

"Word and Action: Reconciling Rules and Know-How in Moral Cognition" in R. Campbell and B. Hunter (eds) MORAL EPISTEMOLOGY NATURALIZED: CANADIAN JOURNAL OF PHILOSOPHY Supp. Volume 26 (2000) (University of Calgary Press, Alberta, Canada) p.267-290

Word and Action: Reconciling Rules and Know-How in Moral Cognition

Andy Clark

Philosophy/Neuroscience/Psychology Program
Department of Philosophy
Washington University
1 Brookings Drive
Campus Box 1073
St. Louis, MO 63130
andy@twinearth.wustl.edu

Abstract

Recent work in Cognitive Science highlights the importance of exemplar-based know-how in supporting human expertise. Influenced by this model, many accounts of moral knowledge now stress exemplar-based, non-sentential know-how at the expense of the rule-and-principle based accounts favored by Kant, Mill and others. I shall argue, however, that moral thought and reason is an intrinsically complex achievement that cannot be understood by reference to either of these roles alone. Moral cognition -- like other forms of 'advanced' cognition -- depends on the subtle interplay and interaction between multiple factors and forces and *especially* (or so I argue) between the use of linguistic tools and formulations and more biologically basic forms of thought and reason.

- 1.

Introduction: Rules and Know-how

How should we conceive the rule of summary, linguistically-formulated rules, principles and imperatives in moral thought and reason? There is a tradition, rooted in Kant (1781) and Mill (1884), but extending all the way to Donagan (1997) and beyond, that finds in such rules and principles the rational essence of true moral thought and hence of genuinely *moral* choice and action. The Kantian, for example, identifies moral actions with actions predicated upon moral maxims which the individual is able to endorse (will) as universal law. Actions thus grounded display genuine moral worth

A rather different tradition, rooted in Aristotle, stresses not the endorsement of summary moral maxims, but a kind of practical wisdom, gained by rich and sound experience and honed by the needs of daily action. It is this tradition, whose modern exponents include Flanagan (1996), that has become most firmly linked to work in cognitive science and (especially, and see below) in Artificial Neural Networks. For in the stress on practical wisdom grounded in rich experience of actual exemplars of real moral problems, it has seemed that “Aristotle’s perspective and the neural network perspective ... converge.” (Churchland, P. M., (1996), p. 106).

These two pictures (the one stressing skilled know-how, the other stressing explicit moral rules and reasoning) are typically depicted¹ in strong mutual opposition. Churchland, for example, goes on to comment that:

This portrait of moral person as one who has acquired a certain family of perceptual and behavioral *skills* contrasts sharply with the more traditional accounts that pictured a moral person as one who has agreed to follow a certain set of rules.

P.M. Churchland, 1996, p. 106.

I shall argue that such bald opposition is a mistake: a distortion of the complex nature of moral reason that obscures the real source of human moral expertise. For human moral expertise, I shall argue, is made possible only by the potent complementarity between two distinct types of cognitive resource (or “mind-tool” – Dennett (1996, Ch. 5)). One is, indeed, the broadly pattern-based, associative learning capacity that we share with other animals and artificial neural networks. But the other is, precisely, the very special modes of learning, collaboration and reason made available by the tools of words and discourse. A mature science of the embodied mind will, I suspect, have as its very center a sensitive account of how advanced thought and reason emerges from the complex and repeated interactions between these two sets of resources.

2. Moral Expertise

One driving force behind the cognitive scientific endorsement of the “practical wisdom” model of moral knowledge is an influential model of expertise. It is a model grounded in both cognitive psychological studies and in attempts to replicate expert skill using artificial neural networks. With this model of basic biological cognition, I have no quarrel - indeed, it is one that I positively endorse (see Clark (1989) (1993)). What I shall be questioning, however, is the concomitant down-playing of the role of actual human talk and discourse in exposing and navigating the moral realm: a down-playing evident in comments such as:

“Stateable rules are not the basis of one’s moral character. They are merely its pale and partial reflection at the comparatively impotent level of language.”

P. M. Churchland (1996a), p. 107.

“The skill development model we are proposing ... demotes rational, post-conventional moral activity to the status of a regression to a pre-expert stage of moral development.”

Dreyfus & Dreyfus (1990), p. 252, 256.

What, then, *is* the cognitive scientific vision that is meant to lead us to such dramatically “anti-sentential” conclusions? It is, in large part, the demonstration that certain kinds of neural network device (of which biological brains are plausibly an example) are capable of extracting and encoding information (knowledge) in forms whose richness, fluidity and context sensitivity far outstrips anything that could be supported by a set of linguistically couched action selection rules, principles or maxims. At the heart of the alternative conception is a daunting story about vectors, prototypes, high-dimensional state spaces and non-propositional, distributed encodings: a story nicely laid out in, for example, McClelland, et al (1989), P.S. Churchland and T. J. Sejnowski (1991) and elsewhere. This computational story dovetails well with cognitive psychological work suggesting that much human knowledge is organized around encodings of prototypical cases rather than via the use and storage of rules and definitions (see Rosch (1973), Smith & Medin (1981)). For one way to think of the way knowledge is encoded in a neural network is to think of the experienced network (the network after extended training on example cases of input and desired output) as

commanding a rich and context-sensitive battery of prototypes ready to be deployed in response to incoming stimuli.

A prototype, as it emerges in this kind of model, is just the statistical central tendency of a set of exemplars each of which displays a range of features. Typical neural network learning algorithms take as input a set of exemplars (or “training instances”) and yield a knowledge-base in which the most typical features and the most typically co-associated feature groups become highlighted and play an especially potent role in driving future recall, generalization and problem-solving. One way to think of the extracted prototype is as a point or region in a space which has one dimension for each possible feature. The learnt prototype of a dog may thus subsist in a feature space whose dimensions include height, weight, aggressiveness, color and so on. Exposed to a variety of dog exemplars, the connectionist system learns the most common features and combinations and hallucinates a “typical dog” – one whose precise feature combinations need not correspond to that found in any actually encountered exemplar, but which reflects the overall statistics of the training set. Churchland (1995) notes that a mature human brain may command a variety of such prototypes: not just ones for concrete objects, but ones for (e.g.) economic, social, moral, political and scientific concepts and ideas. In fact, all human knowledge, according to Churchland, is encoded in this same general way. (for further discussion, see Churchland (1989) (1995), Clark (1989) (1993), Fodor & Lepore (1992)).

The potential importance of all this for our conceptions of moral knowledge and moral reason rests on two specific consequences. The first concerns the grain and nature of the knowledge itself. The second concerns the acquisition of such knowledge and the development of moral expertise. Concerning grain and nature, the now-familiar (see e.g. Churchland (1996a), Goldman (1993), Flanagan (1996), Johnson (1993), Clark (1996)) point is that our moral knowledge may quite spectacularly outrun anything that could be

expressed by simple maxims or moral rules. For the mode of storage and organization, combined with the sheer capacity of biological neural networks, reveals simple sentential formulations as, in Churchland's words, nothing but "a one-dimensional projection of a [high] dimensional solid" (1989, p. 18). The project of reducing the knowledge (know-how) encoded in a biological neural network to a set of summary moral rules is every bit as intractable as that of capturing the knowledge of an expert bridge or chess player in a set of such maxims. Such maxims, as Dreyfus and Dreyfus (1990) suggest, are of great value to the novice. But they cannot replace the finely tuned pattern recognition skills of the expert who (we now assume) deploys a highly trained neural network of great dimensionality, tuned by exposure to countless instances and minor variations, and organized around a multiplicity of stored prototypes representing the fruits of long, hard hours of play and practice. The second important consequence is now also apparent. The development of moral knowledge and expertise is thus crucially dependent on rich and varied moral experience. It cannot be successfully instilled by exposure to simple sets of moral maxims, any more than great (or even good) chess play can be achieved without dense and repeated practice. Seen in this light, the provision of 'moral tales' and (even better) our occasional immersion in rich, often morally complex or ambiguous, novels takes on a new importance. Such "virtual moral experience" affords just the type of training that begins to approximate the kinds of rich, real-world experience needed to instill genuine moral expertise.

Churchland (1995), (1996a) also offers a vision of moral debate and moral progress firmly grounded in this story concerning supra-sentential individual moral know-how. The account has three important features. First, it depicts moral knowledge as fundamentally similar to non-moral (e.g. scientific) knowledge. Second, it depicts moral debate as a clash between the different prototypes to which a real-world scenario might be assimilated. And third, it depicts moral progress as real (genuinely occurring),

but as consisting in something like increasing knowledge of how to match ones behavior to the social framework is which one finds oneself.

The picture, with a little more flesh, is this. The child finds itself in a complex physical and social setting and must acquire a set of perceptual and behavioral skills to flourish. This is not, or not primarily, a matter of learning a set of rules, so much as learning how to act and respond in fluent, highly context-sensitive ways. And it is accomplished by the slow, experience-basic training of complex on-board neural networks. The successful learner develops both “skills of recognition” and “skills of matching...behavior and the moral circumstances at hand” (Churchland (1996a) p. 105). The processes by which we acquire such skills are not, however, specifically moral. They are the very same processes involved in all kinds of skill acquisition and the knowledge obtained is in no way special. Scientific knowledge, according to Churchland, is ultimately practical knowledge: knowledge of “how to navigate the natural world.” Moral knowledge, likewise, is just knowledge of “how to navigate the social world” (op cit p. 106). Moral and scientific knowledge thus emerge as equally ‘objective’ – answering to a set of constraints imposed by a reality not itself constructed by the individual brain that seeks to comprehend it. And each domain allows for genuine - though pragmatic - kinds of progress in thought and understanding, as we (both as individuals and as a species) learn better and better ways to navigate both physical and social space. Moral disagreement, to complete the picture, arises when different agents or groups use different prototypes to characterize a situation. Moral ‘debate’ then consists, largely, in the attempt to assert one prototypical understanding (e.g., the fetus as person) over another (e.g., the fetus as internal growth). Churchland says little about how such debates are resolved, but presumably any good resolution must ultimately turn on the relative usefulness of the prototypes as aids for navigating complex social space.

This concludes my thumbnail sketch of Churchland's connectionist- inspired story about moral thought, argument and progress. It is a story wonderfully grounded in some of the most powerful developments in recent cognitive scientific thought. But it is a story marred, I believe, by a failure to engage with a crucial dimension of human thought and reason in general: a dimension, moreover, whose role in enabling moral thought is especially critical. The missing element, to repeat, is a better appreciation of the complementary role of moral language and practices of moral debate in both the construction and the communal navigation of 'moral space.'

3. Talk and Thought

The connectionist image of moral thought and reason (in particular) and of advanced thought and reason (in general) is incomplete. It fails to do justice to the profound effects of various species of "external scaffolding"² on human cognition. One striking effect of this failure is an impoverished view of the role of public language and public discourse in moral reason.

It is all too easy, I have claimed (Clark (1996)) to allow the stress on supra-sentential moral know-how to blind us to the profound importance of linguistic exchange and moral maxims. Even if it is true that individual moral understanding typically far surpasses anything that a few words can capture, it is *also* true that moral reasoning and decision-making is quintessentially a communal and collaborative affair. It involves the attempt, in a large class of morally crucial cases, to communally negotiate a course of action that meets the differing constraints of multiple individuals and interest groups: identifying these constraints and displaying a course of action as fair and reasonable is at

the heart and soul of much contemporary moral thought and reason (think, for example, of the debates concerning the social acceptability of euthanasia, or the delicate balancing acts involved in setting legal constraints on internet activity). Yet this process -- of negotiating a practical course of action that meets the needs of multiple different groups and individuals -- is linguistically-mediated through and through. It begins with the attempt to articulate, often in quite summary ways, the leading ideas, needs and desires of various parties, and proceeds via the repeated exchange of linguaform formulations until some vaguely acceptable compromise is reached. Such a process, as long as it is guided not simply by local expediency, but also with an eye to more global moral issues, underpins all our attempts to discover fair, communally acceptable, co-operatively developed solutions to moral, social and political problems.

In Clark (1996) I argue, in much more detail, that the marginalization of summary linguistic formulations and sentential reason found in Churchland (1989) (1995) (1996a) and in Dreyfus and Dreyfus (1996) is a mistake. Summary linguistic formulations are not, as e.g. Dreyfus and Dreyfus claim, mere tools for the novice. Rather, they are essential parts of the socially extended cognitive mechanisms which support communal reasoning and collaborative problem solving, and are thus crucial and (as far as we know) irreplaceable elements of genuinely moral reason. They are the tools that enable the co-operative explorations moral space: a space which is intrinsically multi-personal and whose topology is defined largely by the different – but interacting – needs and desires of multiple agents and groups.

But language, I now want to suggest, may play an even more fundamental role in the development of moral knowledge and the exercise of moral reason: a role associated not simply with the communal search for better moral understanding, but with the very process by which the moral realm “comes into view” as an object of human cognitive endeavor.

There is a substantial case to be made that abstract, ‘higher-level’ thought of all stripes is possible – for biological brains whose basic computational profile is much as Churchland suggests – only in virtue of a kind of cognitive augmentation rooted in the use of signs, labels and words. In much this vein, Dennett (1996) speaks of the role of multiple *mind-tools* in enabling human cognitive success, while Clark (1997a, 1997b) stresses the complementary cognitive benefits of stable, linguistic tokens and fluid pattern-associating styles of individual processing. There are two distinct (but deeply related) dimensions to such benefits, and it is worth treating each in turn.

The first dimension concerns learning and the initial “bringing into focus” of the moral domain. Concerning learning in general, it can be shown (Clark and Thornton (1997)) that all known learning algorithms -- both connectionist and classicist -- flounder in the face of a certain kind of complexity in the training data. Very roughly (see the previously reference for a lot more detail) such algorithms (assuming they are essentially unbiased i.e. not provided with antecedent domain-specific knowledge) are likely to uncover only relatively superficial statistical regularities in the input data. To help uncover deeper (“hidden,” relational, and higher-order) regularities, a powerful strategy is to re-code the input data so that complex, “hidden” regularities show up as simple surface regularities. For example, if you re-code two poker hands (9, 10, J, Q, K and 10, J, Q, K, A) as “straight” and “straight” the deeper relational commonality between the two sequences of cards emerges as a simple surface commonality in the labels. The bulk of the work, in such cases, is done not by the basic learning device but by the re-coding, however that is achieved.

Now suppose we view public language, with its convenient stock of moral terms and labels as -- from the point of view of individual reason -- a kind of “found” reservoir of potentially useful re-codings: re-ifications of complex patterns and tendencies in the

underlying web of events. Such items, societally accrued by the painful, slow processes of extended search, trial and error over cultural time, will serve to reduce complex, otherwise “cognitively invisible” patterns to tractable, learnable regularities and will enable individual brains to pursue a deeper and more penetrating exploration of moral space. Such exploration is *possible*, for basic biological intelligences like ours, only in virtue of the repeated transformations of understanding achieved by the use of the linguistic signs and labels. When we learn about promises, duties, rights, obligations and so on, we are not just learning labels for moral relevant states of affairs. We are learning to see the world through a new lens –one in which ever deeper patterns (of interactions and conflicts among entities such as rights and duties etc.) are revealed as potential objects of thought and reason. Language, on this account, is the special tool which brings the moral realm into focus for biological, pattern-based engines of reason, by displaying a “virtual reality” in which morally salient patterns and descriptions are the concrete objects of subsequent thought and reason.

Lest this all seem too speculative, abstract and unanchored in concrete cognitive scientific research, it is worth pausing to consider a simple-but I think striking-demonstration. In recent work with chimps (pan troglodytes), Thompson, et. al. (in press) show that the provision of simple arbitrary tokens as “names” for relationships allows the chimps to discern complex higher-order relationships which are otherwise invisible to the basic (non token-augmented) chimp brain. Chimps trained to associate the relation of sameness (two shoes, two cups, etc.) with one arbitrary tag (a plastic token), and the relation of difference (e.g. a shoe and a cup) with a different arbitrary tag, prove capable of learning to spot the presence of higher-order relations. Of a display containing two cups and two shoes, they can then judge that the *same* relation is displayed (sameness in each case). Ditto for a display containing two pairs of unmatched items (cup/shoe, ball/banana) – here the same relation is again displayed, although it is the relation of difference. Given one pair of matched and one of unmatched items, the

higher-order relation is correctly judged to be difference: one pair displays sameness, the other difference, hence the higher-order relation is difference. Chimps lacking the history of token training cannot learn to solve these higher-order problems. A compelling conjecture is that the higher-order task becomes tractable because each visually presented pair recalls an image or representation of the appropriate plastic token, thus reducing the higher-order task to a lower-order one (are the two mentally recalled tokens the same or different?). The image is thus of a potent cascade in which culturally acquired tags and labels make available a new quasi-perceptual space in which biologically basic capacities of pattern-recognition can be used to negotiate new and otherwise cognitively invisible realms.

Notice, then, that even the process of moral development (learning to see the moral domain itself) involves a complex interaction between linguistic tokens and signposts and our fine-grained (perhaps ultimately prototype based) appreciation of the situations in which they apply, and of the kinds of actions that may or may not be appropriate. For it is our exposure to moral labels (labeling acts as kind, greedy, selfless, misguided and so on) that enables our pattern-sensitive brains to isolate morally salient patterns which might otherwise remain buried beneath the noise of more superficial similarities and dissimilarities. When Dreyfus and Dreyfus (1990) depict linguaform slogans (“don’t tell lies,” etc.) as mere tools for the novice, they have in mind a less potent role: the role of enabling *roughly* acceptable response while we are busy developing (via repeated experience) more precisely honed skills. The developmental image presented above accords moral talk and maxims a more fundamental role: they actually *make possible* the kinds of pattern-and-experience based learning that Dreyfus and Dreyfus celebrate, at least in those cases in which the target patterns are abstruse and easily swamped by other surface regularities.

There is, of course, a second dimension to this kind of (broadly) language-based cognition enhancement. For the line between learning and mature reason is not hard and fast, and the cognitive benefits of linguistic encoding are by no means restricted to the processes by which we come to learn about complex moral reality. Indeed, the picture of linguistic tokens as the stable anchor points around which to organize ever-more-abstract forms of (in this case) moral perception and reason applies equally well to the case of mature thought. A version of this claim is ably defended by Jackendoff (1996), who suggests a number of (what he terms) “indirect” effects of language on thought. The idea is that language makes thoughts and complex trains of thought and argument into objects available for attention and inspection in their own right. The linguistic formulations help stabilize complex ideas in working memory and help keep separate the various elements of complex thought and arguments, allowing further scrutiny and repeated re-visiting from different argumentative perspectives. In this vein, Jackendoff *op cit* asserts that “language is the only modality that can present to consciousness abstract parts of thought like kinship relations, reasons, hypothetical situations and the notion of inference.”

Moral maxims and recipes are thus the anchor points for moral thought and reason. They are the re-visitable islands of order which allow us to engage in exploratory moral discourse, approaching practical moral problems from a variety of angles while striving, nonetheless, to maintain a sense of our targets, priorities and agreed-upon intermediate positions. While Churchland and others are surely correct to insist that the maxims *themselves* cannot do justice to the depth and sensitivity of our moral understanding, they are wrong to depict them as shallow reflections of our real understanding. Instead, they are tools for focusing, holding steady, and refining moral understanding. And without them, our explorations of moral space, if possible at all, would be shallow indeed.

But why should public language be so well fitted to play the role Jackendoff describes? The answer, I think, turns directly on the two roles for linguistic tokens scouted earlier. The first role was as a medium of collaborative effort, and to that end public language is already well-shaped as a tool for interrogation, criticism and the pursuit of reasons. The second role was as a source of new stable entities standing-in for complex patterns and relations. The potency of linguistic formulations as anchor points for mature thought and reason flows directly from these more basic attributes. Public language is the favored means by which pattern-completing biological brains bootstrap their way into the realm of cascading abstraction, detailed self-criticism, and the reflective pursuit of arguments and reasons: clear pre-requisites, surely for any advanced form of moral sensibility.

A partial parallel may exist in the human capacity for mathematical thought. One suggestion recently supported by an impressive array of cognitive psychological and neuroscientific evidence (see Dehaene (1996), Dehaene et. al. (1999)) is that the distinctive human ability to do arithmetic depends on the coordinated interaction of two quite distinct kinds of “mind-tools.” The first is an innate biological competence at low-grade approximate arithmetic: a simple number sense shared with pre-linguistic infants and other animals, to detect a few absolute quantities (one, two, three) and to appreciate changes in quantity and relative quantities (more, less). The second is a culturally acquired and numerical-symbol mediated capacity to describe other exact quantities (22, 103, etc.) despite the lack of any biologically basic capacity to distinguish e.g. thoughts about “53-ness” from thoughts about “52-ness” etc. Genuine arithmetical understanding Dehaene argues, is possible (for brains like ours) only courtesy of the complex interplay between culturally constructed (and initially external) symbolic systems and the rougher innate sense of simple numerosity.

It is not my purpose to critically discuss this suggestion here (but see Clark (to appear)), but merely to display the form of the conjecture. For it once again suggests ways in which basic biological brains may rely on the complementary resources of public symbols and codes to successfully penetrate and navigate otherwise intractable regions of intellectual space.

The larger picture which is emerging is thus one in which neither of the original poles (the Kantian emphasis on explicitly endorsed moral principles or the Aristotelian stress on practical wisdom) captures the subtlety, power and complexity of human moral intelligence. Instead it is the cognitive *symbiosis* between basic, prototype-style, pattern-based understanding and the stable surgical instruments (for learning, criticism and evaluation) of moral talk and discourse that conjures real moral understanding. A mistake to be avoided, in displaying this complex interplay, is that of depicting linguistic formulations as merely low-grade approximations to the finer-grained moral sensibilities realized in on-board neural networks: the mistake of in seeing the linguistic ideas as pale, partial and somewhat impotent reflections (recall Churchland 1996a p. 107) of real moral understanding. To do so is like suggesting that the rake and the plough are merely pale, partial and somewhat impotent reflections of our real ability to farm. Better by far to see the rake and plough as special tools which effectively transform the space in which the farmer's native skills of physical action are deployed and which allow her to tame and exploit terrain which would rapidly repel any unaugmented advances. The image to abandon is the image of simple translation: linguistic items do not benefit us solely in virtue of their problematic translation into some finer-grained inner code. Instead, they act by complementing, and being brought into co-ordination with, the kinds of prototype-based knowledge celebrated by Churchland, Dreyfus and others.

4. Situated Moral Epistemology

A situated moral epistemology, as I am using the term, will offer an account of moral knowledge and moral reason which is sensitive to the critical roles of a variety of non-individualistic elements. Such elements include the social role of language in collaborative problem-solving (Clark 1996), the scaffolding effects of larger scale institutional and organizational contexts (for an ‘economic reason’ based foray, see Clark (1997a)), and (see above) the cognitive and developmental benefits conferred by the framework of moral discourse and moral labels through which our neural engines of reason confront the world.

Among the consequences of any shift towards a more situated moral epistemology are, first, a somewhat different perspective on the issues concerning moral progress and the possibility of genuine moral knowledge, and second, a concomitant shift in our ideas about the natural systems in which moral thought and reason inhere.

Concerning moral progress and the possibility of genuine moral knowledge, it is again useful to compare the situated approach with Churchland’s more ruggedly individualistic stance. For where Churchland stresses the potential for steady moral progress and convergence, a more situated story suggests historical path-dependence and multiple divergent explorations of an exponentially large space of moral possibilities. Churchland’s position, recall, is that moral knowledge is fundamentally like scientific knowledge. Both are best seen as bodies of know-how, represented (connectionist-fashion) in the head, and subject to the same kinds of selective pressure: pressure to successfully engage the natural world, on the one hand, and the social world, on the other. Churchland thus emphasizes:

The practical and pragmatic nature of both scientific and broadly normative knowledge [and] the fact that both embody different forms of know-how: how to navigate the

natural world is the former case, and how to navigate the social world, in the latter.

P. M. Churchland 1995 p. 292

Flanagan (1998) is critical of this parallel, detecting in Churchland as kind of “naïve enlightenment optimism about moral progress and convergence” (op cit p. 35). Where Churchland draws a parallel between moral and basic scientific understanding, Flanagan suggests instead that the moral realm is less constrained and less likely to support strong notions of objectivity, convergence and progress. Moral knowledge, Flanagan plausibly asserts, is intrinsically more local and is better compared to the multiplicity of successful strategies characteristic of a complex ecology. Critique and debate is thus always “perspectival,” rooted in the local moral ecology. And there are enough different, but equally successful, ways of living (niches) to undermine any simple appeal to a “best way” to negotiate social space.

Something like Flanagan’s picture can be illuminated, I suggest, by further reflection on the role of linguistic labels and other forms of external scaffolding in influencing the shape and course of moral thought and reason. For the ultimate source of Churchland’s “blind-spot” is his implicit commitment to an essentially individualistic and non-constructivist naturalization program: one in which moral thought and reason is depicted as a property of individual biological brains, each seeking only the safe navigation of an essentially independent social space.

By contrast, our explorations of moral space are – on my account – highly constrained by the specific linguistic frameworks which both enable and restrict human thought and reason. The provision of labels and tags aids us, as we saw, by creating new “virtual worlds” in which to bring to bear basic biological capacities of pattern-based reasoning. But every choice of moral vocabulary is likewise restrictive, rendering *other*

patterns invisible to all but the most breathtaking (“revolutionary”) exercises of individual thought. This would not matter if the shape of the social world really did enforce certain practices of labeling and talking, as Churchland seems to suggest. But such a view looks to get the descriptive cart well before the constructive horse. For surely, in *at least* equal measure, it is our communal practices of talking, labeling and categorizing that create the special kinds of social space whose successful navigation becomes the ‘goal’ of the child’s brain. And these communal practices, as Flanagan again points out, are dictated by multiple forces and local and historical contingencies. As a result, there simply *is (in one sense)* no such thing as “the social world,” existing independently of our moral frameworks and capable forcing those frameworks into shape. There is, I will allow, a useful (though famously elusive) notion of human flourishing which affords some partial anchor for our moral sensibilities. But here too, there are simply untold numbers of ways to flourish, and (as Flanagan stresses) social niches in which and do so. The overall result is indeed a kind of naturalized, pragmatic pluralism, in which moral knowledge is relative to a niche which is partially, at least, constructed by the very apparatus (of moral talk, labels and categories) with which we confront it.

This idea may be better appreciated by way of an analogy. Consider for a moment the case of “financial knowledge and reason.” Financial reason, as Arthur (1994) nicely points out, involves dealing with an evolved financial ecology that is, in a very concrete sense, partially created by our own activities of labeling and exchanging. You begin with something relatively basic – the trading of valuable items (‘underlyings’, currencies, debts, soybeans, etc.). But with this structure in place, increasingly complex swaps and trades become possible. We can buy and sell futures, which are contracts to deliver an underlying at a fixed later date, or options to buy. These options and futures then *themselves* constitute new kinds of underlying, that can be traded as valuable items

in their own right. Which in turn opens up the space of options on futures and so on. (For the full story, see Arthur (1996))

What I hope to draw from this example is a sense of the power of tags and labels to at once *create and explore* complex spaces. The creation is not unconstrained, nor ex-nihilo. But the spaces themselves become available for exploration only courtesy of the linguistic practices themselves.

Moral thought and reason seems to share at least something of this nature. A recent example is the emergence of whole new domains of moral and legal complexity surrounding the use of the internet and cross-boundary trading (see e.g. Johnson (to appear)). Another case, suggested to me by Carl Wellman, concerns the right to refuse medical treatments even when necessary to prolong life. This right has recently been labeled (the label being recognized, though not yet approved, in *Cruzan vs. Director, Missouri Health Department* (1990) 4 97 US 261 at 277) the “right to die.” Should such a label become approved, it is easy to see, as Wellman suggests, how it may help reconfigure the space of local moral, ethical and legal argument. For it then invites a rights-based argument in favor of physician-assisted suicide.³ In a similar vein, Nancy Fraser (1989) traces the way changes in the way we talk about spousal abuse have gone hand-in-hand, historically, with changes in the way we evaluate such behavior.⁴ No doubt examples could be multiplied. The point, however, is just to suggest – once again – that the forces that shape our communal explorations of social and moral space are highly diverse, and that social practices of talking and labeling are potent tools that simultaneously shape moral space, and make it accessible to pattern-based biological reason. Doing justice to this complex dance must be a primary goal of any situated moral epistemology.

All this has clear implications, it seems to me, for the broader business of “naturalizing moral thought and reason.” For such a project cannot hope to succeed – if I

am right – without recognizing that moral reason is a function not simply of our individual psychological profiles, but of the larger social, cultural, and linguistic systems in which they participate: larger systems which display a marked historicity and path-dependence, and in which change, progress and evolution are determined as much by their own intrinsic dynamics as by the actions or choices of the individuals “within” them. In studying and attempting to understand these systems, we should not be tempted to reduce their dynamics to the dynamics of individual biological brains, but *neither* should we assume all the interesting work is done by the larger social and organizational structures alone. For the moral machine is genuinely a complex ecology in which there has evolved a fit (for better or worse) between individual agents’ practices and expectations, and the social, legal and political institutions and structures in which they are embedded. There has been co-evolution between all the elements of this matrix. Moral debate influences the development of moral (social, political, legal and educational) infrastructure. But moral thought and reason *is itself* constantly and profoundly affected and transformed by this infrastructure (think of the role played by the U.S. Constitution in sculpting contemporary American debates over pornography, gun control and so on). Our pattern-completing brains are thus directed not at some bare, noumenal version of “moral space,” nor even at the more mundane space of practical social living. They are directed, so I claim, at the (communally constructed) space of local moral infrastructure: a space populated by laws, constitutions, formal and informal norms and a persistent host of summary moral maxims, as well as by the less visible, but equally constraining, structures of commerce, industry and educational practice.

5. Conclusions: The Cognitive Complementarity of Rules and Know-How

Moral knowledge, even Churchland admits, must involve a mix between praxis and theoria (see Churchland (1996b)). But having conceded this much, he goes on to claim that “the brain draws no distinction between them: both kinds of knowledge are

embodied in vast configurations of synaptic weights” (op cit p. 305). I have tried to argue for an alternative view: that moral knowledge is made possible by, and consists in, a carefully orchestrated interplay between the kinds of rich, nuanced, know-how directly embodied in neural states, and the genuinely different cognitive tools provided by moral talk and language. Such tools function not by simple translation into ‘neuralese’, but by providing a new and importantly different class of objects on which to target our individual modes of understanding. Moral talk and labels, on this account, transform the space of moral reason as thoroughly as the invention of money transformed the space of trade and “bargaining reason.”

Sensitivity to this transformative role reveals a narrow but navigable pathway between two superficially competing visions of moral knowledge: the rule-based and the know-how/exemplar based approaches to moral cognition. Between Kant and Mill, on the one hand, and Aristotle (and Churchland) on the other, we now dimly perceive a middle territory, in which genuine moral knowledge is *essentially* a function of the complementarity between two sets of mind-tools: the tools of rich, context-sensitive practical wisdom celebrated by Churchland, and the linguaform tools which allow us to deploy those basic biological resources in increasingly abstract, complex and collaborative domains. Human moral reason, thus construed, is structurally akin to human mathematical reason, each depending crucially upon the co-ordination of our basic biological sensitivities with the new tools of public formalism and linguistic debate.

There are other conflicts, too, that may be at least partially reconciled by this kind of treatment. These include the debate over the role of large-scale embedding structures in explaining human action (Marxist ideas about structural determination vs. Mill’s ideas about individual free action, for example)⁵ and the debate between moral realists and relativists. But such topics I leave for another day. In closing, I would stress only the highly preliminary nature of these remarks and speculations. Moral thoughts and moral

talk do, I believe, form a deeply interanimated whole. But the specific shape and nature of the crucial interactions is not yet clear, nor the ultimate implications of such interactivity of our conceptions of moral knowledge and reason. Getting this picture right is surely essential if cognitive science is to make a genuine contribution to moral epistemology.

References

- Aristotle (1972). The Nicomachean Ethics. London, Oxford University Press.
- Arthur, W. B. (1994). "On the Evolution of Complexity". Complexity: Metaphysics, Models and Reality. G. Cowan, D. Pines and D. Meltzer. Menlo Park, Addison Wesley: 5-77.
- Bakhtin, M. (1981). The Dialogic Imagination. Holquist, M. (ed). Austin, University of Texas Press.
- Berk, L. (1994). "Why Children Talk to Themselves." Scientific American, November: 78-83.
- Churchland, P. and T. Sejnowski (1992). The Computational Brain. Cambridge, MA, MIT Press.
- Churchland, P. M. (1989). The Neurocomputational Perspective. Cambridge, MIT/Bradford Books.
- Churchland, P. M. (1995). The Engine of Reason, the Seat of the Soul. Cambridge, MA, MIT Press.
- Churchland, P. M. (1996a). "The Neural Representation of the Social World". Mind and Morals. L. May, M. Friedman and A. Clark. Cambridge, MIT Press: 91-108.
- Churchland, P. M. (1996b). "Flanagan on Moral Knowledge". The Churchlands and Their Critics. R. McCauley. Oxford, Blackwell: 302-306.
- Clark, A. (1989). Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing. Cambridge, MIT Press.
- Clark, A. (1993). Associative Engines: Connectionism, Concepts and Representational Change. Cambridge, MIT Press.
- Clark, A. (1996). "Connectionism, Moral Cognition and Collaborative Problem Solving". Minds and Morals. May, L., Friedman, M. and Clark, A. Cambridge, MA, MIT Press: 109-128.
- Clark, A. (1997a). Economic Reason: The Interplay of Individual Learning and External Structure. The Frontiers of the New Institutional Economics. J. Drobak and J. Nye. London, Academic Press: 269-290.
- Clark, A. (1997b). Being There: Putting Brain, Body and World Together Again. Cambridge, MA, MIT Press.
- Clark, A. (to appear). "Minds, Brains and Tools: Comments on Dennett". Philosophers of Mental Representation. H. Clapin. Cambridge, MA, MIT Press.
- Clark, A. and C. Thornton (1997). "Trading Spaces: Connectionism and the Limits of Uninformed Learning." Behavioral and Brain Sciences 20(1): 57-67.
- Dehaene, S. (1997). The Number Sense. Oxford, Oxford University Press.

- Dehaene, S., E. Spelke, et al. (1999). "Sources of Mathematical Thinking: Behavioral and Brain Imaging Evidence." Science **284**: 970-974.
- Dennett, D. (1996). Kinds of Minds. New York, Basic Books.
- Donagan, A. (1977). The Theory of Morality. Chicago, University of Chicago Press.
- Dreyfus, H. and S. Dreyfus (1990). What is Morality? A Phenomenological Account of the Development of Ethical Expertise. Universalism vs. Communitarianism: Contemporary Debates in Ethics. D. Rasmussen. Cambridge, MA, MIT Press: 237-264.
- Flanagan, O. (1996). "Ethics Naturalized: Ethics as Human Ecology". Mind and Morals. L. May, M. Friedman and A. Clark. Cambridge, MA, MIT Press: 19-44.
- Fodor, J. and E. Lepore (1993). "Reply to Churchland." Philosophy and Phenomenological Research **53**: 679-682.
- Frazer, N. (1989). Unruly Practices: Power, Discourse and Gender in Contemporary Social Theory. Minneapolis, MN, University of Minnesota Press.
- Goldman, A. (1986). Epistemology and Cognition. Harvard, Harvard University Press.
- Goldman, A. (1993). "Ethics and Cognitive Science." Ethics **103**: 337-360.
- Jackendoff, R. (1996). "How Language Helps Us Think." Pragmatics and Cognition **4**(1): 1-34.
- Johnson, D. (to appear). "Emergent Law and Order". The Biology of Business. J. Clippinger (ed).
- Johnson, M. (1993). Moral Imagination: Implications of Cognitive Science for Ethics. Chicago, University of Chicago Press.
- Johnson, M. (1996). How Moral Psychology Changes Moral Theory. Mind and Morals. May, L., Friedman, M., Clark, A. Cambridge, MA, MIT Press: 45-68.
- Kant, E. (1781 (=A)). A Critique of Pure Reason. London, MacMillan.
- Kant, E. (1787 (=B)). Critique of Pure Reason. London, MacMillan.
- Kant, E. (1977). Critique of Practical Reason. Indianapolis, Bobbs-Merrill.
- McClelland, J. L. (1989). Parallel Distributed Processing -- Implications for Cognition and Development. Parallel Distributed Processing -- Implications for Psychology and Neurobiology. R. Morris. Oxford, Clarendon Press.
- Mill, J. S. (1884). A System of Logic. London, Longmans, Green & Co.
- O'Neill, O. (1989). Constructions of Reason. Cambridge, Cambridge University Press.
- Rosch, E. (1973). "Natural Categories." Cognitive Psychology **4**: 324-350.
- Smith, E. and D. Medin (1981). Categories and Concepts. Cambridge, MA, Harvard University Press.
- Sterba, J. (1996). Justifying Morality and The Challenge of Cognitive Science. Mind and Morals. May, L., Friedman, M. and Clark, A. (eds). Cambridge, MIT Press: 239-256.
- Thompson, R., D. Oden, et al. (in press). "Language-naive Chimpanzees judge relations between relations in an abstract mapping task." Journal of Experimental Psychology: Animal Behavior Processes.
- Todorov, T. (1984). Mikhail Bakhtin: The Dialogic Principle. Minneapolis, University of Minnesota Press.

Vygotsky, L. (1986). Thought and Language. Cambridge, MA, MIT Press.

Wellman, C. (1999). The Proliferation of Rights. Boulder, CO, Westview Press.

¹ For an exception, see Sterba (1996).

² This broad notion of external scaffolding is based on Vygotsky's (1962/1986) observations about the role of linguistic rehearsal in learning, but extends the image to include all the kinds of social, linguistic and technological support that transform the problem spaces confronting the naked biological brain. See Clark (1997b) for the full story.

³ For a useful discussion of the emergence of new legal rights, see Wellman (1999).

⁴ Thanks to Joel Anderson for drawing this example to my attention.

⁵ For an excellent treatment, see Andrew Sneddon, "Agents and Actions," Unpublished doctoral thesis, Queens University, Canada.