The 3 I's

One way of modelling how the human brain works is as the interaction of three modes of response: "Instinct", "Intuition" and "Intellect" – the 3 I's. (This somewhat resembles Plato's allegory of the soul in terms of a charioteer guiding two horses, but the resemblance is by no means complete – the control is not just one-way, for example, as it is in Plato's picture.) The thesis advanced here is that human *moral* responses cannot be attributed exclusively to one of these three modes, but are an example of their interaction.

The terms themselves need clarifying, as they are not always used in the sense I intend here.

- **Instinct** By "Instinct" I mean those innate responses which are determined genetically and shared with some or many other species. In fact, it is often easier to see them in other species since in the human they rarely occur without modification by the other two I's. The loyalty to the pack leader of social animals, the care for the newborn in animals whose newborn young are not fully equipped to fend on their own, hostility to competitors for resources, dutiful performance of allotted roles in highly organized groups: these are just some of the "instincts" that appear to be "hard-wired" in many species. The term "drive" is more or less synonymous. The sympathetic recognition of distress in fellow members of some species, elephants being a notable case in point, also appears to be implicit in the genetic design – "mirror neurones" being one of the mechanisms employed in our own species at least. Neurologically, instinctive responses are typically associated most closely with the "lower", older, hence more universally shared parts of the mammal brain, though with different details according to the species. They may well be implemented through neurochemical means – adrenalin or seratonin, for example.

  Psychologically, there is much overlap between "instinct" and what we describe, particularly in the human context, as "emotions". In fact, in cognitively primitive species, the instinctive triggering of neurochemical states may be the only thing that we can recognize as analogous to emotions. In more advanced species, the kind of response that I label below as "intuition" also becomes involved in the triggering, but it is the neurochemical response that makes the drive recognizable as "emotion". Thus, in a phrase, it seems that emotions have an evolutionary origin in instinctive responses.

  Some of these instinctive responses resemble moral responses, though it is arguably anthropomorphic to insist that they really justify that term in other species. In our species, however, the human-specific versions of these instincts appear to play a very important role in our moral sense, namely that without the contribution of the instincts, moral judgements would have no motivating force. An emotionless person might well understand the intellectual consequences of the Golden Rule, but have no desire to act upon them. (The very word "desire" indicates the emotional content).

- **Intuition** By "Intuition" I mean a learnt response that is not open to minute conscious examination, though it influences what comes to the conscious mind when we attend to a particular question. Another way of saying roughly the same thing is that it cannot be precisely expressed through any kind of language. In human morality, intuition comes to the fore when we have to make judgements, which are typically not the direct result of a mechanical procedure which we would know how to program but are what seems right after a period of reflection and calling different aspects to mind. In terms of Kahneman's "Thinking,

Fast & Slow", the intuition is also what does the fast thinking. Thus in the case of everyday choices with a moral dimension that have to be made "on the fly", it is the intuition that guides our response at the time. Neurologically, intuition in the human brain appears to rely greatly on pattern-recognition, often of an extraordinarily complex kind. This affects the relationship to instinctive responses, since emotions are not just felt but recognized, remembered and anticipated by the human mind.

- **Intellect** By "Intellect" I mean the faculty that enables overt reasoning, whether formally logical or by pointing to analogies or to agreed facts, for example. The human intellect in particular relies on the abstract symbols of language, in the broad sense that includes mathematical and musical symbols. In human morality, the intellect enables discrete facts, rules, and arguments to be brought to bear to influence the intuitive picture of the world by which we make judgements. One way it can do so is by incorporating a belief in certain formalised rules, whether laid down by some authority or justified by common experience. The intellect also of course plays a major role in communication between individuals, although when that communication has emotional force the other I's are involved as well. In Kahneman's sense, the intellect does the slow thinking, so is involved when we have to wrestle with a dilemma, though the intuition is involved as well: different intuitions are brought to bear as intellectual processes result in conscious attention shifting from one aspect to another. Neurologically, the correlates of conscious, intellectual thought are the subject of only very preliminary research, but one promising line of enquiry suggests that it is a manifestation of the massive integration of massively diverse processes.

- I will use the term '**rational**' to refer to the 'Intuition' and the 'Intellect' acting together, and not just with reference to overt reasoning.

The assumed interaction between the 3 I's can perhaps be best elucidated by considering a variety of examples, necessarily somewhat speculative. Only typical behaviour is assumed – for example, infants, psychopaths, and the brain-damaged are not included.

Human altruism and the Golden Rule

Many scientists and philosophers have sought an "explanation" for altruism in evolutionary terms, but it may also be the case that human altruism has an entirely different explanation in the specific features of human cognition – or the case may be some mixture of the two. What is almost or perhaps completely unique about human altruism is that it is potentially universal, directed by some people towards complete strangers or even to members of other species. The features of human cognition that I am referring to are, firstly the reliance on a continually evolving and hugely complex mental picture of the world, and secondly, the ability to stand outside that model, and see the self as though in the third person. The first feature appears to be shared with some other species, the second seems to unique, at least in the degree of development, in the human – notably excluding infants. Thus preferences, desires etc. relating to behaviours or outcomes can be *felt from a third-person stance*. It seems likely that this arises from the way the innate generic architecture of the human brain wires itself up in response to normal early experience, that is the capability is innate but the implementation is intuitive. At a later stage this pattern of response may be recognized and formulated verbally by the intellect as something like the "Golden Rule". This in turn may reinforce the intuition, so that asking how an action would look like from another's perspective becomes automatic rather than reasoned on every occasion.

Loyalty and Xenophobia

With the mechanisms supporting altruism and the Golden Rule in place, the hard-wired feature of the ancestral brain that supports emotional attachment to members of the same tribe/troupe may plausibly be hijacked later in evolution to reinforce concern for those understood *intuitively* rather than *instinctively* to be "one of us". The other side of the coin, of course, is xenophobia. The conception that there are beings unknown to us personally to whom we owe obligations as "fellows" can be understood as defining a categorical boundary between those who are "one of us" and those who are not, and who are therefore excluded from the obligations. In the course of human history advancing communications have gradually led to increasing knowledge of different tribes, communities, lands etc. so that the intuitive picture built up by many people – and, due to education, the intellectual picture – has involved a gradually enlarged categorical boundary until today many people believe there is no basis for excluding any members of the human race from altruistic concern in principle.

In fact, this principle has been so hammered home that some people take a critical or condescending view of our ancestors to whom xenophobia was entirely natural, but this must be an anachronistic judgement. So far as those early ancestors knew, it *could* have been the case that other races were so different as to be unable to share any common moral understanding. As it happens, evidence over the years has demonstrated very powerfully that the entire human race shares so much (culture apart) that a strong case can be made against xenophobia of any description, but I cannot see how that can make it 'intrinsically' wrong; it is just wrong for our species, on our planet – on my best intuitive judgement, after reflecting on what I intellectually know. (I can therefore wholly *dissent from* the judgement of my ancestors without being justified in criticising them, let alone feeling morally superior).

Many people believe intellectually that, even without accepting a xenophobic view, there are benefits in committing to a rule which places specific obligations on us to fellow members of certain groups, most obviously nations. This depends on an intuitive model of the role of nations at the present stage of history.

Affection and Nepotism

Some people interpret utilitarianism as an injunction to treat everyone equally in every respect, which almost implies loving no-one, or at least not acting upon love for anyone lest this unjustly prioritises their welfare above that of strangers. Most people however take the "rule utilitarianism" position that the world is a better place if love and affection for selective members of one's circle are acted upon, provided that the actions are not of certain 'inappropriate' kinds. Nepotism is an example of the exercise of family affection in a way that is widely held to be morally inappropriate, the difference being that the preference is being made by somebody acting to exercise some power conferred upon them rather than in their personal capacity. What this highlights is that the same *emotional* state, affection for a relative, can be *morally* right and wrong at different times: the rightness or wrongness doesn't stem exclusively from that particular emotional state, but from its combination with a rational understanding of how the chosen action relates to a wider picture. On the other hand the wider picture only affects the action to the extent that there is a *desire* for justice, so the rightness or wrongness doesn't stem exclusively from the rational understanding either.

Animal welfare

The case for animal welfare is somewhat like the case against xenophobia: it is founded on a recognition of some degree of similarity of at least some other species to our own. For example, it may be believed that other species suffer in a way that is more or less analogous to the way we suffer. This belief may be supported intellectually by some empirical facts – from similarities of behaviour to common brain structures involved – but with obstacles to a purely rational conclusion such as the lack of verbal communication with other species, this belief must depend eventually on an intuitive judgement, which may both prompt and be prompted by instinctive responses as well as the intellectual analysis.

A case such as a cat torturing a baby bird for fun highlights the tension that may exist between the 3 I's. Our instinctive response may be one of sympathy with the bird and condemnation of the cat, but many people would regard that intellectually as an inappropriately anthropomorphic reaction, and would consciously suppress the instinctive reaction (while still acknowledging sympathetic distress). In this view, the cat, despite domestication, still belongs to a world of nature that we intuitively understand as an amoral world 'red in tooth & claw'.

Implications

There is no intended suggestion that this theory of the nature of morality can of itself provide answers to moral dilemmas – on the contrary, since it underlines how complex morality can be. It is rather a warning against oversimplified claims that leave one or two of the 3 I's out of the picture. At best it can provide a sort of check-list of things to bear in mind: how do I feel about this, is this feeling biased by personal interests or is it consistent with how I believe the world works in other contexts, are there relevant rules that I have mentally signed up to? All things which are common sense if not hindered by restrictive dogmas.

APPENDIX: The 3 I's compared with traditional moral theory

Philosophical History

The strongest schools of modern philosophical thought about the basis of morality are probably "moral rationalism" stemming from Kant and "moral sentimentalism" (in the technical, not the fluffy kitten sense) stemming from Hume.

Kant argued that right and wrong has nothing to do with feelings – kindness to others motivated just by the nice warm feeling it gives you may be pleasing, but it doesn't count as being *moral.* Morality has to be based on rational analysis. There is certainly some force in this argument. In cases such as nepotism (as discussed above) and giving a false alibi for a friend, the motivation may be precisely the feeling that one is showing kindness or affection to the beneficiary, but it is generally held that this behaviour is immoral – such an emotional approach should be trumped by what is essentially a rational, third-party analysis of the relationships concerned and their consequences. However, an obvious problem with Kant's approach is the one I have mentioned in discussing "Instinct" – that the outcome of a rational analysis provides within itself no motivation to act upon it. We must at least have an emotional commitment to rationality to recognize Kant's "categorical imperatives" as imperative for us. Moreover, the kind of analysis that Kant

recommends involves judgements that imply values, which in turn must come from somewhere, and where else than from our feelings?

Hume by contrast insisted that "reason is and ought only to be the slave of the passions" and that an "ought*"* (or "ought not") statement cannot justifiably be derived from an "is" (or "is not") statement. Again, one can see where he is coming from: these are essentially the objections to the Kantian school referred to above. Reason without felt values is the province of a text-book in logic rather than of a human being, and attempts to derive *oughts* from *ises* always seem to end up begging the question, since the "ought" requires a motivation which is not present in the "is". Hume did acknowledge that facts ("*is* statements") can be combined with moral values ("*ought* statements") to derive further moral values. However, this does not seem enough for cases such as nepotism and false alibis and bird-torturing cats where the introduction of overt reasoning effectively reverses the judgement that "the passions" on their own tend to imply. At least, they seem to act on their own. Looked at more closely, what the reasoning reverses is a debatable, otherwise unexamined intuitive picture that the emotions are responding to.

It has increasingly seemed to me as I have studied the topic that Kant's arguments are stronger in proving that emotion is not *sufficient* to establish morality than in proving that it is not *necessary*; and conversely Hume's arguments tend to prove the insufficiency of reason rather than its lack of necessity. So the truer position seems to me to be that *both are necessary*: that morality arises from a *conjunction* of emotional and rational facts. Moral right & wrong seem to be names we give to certain situations of judgement where our gut and our neo-cortex act together, so to speak. This is more than a case of pointing in the same direction; without a contribution from both capabilities they don't point in *any* direction. So the question of which is the "final arbiter" can never arise. The contending judgements being chosen between are *both* combinations of thought and feeling when examined closely enough.

In the previous paragraph I considered rationality in the broad sense as a single capability. Now consider that modern philosophy and (especially) psychology has pointed to the inevitable interdependence of overt and intuitive reasoning, two different ways in which the neo-cortex operates. Then we arrive at the tripartite conjunction of three types of capability examined here: the three I's.

Objections: subjective and objective theories of morality

One philosophical objection that might be raised against the above treatment (main text) is that it might be seen as committing the "naturalistic fallacy". This is the attempt to derive moral principles from an examination of "natural" behaviour (an example of the derivation of *ought* statements from *is* statements that Hume challenged). Such an attempt can obviously be flawed in the sense that, say, lying and cheating are perfectly "natural" types of behaviour, which does not make them morally right. However what I have attempted to do is to examine, not 'any old' natural behaviour, but mental behaviour *specifically when trying to make moral choices*, so that moral premises and motivations are imported at the start of the process – thus avoiding the charge. That is, I am observing merely that is natural for humans to use categories of right and wrong, and to seek to distinguish between them, so it is reasonable to enquire what kinds of mental process give rise to these categories. The one kind of 'is' statement from which an 'ought statement' can be legitimately, indeed tautologically, derived is a true statement about the nature of obligation. Suppose you accept my account of the '3Is' basis of morality. To the question, "Yes, but why *should* I behave in the way that my 3 I's jointly decide is correct?", I can then reply

"because that what is what you must mean by 'should': you have already accepted that there is no higher standard of obligation that your question can refer to. If I am right you are already motivated to behave this way, without needing rational persuasion".

A philosophical objection likely to be made by philosophers in the Kant tradition is that my account denies the "objective" nature of morality. Kant even wrote that "The ground of obligation ... must be sought *not in the nature of the human being* or in the circumstances of the world in which he is placed...". Taken at face value, this seems perverse since his rational analysis places great stress on such concepts as human autonomy, dignity and freedom of will, which are surely a part of "the nature of the human being". Possibly what he had in mind was more the widely supported argument that the ground of obligation is *not a matter of current human subjective judgement* – as Dworkin put it, whether torturing babies for fun is wrong does not depend on whether somebody thinks it is, or even whether everybody thinks it is.

In the form stated here, there is an excellent case for this view, because judgements are never final. If I (or the human race) thought that torturing babies for fun was OK yesterday but think it is wrong today, and if morality were ultimately no more than a matter of judgement, then it would follow that such torture was *actually* OK yesterday and is wrong today. But we intuitively understand morality as something that is corrigible – if I change my mind, I do so for some reason, and believe my new conclusion is better than the old one. So there must be some ground *outside* my subjective judgement for making such a change.

Or consider a debate between two people with opposite moral views – such as a pacifist and a believer in just warfare. We know that in practice there is little likelihood that they will come to an agreement, but equally, both debaters will consider that their views are facts about the point at issue, not facts about themselves (by contrast to liking on not liking Marmite, say). They will both believe their views are *right*, and probably feel frustrated by the knowledge that they cannot express them as persuasively as they feel they deserve to be expressed. All this supports the notion that a moral belief is a belief about something other than the current personal predilections of the believer.

This explains why morality is considered by many people to be about something "out there", external. And yet, as I argued in my position paper "The Basis for Secular Ethics", morality is unimaginable in a world with no conscious beings, so "out there" cannot for us be outside human consciousness: it must arise from what is universal about human consciousness, which is 'internal' from our point of view as being a representative of a type, but 'external' from our point of view as an individual self. So even though the ground of obligation is not a matter of *current* subjective judgement, that does not mean it is not a matter of *potential*, and in some sense '*ideal*', subjective judgement. The concept of the 3 I's can now help to clarify this.

- Firstly, as our *instincts* are by definition predetermined, literally "in our DNA", those that contribute to our moral sense are part of the universal human basis of morality. If some genetic variation results in a lack of moral sense, (innate psychopathy) we regard the person concerned as an exception, as we do an infant, not as belonging to the universal type.
- Secondly, the *intellectual* contribution is universal to the extent that it is universally communicable. If another person reaches a different moral conclusion *solely* because one of us is lacking a fact or overt argument that the other has in mind, this difference can in principle be resolved by comparing notes.

- The problem comes from the *intuitive* contribution. Take the case of the argument over pacifism. Both parties will typically feel that *if only* they could make the other person "see the world in the same way", then they must accept the conclusion they are denying. That is to say, anybody whose intuitive understanding of a situation is the same would eventually (after rational argument and fact-sharing) evaluate it in the same way, so if only everyone's intuitive pictures of the world could be pooled in a universal intuition, there would be no moral dilemmas! But consider what this actually means. Each individual's intuitive picture of the world is an expression of their individual "connectome", the set of billions of neural connections that they have built up over time. So the ultimate pooling of intuitions requires the integration of billions of connectomes in a sort of ultra-connectome – to be revised and re-integrated whenever any of the individuals concerned gains new information! Although the impracticality of this ideal is beyond description, there are two good reasons for thinking about it. Firstly, it helps to explain the intractability of moral disputes, and secondly, although the *ultimate* ideal is far beyond reach, it provides a pointer to the *direction* of movement that is needed. If two people are able to share just a little bit of their intuitive models of the world, the prospect of their moral judgements converging is enhanced. Such sharing can come from many sources – a photograph, a play, a metaphor or other vivid verbal expression, and so on. If any 'insight' (fragment of intuition) leads to me to revise some moral judgement, by definition that judgement is improved, in the sense of being based on greater insight than before, even if it is eventually reversed by a further insight.

Thus, rather than defining "moral truth" by its absurdly inaccessible ultimate state, it is more useful to define it as being that which *can be approached more closely* in a normal human mind (the instinctive contribution) by greater insight (the intuitive contribution) and greater relevant knowledge (the intellectual contribution).

Refs:

Hume*: "A Treatise on Human Nature", Book 2, Part 1 §1

Kant*: "Groundwork of the Metaphysics of Morals", §4

* e.g. cited by Baggini, "The Edge of Reason", Ch. 7

Dworkin, "Justice for Hedgehogs", Ch. 1