapprove of it in any way plausibly is guilty of a linguistic infelicity, for there is something insincere about calling an action 'wrong' when one does not disapprove of it. This is reflected in common sense by such truisms as "actions speak louder than words," and the way in which we infer that someone does not really believe something is wrong when they show no sign whatsoever of being disposed not to perform such actions. Again, these claims are controversial, and I am forced to oversimplify the opposing terrain dramatically but once again these are claims the expressivist must defend in any event. If those claims are defensible then

This, in turn, makes it likely that background practices needed to underwrite inferences like, "She said war was morally wrong so if she is sincere then she disapproves of war" are in place. By contrast, a cognitivist must hold that background practices license inferences like, "She said war was morally wrong so she must think war has such-and-such properties." Given widespread and deep moral disagreement, it is unlikely that the relevant conventions underwrite such inferences.[28] On an Augustinian account, children learn to associate a descriptive and an expressivist rule with 'morally wrong'. The expressivist hypothesis is that deep moral disagreement means that only the expressive meaning is widely shared enough to fix the meaning of 'morally wrong'.

At this point we should return to Gert's objections to Augustinian accounts. His first objection is that Augustinian accounts could not explain why expressivism is so obviously true for words like 'funny', 'disgusting' but at best highly controversial in the case of moral words. However, the very fact that Augustinian accounts of language learning are *themselves* more controversial than models that rely on parental correction of the child's utterances could explain this asymmetry. Insofar as Gert's argument highlights the ways in which the expressivist must rely on an Augustinian conception of language learning, moral expressivism might simply inherit the controversy surrounding Augustinian conceptions. Nor is it clear why the ability to explain this asymmetry is such a weighty desiderata. Considerations of how language is learned are only one of a wide range of considerations one might invoke to explain why expressivism is more obvious for one vocabulary than another.

Gert's second worry is that the attitudes expressed by moral utterances according to expressivism are too subtle for young children to recognize. Gert's worry here does not seem to be based on any particular empirical evidence about children's abilities, but it

---

they can be used to explain how the facts about how a child learns the meaning of moral words poses no new problem for the expressivist. Otherwise, expressivism is doomed quite apart from language learning.

is worth considering. To some degree this worry is impossible to address without a more detailed account of the version of expressivism on offer. Nonetheless, the account developed here implicitly provides a recipe that any independently plausible version of expressivism could use to meet the worry. Initially, the child might simply learn to associate uses of 'morally wrong' with expressions of anger or displeasure in some very general sense. However, a child's grasp of the meaning of moral words can evolve as the child matures, just as a child's appreciation of other sorts of words can and does evolve as the child matures.[29] In particular, children's understanding of the negative attitudes expressed by moral utterances may become more refined. They might, for example, come to associate 'morally wrong' not merely with anger but with the sort of empathic anger Hoffman describes. Alternatively, they might eventually realize (implicitly) that uses of 'morally wrong' serve to express the acceptance of the kinds of norms Allan Gibbard has discussed in some detail.[30] Indeed, on Gibbard's account there may even be an evolutionary basis for supposing human beings eventually acquire the ability to recognize the attitude he refers to as 'norm acceptance' though this particular account is controversial. Again, I do not mean to tie the account of language learning developed here to any particular version of expressivism. The more general point is that young children can initially learn to associate perhaps crudely understood forms of disapproval with uses of 'morally wrong' and later refine their understanding considerably.

Gert's third objection is that Augustinian accounts do not fit well with the empirical evidence. This concern has also been addressed to some degree already, as I have adverted to further empirical evidence that supports the Augustinian conception in general and the unique plausibility of such conceptions in the moral case. However,

---

[28] Here I put to one side the possibility that 'wrong' just refers to a sui generis property of some kind, but I return to this suggestion when discussing MacNamara's position; see below.
[29] For a discussion of the ways in which children's understanding of emotions changes as the child matures, see Klinnert, et. al.
[30] See Gibbard, *Wise Choices, Apt Feelings* (Oxford: Oxford University Press, 1990).

Gert's worries should be met on their own terms. Most of the evidence Gert cites concerns personal pronouns, and hence is more directly relevant to his critique of speaker relativism as opposed to expressivism. The main point here seems to be that some children begin by making the mistake of thinking 'I' and 'you' are proper names and refer to themselves as 'you' and others as 'I'. Unless these confusions are corrected such children have great difficulty understanding the real meanings of indexicals. So at least in these cases, the pure Augustinian conception of language learning through passive observation breaks down. However, these kinds of mistakes are actually very rare, as Gert himself allows, which in itself is evidence in favor of an Augustinian account. Insofar as most children somehow manage to pick up the meanings of indexicals without ever needing to be corrected in the first place we have reason to think that such correction is not as essential as Gert argues. Furthermore, autistic children are especially likely to make these kinds of mistakes with indexicals, which is just what the Augustinian conception would predict.[31] For autistic children have a great deal of trouble inferring the mental states of others from their behavior, facial cues, etc., and Augustinian accounts rely heavily on children's native ability to make such inferences. As Bloom puts it, on an Augustinian account, "children use their *naïve psychology* or *theory of mind* to figure out what people are referring to when they use words. Word learning is a species of…mind reading."[32] So the fact that autistic children have more trouble learning the meanings of indexicals serves to confirm rather than undermine Augustinian theories of language learning. Not all children who have trouble learning indexicals are autisitc; blind children also have considerable trouble with them. However, some research suggests that this is because visual coorientation may be an important precursor to the development of a theory of mind, in which case blind children would have trouble for

[31] See Bloom, p. 125.
[32] Bloom, p. 61.

the same kind of reason that autistic children have trouble.[33] This fits well with Augustinian theories as well. Alternatively, blind children may have trouble with indexicals simply because it is harder to tell who is speaking to whom, and this would present a problem for the blind child on any plausible account, Augustinian or otherwise.[34] Finally, some empirical evidence as to how children learn the meanings of indexicals strongly supports the Augustinian conception in a much more direct way. Yuriko Oshima-Takane's research is especially salient here. In a study in which some children were directly addressed by their parents and others merely overheard parents talking to one another,

> …the most striking result was that not a single child in the addressee condition showed the correct imitation pattern during the modelling sessions. All four children who showed the correct imitation pattern were in the non-addressee condition...Since no children in the addressee condition showed the correct imitation pattern, the correct imitation by the four children in the non-addressee condition lends support to the hypothesis that systematic opportunities for observing the use of *me/you* in speech not addressed to children facilitate correct imitation.[35]

So far from speech addressed to children (including corrections) being necessary for their learning the meanings of indexicals, such speech seems *counterproductive*, given this evidence. Admittedly, some of the other evidence here cited by Gert seems to point in the other direction. However, one of the main studies Gert cites relies primarily on French children and the authors note that the complexity of French pronouns as opposed to English pronouns may account for some of their data. More importantly, even in this study very little confusion was observed for non-addressed uses of first

---

[33] See R.P. Hobson, "Through Feeling and Sight to Self and Symbol," in U. Neisser (ed.) *The Perceived Self: Ecological and Interpersonal Knowledge of the Self* (New York: Cambridge University Press, 1994), cited in Bloom, p. 126.
[34] See Bloom, p. 126.
[35] Yuriko Oshima Takane, "Children Learn from Speech Not Addressed to Them: the Case of Personal Pronouns," *Journal of Child Language* 15 (1988) 95-108, pp. 105-106.

personal pronouns like 'I' and 'my' and a grasp of these pronouns arguably is most essential to the speaker-relativist and expressivist accounts.[36]

Finally, Gert appeals to MacNamara's discussion, according to which children learn to attach 'good' and 'bad' to descriptive predicates in their language of thought. Actually, it is dialectically bizarre (or at least ironic) to appeal to an account like MacNamara's in the context of a critique of Augustinian theories of language learning. For Augustinian theories were ridiculed by Wittgenstein and others precisely because they presupposed a language of thought. Moreover, Gert is certainly right when he concedes that MacNamara's theory "has a descriptivist bias that begs some important questions" (308, n. 19) in Gert's favor. Gert suggests that we should pay heed to what MacNamara has to say anyway because we should "take seriously the views of those who are actually familiar with the relevant empirical data" (308). However, we should not defer to such experts when the basis for their position is largely philosophical rather than empirical, and MacNamara's main argument here is entirely philosophical. MacNamara's argument for his position begins with a nod to G.E. Moore, "Besides, the set of good actions cannot be defined in terms of some non moral area of discourse – see Moore (1903, chap 1)."[37] Simply assuming that 'good' and other moral predicates refer to a property, MacNamara then asks how children could learn the meaning of such words. Since moral properties are irreducible he argues that the only way children could learn such words would be if they had a primitive word for them in their language of thought:

> Neither is the word 'good' ('bad') definable in terms of pleasure or utility or any expressions more basic than itself, as Moore (1903) shows in the first chapter of *Principia Ethica*. It follows that it must be unlearned. Not that the English word 'good' ('bad') is unlearned but that some equivalent predicate in the language of thought must be.[38]

---

[36] Pascale Girouard, Marcelle Ricard and Therese Goin Decarie, "The Acquisition of Personal Pronouns in French-Speaking and English-speaking Children," *Journal of Child Language* 24 (1997) 311-326, p. 322.
[37] John MacNamara, "The Development of Moral Reasoning and the Foundations of Geometry," *Journal for the Theory of Social Behaviour* 21 (1991): 125-150, p. 125.
[38] MacNamara, p. 143.

This argument is purely philosophical, and MacNamara's impressive grasp of the relevant empirical data therefore does not bolster it. The argument is an elimination argument, but falls prey to a classical problem for such arguments by not considering all relevant possibilities. In particular, the argument ignores the expressivist possibility that moral words serve primarily to express non-cognitive states rather than to refer to properties. This omission is serious, particularly given MacNamara's emphasis of the Open Question Argument. Expressivists historically gained at least as much mileage from that argument as non-naturalists. Indeed, Moore himself eventually conceded that he could not see how his argument supported non-naturalism better than expressivism.[39] The history of 20th century meta-ethics began with Moore's *Principia* but did not end there.

Conclusion.

Gert's challenge is a welcome reminder of the importance of children to any plausible meta-ethical theory. It also highlights new areas for further interdisciplinary work between philosophers, linguists and developmental psychologists. Even if Gert's challenge can be met in the way I have argued it can, it raises important and neglected issues worthy of further consideration.

[39] See his "A Reply to My Critics," in P.A. Schlipp (ed.) 1942. *The Philosophy of G.E. Moore.* The Library of Living Philosophers (Northwestern University, Evanston: 1942)