5 Symbolic Thought and the Evolution of Human Morality

Peter Ulric Tse

How Human Symbolic Thought Came into Being

While the cognitive, perceptual, and emotional capacities of humans and apes share much in common (e.g., numeric thought: Hauser, Dehaene, Dehaene-Lambertz, & Patalano, 2002; certain substrates of language: Hauser, Chomsky, & Fitch, 2002), there is a core set of abilities that makes humans fundamentally different from apes and other animals in kind, not just in degree. There are no documented cases of other animals dancing to rhythmic sounds, intentionally producing art, or spontaneously generating and manipulating meaningful symbols. Although some animals have learned to use what appear to be symbols in a rudimentary fashion (e.g., Savage-Rumbaugh, McDonald, Sevcik, Hopkins, & Rubert, 1986; Savage-Rumbaugh, 1987; see also Kaminsky, Call, & Fischer, 2004; Pepperberg, 2002), they do not generate them spontaneously in the wild and cannot do so in the flexible, generative, and recursive manner of humans (Petitto & Seidenberg, 1979; Terrace, Petitto, Sanders, & Bever, 1979). While chimpanzees and bonobos can be trained to associate a meaning with an arbitrary sign, establishing such an association typically takes many trials (Petitto & Seidenberg, 1979). Humans still have associative learning in common with other animals, but in addition humans have one-shot learning of associations among arbitrary categories of things and events. While there are instances of one-shot associative learning in the animal world, such as the Garcia effect, where animals learn to avoid the food last eaten before becoming nauseous (even when the food could not have caused the nausea), cases such as the Garcia effect are the exception that proves the rule. One-shot learning is possible in such special cases because certain animals are hardwired to make particular kinds of one-shot associations, such as between nausea and food. In general, however, animals, unlike humans, are not hardwired for one-shot learning if the learning takes place

between arbitrary classes of events. Repetition is typically necessary to learn arbitrary associations for mammals other than ourselves because this type of learning is the learning of probabilities of event co-occurrence and predictiveness. Associative learning through repetition is a far cry from the one-shot learning of a symbol's referent by a human child or the ease with which a child can change a symbol's referent at will. Such learning can occur in the absence of co-occurrence between a thing and that to which it refers, and indeed it can occur in the absence of any direct experience with a thing at all, as when a child learns the meaning of the word "ancestor" or "heaven." Indeed, the cognitive/neural changes that made one-shot learning among arbitrary events possible also made symbolic cognition possible. Our nearest animal relatives may be able to learn arbitrary associations, but they are not using symbols in the way that we humans use symbols. Beyond lacking syntax, they are not typically able to effortlessly assign or reassign an arbitrary meaning to a given symbol once an association has been learned. It is this arbitrary and flexible relationship between a symbol and its referent(s) that is the hallmark of true symbolic thought, setting it apart from mere association, and it is one that all other known animals appear to lack. While it has often been emphasized that syntax and language are what most separate us from higher apes, syntax and language are most likely more recent developments[1] that emerged long after the more fundamental ability to understand and use arbitrary symbols nonsyntactically. In short, the capacities that set humans apart in kind, not just in degree, from all other known animals include capacities and propensities for art, music, dance, analogical reasoning, abstract thought, the spontaneous generation and use of symbols, and the ability to reason abstractly about others (e.g., Povinelli & Preuss, 1995; Povinelli, Dunphy-Lelii, Reaux, & Mazza, 2002) and about events, as well as the ability to manipulate symbols recursively and syntactically. The first part of this paper explores the hypothesis that all these modes of human behavior and cognition share a common root cause.

The central claim developed here is that neuronal circuits underlying perceptual, motoric, and cognitive capacities that were functionally distinct and modularly encapsulated (Fodor, 1983) in our earlier chimp-like ancestors came to interact through a new type of attentional binding in our more recent ancestors. This paper addresses the architectural changes in information processing that permitted this breakdown of modularity,[2] without getting too deeply into any "engineering" solutions at the level of neuronal circuitry that might have realized these architectural changes. However, one or more neural/genetic changes must have occurred to

account for the emergence of uniquely human modes of behavior and thought at or after the divergence of our genetic line from that which led to chimpanzees and bonobos. Because of this change, operators previously limited to operands within a given module could now operate over the operands of other modules. Cross-modular binding and entrainment are seen as the root cause underlying the birth of uniquely human modes of cognition.

The Nature of Symbols

It is not necessary to review the complexities of semiotic theory to understand that a symbol has various aspects. One is the perceived "sign," which exists as a representation that is seen, heard, touched, or otherwise experienced. For a given person, a sign can stand for one or more referents, which can be thought of as the meaning of the sign. A sign's referent(s) can belong to many different types of mental representations, which themselves need not be signs. If the mental representation is one of an object (or type of object) in the outside world, then the sign symbolizes that thing in the world. When a sign has a referent (i.e., a meaning) for a particular perceiver, it is a symbol for that perceiver. A symbol is assumed to possess two key defining properties: (1) a symbol is a mental representation that can be stored in long-term memory or held in short-term memory that can stand for one or more arbitrary stored or online representations; (2) a symbol can be flexibly remapped to an existing or new referent without a need for many trials of learning to build up an association. In other words a symbol is arbitrary. Its meaning is not based on the likelihood or degree of co-occurrence between a sign and its referent. It is simply assigned. Several animals may have some capacity to process information in the sense of (1), but only humans, as far as we know, possess the capacity to process symbols in the sense of (2) as well. A mind capable of (1) but incapable of (2) can learn complex associations between an object or event and some referent, but it is "associational" rather than truly symbolic. Many animal minds are exquisitively sensitive to probabilities of co-occurrence. If one object or event tends to predict or accompany the occurrence of some other object or event, cognition capable of (1) can make and learn an association between these objects and events. Classical conditioning is an example of an ancient form of associational learning. In contrast, a three-year-old child can pretend that a block is a truck and then, a moment later, pretend that it is a monster. This capacity to instantly remap the referent of an object file is unique to humans and is at the heart of why our cognition can be truly symbolic. Even if it turns out that

symbolic processing is not unique to humans, this would in no way alter the present theory, since the goal here is to explain the changes in cognitive architecture that led to human symbolic cognition. If dolphins, bonobos, or martians turn out to be symbolic as well, that would be nice, but it has no apparent bearing on how we became symbolic. It so happens that humans appear to be unique in our symbolic cognitive capacities, but this is not in itself all that interesting. What is interesting is understanding how our minds came to be the way they are.

Attention and Object Files

An "object file" (Kahneman, Treisman, & Gibbs, 1992) is a metaphor for attentional processes that combine multiple types of information existing over various modalities into a common bound representation of an object. An object file is an attentionally tracked "figure" (Lamy & Tsal, 2000) integrated as a temporary episodic representation in a short-term memory buffer (Kahneman, Treisman, & Gibbs, 1992; Schneider, 1999) that maintains a coherent identity even as the particular contents defining that identity change over time. While there are a variety of conceptions of what an object file is (Pylyshyn & Storm, 1988; Carey & Xu, 2001; Scholl, Pylyshyn, & Feldman, 2001), all conceptions have in common a psychological entity that keeps track of an object over time, within some space. This space need not be spatial space. It could be any space that obeys a metric. For example, one can track in musical space. One can listen to a symphony and track the oboe. Once can then listen to the same symphony again and this time track the lead violin. In both cases the sensory input is the same. What differs is the nature of the object file one constructs. Object files can only exist in the context of a memory buffer that keeps track of the bound representation. Even though the term "object file" contains the word "object," an object file need not correspond to an object in the outside world. It could just as well correspond to a thought or plan that one is keeping track of in the presence of complete sensory deprivation. Nonetheless, the prototypical example of an object file is one associated with a tracked visual object, and it is in the domain of concrete object tracking that object files may have had their origin. The contents of an object file may change, and the labels attached to that object file may change, as when one says, "It's a bird; it's a plane; it's Superman!" However, the object endures as a single object over time by virtue of having a single object file associated with it, not by virtue of having constant contents or a constant label. The contents of an object file are thought to be mid to high level. That is, there is widely thought to be a preattentive stage of

representation that cannot be attended and whose contents cannot be added to an object file (Wolfe, 2003; Treisman & Gelade, 1980). Attention can only operate over automatically preprocessed representations available after the operators of early perceptual systems have transformed raw sensory input into mid-level representations. Possible object file contents can be perceptual features, such as color and texture, that may exist on feature maps (Treisman, 1992; Quinlan, 2003), mid-level structures such as surfaces (He & Nakayama, 1992), abstract identity tags (Gordon & Irwin, 1996), or higher level conceptual information (Gordon & Irwin, 2000). The contents of an object file are not static. They are also not simply dictated by the flux of bottom-up sensory input. They can be changed through the manipulations of numerous cognitive operations, such as mental rotation or attention, as when one attends to any of a number of well-known bistable figures to bring out the appearance of one figure and then another. The sensory input has not changed, but the contents of the object file have changed dramatically because of the action of a cognitive operator on sensory operands.

One of the central ideas developed here is that symbols and symbolic thought are inherently attentional in nature, because they involve (now or involved in the past) the binding of a sign with arbitrary referents, whether temporarily in working memory or more durably in long-term memory. With unintentional repetition or intentional practice, arbitrary representations can become bound together in memory. This is the relatively slow process of associative learning that humans share with other animals. With attention to joint object files, however, arbitrary representations can become bound together in memory after one instance.[3] Attentionally bound symbols and referents can be "chunked" or bound together in memory, recalled as a unit, and later processed without the need for further attention. However, at the stage of encoding, one-shot binding of symbol and referent occurs because symbols and referents occupy a common object file. Symbols are inherently attentional because binding (within a tracked object file) is just what attention is.

How Object Files Changed: Cross-Modular Binding

Whereas many nonhuman animals appear to have the capacity to monitor, select, ignore, track, and otherwise attend to objects and therefore must have analogs of human object files in their cognitive architecture, human and animal object files differ. A dog's object file of a tree, for instance, presumably contains color and shape information represented in different neural populations or maps, indexed by location, as well as other

information that the dog has learned to associate with this particular tree or trees in general. A dog's object file can be multimodal, in the sense that auditory, haptic, olfactory, visual, and other information can all occupy the object file of the tree. All types of information, however, are about the tree. A dog's object file is encapsulated in the sense that it cannot contain information about irrelevant objects or events. Object file encapsulation presumably helps animals survive, because irrelevant information is not represented, permitting the animal to remain undistracted and unconfused by matters irrelevant to survival and the matter at hand.[4] Human object files, unlike animal object files, can contain *any* information which can be attended or downloaded from memory into the object file. As such, human object files can enact truly cross-modular binding. Because any representation can be downloaded into the object file of a tree, a tree can be taken to stand for "my friend Bob," "truth," or anything else. This requires tagging the tree component of the object file as real (i.e., pointing to the world) and the Bob component as not real or referential (i.e., pointing to a representation that need not be in the world). In the absence of such a tag, a person might take a tree to really be Bob. Cognitive modularity and encapsulation protect animal minds from cognitive "noise" and the obvious dangers of misrepresentation, psychosis, delusion, and hallucination.

Whereas animals seem to be capable of internally modeling events that might happen or might have happened, humans go beyond this "literal" capacity to one of imagination that can model events and objects that could never happen and could never exist in the real world. Imagination became possible when arbitrary contents and operators could be downloaded to a common object file. For example, the representation of wings could be downloaded into the object file containing the representation of tree. The operator that places one object onto another could then be accessed to create a new representation of a tree with wings. Such a construction was not possible in the more chimp-like minds of our early ancestors.

A human object file of a tree can contain everything that a dog's would, plus information that has nothing to do with this particular tree or any tree. For example, a man looking at a tree can simply decide that it stands for his wife. This is accomplished by downloading representations of his wife into the object file holding the tree information. That information can be as minimal as a label or pointer to the concept wife, or it could be much more specific to his particular wife and as elaborate as circumstances demand. Importantly, this downloading to the present object file need not

take place in a conscious manner or a manner dictated by the plans or goals of a central executive. It can happen entirely automatically and unconsciously. Because of automatic cross-modular binding, he may see a tree standing next to a bush, and this may remind him of his wife and son because of their similar size relationships. He did not plan this to happen. It happened in a stimulus-driven manner because similarity in one domain automatically triggered activation on other maps encoded in terms of the same relationships, and these new activations, in this case corresponding to representations of his wife and son, were automatically downloaded to the object file containing the representations of the tree and bush, which then could function as symbols of his wife and son. He might say that the tree and bush remind him of his wife and son, but this fact became available to his conscious report after the link between disparate representations had been made. Thus, symbols, although inherently attentional in nature, because initially mediated by the attentional construct of an object file, are not necessarily volitionally constructed. It must be emphasized that symbolic thought, though inherently attentional, need not be driven by volition or intentions, although it can be. Just as a motion in the visual periphery will automatically draw one's attention, symbolic thought can be carried out in an automatic or nonvolitional mode. In particular, reminding is a consequence of the basic architecture of human symbolic cognition. A dog cannot be reminded of anything by a tree and a bush other than things that have a direct link in its experience or in the world with a tree and a bush. Animal cognition, lacking any basis for symbolism, reminding, or metaphor, is inherently literal.

Cross-Modular Binding Gave Rise to New Types of Mental Dysfunction

Human cognition, though much richer than that of a dog or chimpanzee, runs the danger of associational breakdown where everything reminds one of everything else, or stands for anything else, and also runs the danger of hallucination, where the border between what is real and self-generated breaks down. Quite simply, our minds can break down in novel ways that the minds of most animals cannot. It is possible that various types of psychosis, schizophrenia, and many other human mental disorders are the result of the malfunctioning of neuronal systems or cognitive operations which in mentally healthy individuals permit the possibility of symbolic, analogic, creative, and imaginative thought. Examining the neural and genetic bases of disorders of these types of higher cognitive processing holds great promise in revealing the genetic basis of the change(s) in neural circuitry that must have given rise to cross-modular binding.

Symbolic Thought versus Syntactic Thought

The ability to use and recognize symbols must have preceded the evolution of language because whereas syntactic processing is necessarily symbolic, symbolic processing need not be syntactic. Language by its very nature involves the utterance and manipulation of symbols. It is therefore unlikely that language as such was the capacity that first emerged to define protohuman symbolic cognition. According to the cross-modular binding hypothesis, primitive types of dance, art, music, humor, analogy, and symbolic reasoning evolved early, followed potentially much later by the emergence of syntactic operations over symbols. Indeed, binding and the ability to form associations across previously encapsulated classes of information may have at first been relatively rudimentary capacities. The first symbols used by our ancestors may have been no more than sticks and stones used to stand for and mimic objects, animals, or people. And the human body itself could easily have served as a symbol, because pretending to be something or someone else would require placing the body and the referent in the same object file. Indeed, some recent authors suggest that early human communication took the form of mimesis rather than language (Corballis, 2003; Donald, 1993). No such use of symbols would have left material evidence in the fossil record. It would therefore be a mistake to assume that symbolic reasoning emerged with the first evidence of art in the fossil record. However, once even rudimentary symbolic reasoning was in place, selective pressure could operate on this new capacity, leading over time to the complex forms of symbol manipulation found in modern humans.

Cross-modular binding freed object files to contain information from other modules. The "primary" content of an object file corresponding to a perceived sign became a symbol when information to which the sign referred could be added to the object file. Although an initial binding between symbol and referent requires attention in the case of one-shot learning, once encoded together as a "chunk," a symbol can access its meaning directly, even in the absence of attention, because it is bound with its meaning in memory. Thus, a symbol need not invoke a joint attentional object file if it has already been "chunked" and laid down in long-term memory. That is, such prebound symbols can be processed in the absence of attention. However, even a prebound symbol must have once been an attentional construct, because only attention can bind disparate types of information within a common object file.

Cross-modular binding may also be able to account for the emergence of syntax, which presumably evolved over a long period after the emer-

gence of primitive symbolic expression. The syntactic symbol manipulation found in language and mathematics has as its core the recursive application of rules for the manipulation of symbols (Chomsky, 1965; Hauser, Chomsky, & Fitch, 2002). Other animals apply computations recursively (e.g., to the representation of their own position when they navigate or when they carry out sequential motoric acts with their hands), but these nested computations appear to be largely modular. Hauser, Chomsky, and Fitch (2002) have recently suggested that such modular and domain-specific recursive computations may have become domain general over the course human evolution. That is, the recursion found in one module could be applied to the contents of other modules. This hypothesis is consistent with the general principles developed here. Binding across modules may have permitted computations carried out over, say, a recursive navigational map, or the body's nested sequencing movements through space, to be applied to the contents of other maps. Once recursion could be applied to symbols, symbols could be combined recursively and therefore generatively. Natural and sexual selection could then act on this capacity, resulting ultimately in our present capacity of language. The focus here, however, is to explain not the emergence of syntax and language but the emergence of the symbolic cognition that must have preceded syntactic and linguistic cognition.

Cross-Modular Neuronal Entrainment

Although this is a cognitive theory of cross-modular binding, some speculation concerning the neuronal mechanisms underlying cross-modular binding may be useful. It should be emphasized, however, that cognitive architectural changes described up until now could be correct even if the neuronal mechanism outlined here turns out not to be correct.

Until this point cross-modular binding has been talked about as cross-modular attentional binding involving symbolic referents. However, some forms of cross-modular excitation or inhibition may have happened in a bottom-up manner not requiring the imposition of symbolic referents. One of the most basic and perhaps earliest consequences of cross-modular information transfer was entrainment across previously encapsulated modular representations. That such entrainment occurs in the human mind is an important clue to the manner in which modular encapsulation broke down. One fascinating clue is that some form of dancing occurs in all human cultures. A martian who descended to earth would probably find it curious that only one species on earth undulates its body in synchrony with patterns of air pressure in the 1-Hz range. He might find it

equally curious that this species does not spontaneously undulate in the presence of rhythmic pulses of odor, light, or taste. This asymmetry offers clues to the neural substrate of cross-modular binding. Dancing affords the simplest example of a novel behavior that arose because of binding across modules. After the mutation(s) in question, neurons tuned to auditory stimuli could entrain motor neurons, allowing auditory rhythms to trigger rhythmic motoric behavior. Prior to the occurrence of the gene mutation(s) that presumably caused the rewiring of neuronal circuitry that permitted cross-modular binding, even rudimentary dancing behaviors in response to auditory rhythms would have been cognitively and motorically impossible, because neurons within the auditory module would not have entrained motor neurons. Such entrainment presumably requires (cortico-cortical or corticothalamocortical) excitatory connections that presumably did not exist in the brains of our chimp-like ancestors and do not exist in the brains of other animals. The absence of such connections is the reason why no nonhuman animal dances or can dance to rhythmic sound. Although they can hear the sounds, they quite literally cannot be entrained to the beat.

Neural entrainment does not require the existence of object files that contain symbolic referents. Neural entrainment merely requires coexcitation of neuronal activity across and within modules. For example, the gesticulations that modern humans make while talking are precisely synchronized with speech (McNeill, 1985) and convey meaning that supplements language, particularly when the speaker is conveying information about space or emotions (Goldin-Meadow, McNeill, & Singleton, 1996). If modern human gesticulations are indicative of prelinguistic gesticulations, then hand, arm, and head motions may have become entrained to the emotional content and rhythm of existing vocalizations long before the evolution of spoken language or syntactic symbol manipulation or even symbolic thought (cf. Corballis, 2002; Donald, 1991).

The cross-modular binding hypothesis operates under the assumption that the neuronal encoding of events on different modules will share common spatiotemporal characteristics (such as bursting, or firing with a particular frequency or pattern) if the stimuli to which the corresponding neurons are tuned have common spatiotemporal characteristics (such as abruptly changing, slowly changing, intense, or nonintense). For example, the binding hypothesis would predict that the activation on the auditory module "wawawawa" should entrain and therefore be associated with smooth or undulating hand motions (motoric) or drawn curves (visual) in all human cultures, as opposed to the more abrupt hand motions or jagged

shapes that would be associated with a sound like "tiktiktiktiktik." This would follow because auditory modulation is smooth in the former case but abrupt in the latter.

Neural entrainment and binding across modules can only occur if there are direct or indirect neuronal connections that permit entrainment to take place. One prediction that follows from the binding hypothesis is that there are corticocortical, corticothalamic, or thalamothalamic connections between neurons within different modules in humans that are not present in chimpanzees or other nonhuman apes. It is not simply that there is more connectivity in the human brain; it is of a different kind. Modules that function relatively independently in chimpanzees, because they are not directly connected, should share direct uni- or bidirectional axonal connections in humans or share indirect connectivity via another area, such as the thalamus. Diffusion tensor imaging and fiber tract tracing of the different species' cortices and thalami may be able to ascertain whether this prediction is correct.

Ramachandran and Hubbard (2001) have recently proposed that synesthesia is due to "cross-wiring" between the brain areas that subserve the percepts that synesthetes simultaneously perceive. For example, synesthetes who perceive different black numbers to have different colors may have fibers that link the area that codes for numbers and the area that codes for color. The authors point out that both areas lie near the angular gyrus. Because synesthesia runs in families and may be X-linked (Bailey & Johnson, 1997), it is possible that the presumed excess fiber proliferation or faulty pruning that underlies synesthesia is genetic. This proposal is related to the present cross-modular binding hypothesis to the extent that coactivation across maps arises from novel cross-connectivity. Whereas synesthesia would arise from novel cross-connectivity not found in most other people, binding arose from novel cross-connectivity that did not exist in earlier hominids. Once this trait arose, it was presumably selected for until it spread throughout the entire population. Thus, binding across modules is, according to the current hypothesis, a normal part of human cortical organization, and not a rare occurrence as is synesthesia. The current hypothesis is one about how the normal human brain developed, not just one about unusual, synesthetic brains. Moreover, their "cross-wiring" proposal does not offer a model of how neuronal connectivity gives rise to synesthesia. The binding hypothesis, in contrast, ascribes a specific function to cross-modular connectivity. In particular, neural entrainment across previously encapsulated modules is presumed to arise because of synchronous patterns of neuronal discharge. And cross-modular

binding occurs in the general case because of operators and operands shared within a common object file. Future research will have to determine whether the genetic basis of synesthesia is related to the proposed genetic basis of the connectivity that underlies cross-modular binding in normal human brains. There is no reason to think that it is.

The Birth of Analogical Cognition

On this account, the human capacities for metaphor and analogy arose from cross-modular binding because events represented within one module could entrain and therefore spontaneously remind one of events represented by neurons within another module. The cross-modular binding hypothesis would predict that there are analogies that transcend cultures (cf. Lakoff & Johnson, 2003), because the root of those analogies is the entrainment of neurons across modules that are responding to corresponding properties of stimuli. For example, if neurons tend to respond with an abrupt burst of firing to the onset of a loud sound, they will entrain neurons that respond with an abrupt onset within, say, the visual module. Similarly, neurons that respond to more intense auditory stimuli by firing more would entrain neurons that respond correspondingly to visual input. Loudness and abruptness of sounds should therefore be considered analogous to brightness and abruptness of visual onset across all human cultures. In contrast, an abrupt and loud sound and a gradually modulating or dimming light should not seem analogous to any humans, regardless of culture. However, it is not only the spatiotemporal (e.g., intense, abrupt, grouped, or continuous) and physical (e.g., pitch, color) characteristics of stimuli that would entrain neurons tuned to corresponding characteristics within other modules. That is, neural entrainment need not be only sensorimotoric. The semantic, cognitive, and emotional characteristics of stimuli could drive cross-modular entrainment as well. Thus, one might predict that the feeling of sadness or depression would be associated with lowness or downness across all cultures, and that joy would be associated with highness and upness. The same might be said for intelligence and quickness or sharpness, on the one hand, and stupidity and dullness or slowness on the other. Countless similar examples can be given. Although such examples seem amusing, it really is an empirical question whether all cultures make such associations.

A deeper kind of analogical correspondence than sensory-sensory correspondence would have emerged as sensory modules entrained cognitive modules. For example, a tall tree among bushes might remind one of a tall

person surrounded by short ones. It is this reminding-of that comprises the basis of analogical or metaphoric thought. Analogical thought is in turn a by-product of the synchronous firing, coupling, or entrainment of distal neuronal populations across modules. In order for such analogies to be automatically evoked by a given stimulus, the manner in which, say, tallness is encoded in the firing of neurons on the sensory module in question must be similar enough to the manner in which corresponding qualities are encoded across entrained modules that entrainment is possible. This raises the prospect that there is a shared neuronal pattern of firing for similar qualities of objects and their relationships, even when the neurons representing that similar information exist within different modules.

The emergence of binding across modules triggered not only the birth of symbolic thought by permitting arbitrary objects to stand for arbitrary referents but also triggered automatic cross-modular activations. Analogy and metaphor are now defining aspects of human cognition. Indeed, they are perhaps almost as important as the ability to express and entertain ideas using symbols. Analogy emerged for the same reason that symbolic thought emerged. After the emergence of cross-modular binding, an attended object represented within a given module could activate unattended information represented within other modules, which could then be added to the object file. That is, an object file could hold not only the module-appropriate features of the object but also activations, associations, or referents active within unrelated modules. For example, if an object moved abruptly, it could remind one of something else that changed abruptly, and if it moved slowly, it might arouse an entirely different set of associations. In this manner, a given event or object represented within, say, the visual or auditory module could come to be associated with arbitrary activations within other modules.

Prior to the emergence of cross-modular binding, our ancestors attended in a "literal" fashion to information about objects carried by neuronal firing within functionally isolated modules. After genetic mutation(s) in genes presumably encoding aspects of how neurons form into specific neuronal circuitry during development, binding became cross-modular. A given attended stimulus could automatically evoke neuronal firing on unrelated modules, or within neighboring nodes within a module, which could in turn be attended or not. Thus, in addition to being capable of being reminded of something by something else superficially quite different, the human mind is inherently prone to making semantic associations

as one activation triggers another in nodes connected to it within a network of semantic representations. In order to protect itself from associational breakdown, human mental architecture had to develop mechanisms to suppress meaningless associations and enhance meaningful associations by suppressing and enhancing the degree of firing of entrained neurons within unrelated modules or within connected nodes. Thus, human object files may differ from those of other animals not only insofar as containing operands and operators from different once-encapsulated modules but also in their capacity to overcome or preempt distraction and in their capacity to track. One possibility is that certain executive circuitry evolved to play this role.

Cross-Modular Operators and the Birth of Human Aesthetics

To reiterate, in addition to permitting symbolic thought, a key consequence of the emergence of cross-modular binding was that operators could be bound with new types of operands, such that operators from one module could operate on the operands of previously encapsulated modules. That is, not only concepts and percepts but also operators could be downloaded into an object file. As already mentioned, nestable motoric operators could operate upon symbols rather than just physical actions, giving rise to the possibility of syntax. Another example would be that operators designed to decode the emotional content of vocalizations could now operate over nonvocal sounds, giving rise to the possibility of music. And basic aspects of art and human aesthetics may have emerged as a consequence of operators that became "disencapsulated" or "universalized." For example, operators that evolved to discern genetic health in a potential mate could now operate over visual scenes, which could then attain the status of eroticism or beauty, although obviously a scene cannot be a mate. According to this view, a macaque can assess the emotional content of a vocalization or the beauty (i.e., genetic health and fertility) of a potential mate's face and body but could not apply those operators to things other than vocalizations and bodies. Operators in a macaque are dedicated and domain-specific operators, whereas in humans they became domain general. A macaque or dog or cow can hear sounds but cannot experience music.

Because cognition and behavior leave no direct physical record, it is impossible to know the first consequences of cross-modular binding in our ancestors who first had the mutation(s) that led to this new form of binding. It could be that behavioral changes occurred immediately or that they built up over hundreds of generations as natural selection acted upon

them. Some degree of speculation is unavoidable when hypothesizing about cognitive and behavioral changes brought about by connectivity changes in our ancestors' brains. However, we can examine types of cross-modular binding in modern human behavior for possible clues to the changes that may have occurred in our ancestors' minds and behaviors. Modern human behavior is rich with evidence of cross-modular binding, and it is this evidence that ties together numerous human behaviors that seem unrelated.

Of course, encapsulated modular operators could have evolved to great levels of complexity while still being limited to a particular class of operands. For example, it is possible that the mental operator that permitted the internal visualization of shape required for stone tool production developed in complete isolation from other mental operations. It could be that it was limited to the operand of stone. Indeed, it is surprising that no bone or antler tools are found until very late in the fossil record. However, once made domain general, this operator could be applied to very different operands than stone and perhaps could be used to generate more than tools. Once made domain general, such a capacity of three-dimensional visualization could permit visualization of forms and combinations that could not exist and thus could have played an important role in the birth of human imagination.

Cross-Modular Operators and the Birth of Abstract, Religious, and Causal Cognition

The essence of abstraction is the ability to detect patterns that are not evident in the immediate sensory input. That is, abstraction requires the representation of currently unobservable entities. Pattern extraction is common to the perceptual systems of many animals, but in nonhuman animals, pattern extraction may be modularly isolated from the other contents of cognition and, thus, domain specific. Once modularity broke down because of cross-modular binding, the computations underlying pattern extraction and recognition could be applied to data that were not perceptual in nature. An example of perceptual pattern recognition that became abstract pattern recognition in humans involves the inference[5] of causality. Other animals may be able to detect patterns of cause and effect in the flow of sensory input, because they carry out computations over sensory input that evolved specifically to solve this task. However, other animals may not be able to detect causal relations that are not evident in current sensory input. Because animals seem to be limited to causal pattern extraction from the current flow of sensory input, they may be

limited to detecting physical causal relations that arise from spatiotemporal contiguity and simultaneity. In the human lineage, however, the computations dedicated to the extraction of patterns of causal relation in the sensory domain came to operate over the contents of other modules, permitting the extraction of patterns of causal relation that transcended spatiotemporal contiguity. For example, once causal relations could be detected in the contents of memory, patterns of cause and effect could be recognized that transcended the here and now. One could become aware of a time before one's life or after one's death. One could entertain the fact of one's impending death. Simply put, whereas other animals can "connect the dots" perceptually, they cannot do so abstractly. Humans, in contrast, came to connect the dots to such an extent that hidden causes such as invisible beings (e.g., gods, demons) and modes of causality (e.g., karma, luck, the evil eye, curses, prayer, fate) were invoked to explain visible events, such as the weather, or life and death. Religion began in our lineage when pattern recognition operators became domain general. For example, the operator that determines whether sensory input comprises an animate being is an important operator for all mammals, since animals can be predators, prey, or mates. After cross-modular binding permitted this operator to become domain general, animacy and associated mental traits such as volition, personality, and intentionality could be seen in many patterns of events, even those that we now take to be patently nonanimate. Once the human mind became symbolic, it became a "congenital animist." Another example would be that operations that evolved for perceptual grouping on the basis of similarity, or the segregation of figure from ground, came to be applied to the contents of other modules with the breakdown of the modular mind. Once patterns of similarity/difference could be detected in information that was not perceptual, the human mind gained the capacity to become "unglued" from the here and now, free to connect the dots in novel ways. In fact, however, our ancestors probably connected the dots in characteristically animalistic ways, for example, tending to place events in a framework of causality governed by invisible animate beings simply because we had inherited this way of organizing events from our presymbolic ancestors, who organized events in terms of the actions and intentions of other animals. We generalized such tendencies to beyond the here and now. Human religiosity emerged with symbolic thought by applying an ancient and presymbolic way of inferring causality and perceiving patterns in the flow of events to operands that went beyond events in the currently perceived environment.

A theory of mind may have emerged with the capacity to think abstractly. As Povinelli et al. (2002) have written:

> (A)lthough chimpanzees are excellent at exploiting the observable contingencies that exist between the facial and bodily postures of other agents on the one hand, and events in the world on the other, these animals may not construe others as possessing psychological states related to "seeing" or "attention." Humans and chimpanzees share homologous suites of psychological systems that detect and process information about both the static and dynamic aspects of social life, but humans alone may possess systems which interpret behavior in terms of abstract, unobservable mental states such as seeing and attention.

In order to represent others' invisible mental states, one must be able to detect patterns, including patterns of causality, in invisible, nonperceivable entities. Once the ability to think and perceive abstractly evolved, patterns of invisible behavior, such as thought, intention, perception, and attention in others, could be recognized and inferred. Before the emergence of abstract thought afforded by the application of computations originally dedicated to perceptual pattern recognition, no true theory of mind was possible.

Changes in the Nature of Play

The nature of play would also have changed as a consequence of cross-modular binding and the emergence of symbolic reasoning. In nonhuman mammals, play involves pretending to enact adult behaviors, such as fighting and hunting. In humans, however, play involves not only this but also pretending to be something that one is not. Because one can oneself symbolize anything, one can pretend to be, for example, an animal, another person, or an object. One can even pretend to be something that does not exist, such as an elf or an egg with wings. There are no examples of this kind of symbolic play in other mammals because no other animals think symbolically or even analogically, since both types of thought entail a breakdown in modularity. Once the body could be used as a symbol, the body could be used to communicate. Donald (1991) has suggested that our ancestors went through a stage where mimesis was the primary means of communication. Once our ancestors understood that objects could have arbitrary referents, the body itself could be exploited as a symbol to convey arbitrary referents. Thus pointing, mimesis, and pretend play may all have emerged with symbolic thought and cross-modular binding and operations.

Symbolic Thought and the Evolution of Human Morality

Symbolic Thought Gave Rise to Morality

The emergence of symbolic thought had profound consequences for human moral cognition. You might say that the birth of symbolic thought gave rise to the possibility of true morality and immorality, of good and evil. Once acts became symbolized, they could now stand for, and be instances of, abstract classes of action such as good, evil, right, or wrong. Symbolic thought permitted new dimensions of behavior, for example, the expression of territoriality over the ownership of an idea rather than just a concrete thing such as turf, a bone, or a mate. Thus, while a monkey has affection, social intelligence, likes, dislikes, fear, inhibitions, territoriality, deceit, aggression, vengefulness, and other predispositions that govern behavior, these are not morality. The more prosocial of these modes of feeling and thinking can be thought of as "preadaptations" for morality, but they are not morality. A monkey lacks moral judgment, prohibitions, norms, principles, laws, approval, disapproval, injunctions, the concept of good, the concept of wrong, virtues or vices. A monkey lacks morality because it lacks symbolic cognition. This is why animals in general are amoral rather than immoral or moral. They cannot conceive of doing the right thing or the wrong thing. This is why we do not put animals in prison when they do something we do not like or which violates our sense of what is good or right. This is why we do not reason with animals and would not even if we could communicate with them at the level of concepts. When they do something that offends us, either we stop them, or we punish them in a way that they understand, usually by doing something that they find aversive as soon as possible after the offending act, so that we can build up in their minds certain associations between certain types of behavior and certain consequences.

Moral and Immoral Acts as Instances of Symbolic Categories in Minds

An act becomes immoral for us, and we disapprove of it accordingly, because it comes to symbolize or stand for other similar acts and thereby becomes a member of the abstract category "bad," "wrong," or "evil." That is, morality is rooted in both our capacities to symbolize and to generalize to a level of categorical abstraction. A person, once symbolized, has done more than stolen a particular piece of, say, meat. He has become a thief. The piece of meat is not just a particular possession, but an instance of the class of things called "property," which has associated rights. Once taken, it is stolen property that happens to be a piece of meat. Once symbolized,

individual acts and things can become instances of classes of events or classes of things that are good or evil, acceptable or unacceptable, sanctioned or not.

Consider the nonsymbolic mind of a little dog. How does he regard the bone that he is gnawing? It is his because he has it, not because it is his property or his in some other abstract or symbolic sense. He knows that other dogs want to get his bone, and he does his best to stop them. If a big dog comes along and forces his bone away from him, it is no longer his bone because he no longer has it. It is the big dog's bone because now the big dog has it. The little dog might be angry and frustrated that he lost his bone, but he cannot think "the big dog has my property" or "I own that bone regardless of who has it now." He can think "the big dog has a bone that I had, and I want that bone back."

In contrast, I can own an object regardless of who has it now. This is because I and other members of my society can place a representation of me along with a representation of the object in question in a common object file. A dog cannot do this, and as far as we know, no other animal can either. Humans are so advanced that they can jot symbols down as words, which can then create such an object file in the mind of a reader at any future time or place. Writing and other forms of data storage create the impression that ownership takes on a special kind of formal existence that would continue even if all people were killed. In fact, ownership (or any other abstract or legal concept) only exists insofar as such object files can be created in my mind or some other human mind. If I lose the ability to place a representation of myself along with a representation of my house in a common object file, do I own my house? Well, yes, in a legal sense, but certainly not in my own mind. If all people forever lose the ability to create such a joint object file, do I own the house? In this case, the answer is no. A legal/moral status, such as ownership, is not platonic. It requires the processing of symbolic minds and only exists in such minds.

The Domain of Moral and Immoral Acts Can Be Symbolic

Another effect of symbolic processing is that acts can be committed in the symbolic domain. For example, if I hate Mr. X, rather than attack him, I might slander his corporation or theory, because his corporation or theory stands for him. The ancient impulse toward territoriality can now be expressed and triggered over abstract representations such as infringement of perceived national interest or theft of intellectual property. The ancient impulse toward aggression can now be triggered by threats to abstractions

such as one's nation. One can express aggression not only to individuals in the symbolic domain but toward abstractions such as social movements or ideas. One can hate slavery, or a slavish mentality, or one can love the teachings of Buddha. One can hate a mentality and yet not hate the person who is possessed of that mentality. None of this is possible for nonhuman animals.

Because the human mind is capable of virtually unlimited power and freedom in the domain of symbolic, abstract thought, there has been a tendency throughout history to devalue that which is not symbolic, in part because the nonsymbolic often limits our power and freedom. This is especially true of the body. The body, with its filth, lusts, and mortality, is rejected for its impurity and decay. It comes to symbolize impurity and death. Its urges are labeled immoral by many religions. It is viewed as that which must be conquered in order to be moral, spiritual, or pure. Many Christians view sex and masturbation as evil for this reason. A similar mind-set leads many religious men to reject women, with their monthly bleeding and violently birthing bodies, as inherently impure or even inherently immoral. Add to this rejection of the female body their own suppressed lust for such bodies, and you have a sure recipe for the neurosis, hypocrisy, and misogyny apparent in many men who assert their higher morality. Even in religions that embrace the body and its urges for pleasure and fulfillment, such as paganism, hedonism, or bacchanalianism, the body is symbolized. In these cases it comes to stand for something good, namely, life, the divine, or love. A fundamentalist Christian will typically reject people who embrace the body in this way as decadent or evil. Similarly, a person who embraces the body will view the Christian suppression of the body as wrong or evil. Who is right? The answer is that no one group has succeeded in imposing its morality upon all other people, causing untold suffering and death, although many have tried. The reason that no one has succeeded is that one's morality, like one's religion, and the God that one worships, is chosen based on what feels good, comfortable, and right. One's moral system largely follows one's taste and personality. And personalities differ.

In this regard, William James had a deep insight into human nature. In his book *Pragmatism and the Meaning of Truth* (1907/1908) he outlines a philosophy of truth built upon that of Charles Sanders Pierce, called "pragmatism." According to this view, truth is not what philosophers have traditionally thought it is. It is not the accurate correspondence between a proposition and that to which it refers in reality. It is also not coherence of meaning between a proposition and a model or theory in which that

proposition sits. Rather, for ordinary people, practically speaking, truth is the cessation of doubt. James asserts that for most people, the feelings of uncertainty and doubt are unpleasant. So people search for the end of their doubts. They search for certainty. And when they have locked onto some belief that they are certain is correct, they feel comfortable, vindicated, set free, and, sometimes, divine. They feel safe and saved. This arrival at safety and certainty is what "truth" means in James's theory. Once locked into a belief system, the only threat to personal safety and comfort is the doubt of others. This accounts for the vehemence of many proselytizing zealots. They are proselytizing not for God or for the sake of the people they desire to save, though they may believe this. According to James, their vehemence and conviction comes from their need to be certain, which in turn comes from their own discomfort with doubt and dread. In contrast, James points out that there are certain types of people, often writers, artists, thinkers, explorers, or scientists, who are at home in doubt and uncertainty and who actually enjoy exploring the new possibilities associated with the unknown. Such personalities reject rigid dogmas. For them, truth is not the cessation of doubt but the discovery of what is. Whereas the former personality type might adopt an off-the-shelf fundamentalist belief system, the latter personality type will reject such dogmas and perhaps construct one of their own. Or they might not adopt one at all, since such personalities feel comfortable in admitting that they do not have all the answers.

Just as the type of truth that one embraces is rooted in one's personality dispositions, the morality that one embraces and attributes to God or the universe, or that one formalizes in an "objective" body of enduring law, is an external projection of private intuitions and desires. One starts out with these intuitions and desires and then builds up theories that justify them. It is not surprising, for example, that John Rawls, in his theory of justice, ends up with an ideal governmental and legal model that looks pretty much like American Jeffersonian democracy, because this is the intuitive endstate that he wanted to justify with his clever arguments and rationales. When he presents his theory, it seems that the conclusion emerges from the reasoning, but in fact the opposite is the case. Any moral or legal system constructed on the basis of reason alone which did not feel right would be rejected by most people, no matter how rational. Human moralities, in short, are rooted in irrational urges and clothe themselves in rational justifications from which they claim to derive in order to disguise this. The reason that we have conflicting moralities and conceptions of good and evil is that we have conflicting desires and feelings about the

various domains of choice and activity that moralities seek to govern. What all human moralities have in common is that they are symbolic. Conservative Christians may symbolize the body as filthy and in need of suppression. Pagans may symbolize the body as a fount of life in need of expression and fulfillment. However, to neither group does the body symbolize nothing as it does for a dog. For a dog, its body is just its body, representing nothing. For humans, such a state may only exist in infancy before the advent of language and pretend play.

Understanding the Multiple Symbolic Roots of Human Evil

In order to talk about morality, it is necessary to talk about good and evil, because morality is fundamentally concerned with maximizing good and minimizing evil. It is doubtful whether "evil" can be defined in the abstract, in a manner free of any individual's frame of reference. Here "evil" will be operationally defined relative to a subject's mind to include that which harms or could harm that mind or that which that mind cares about. "Good" will be defined in a similar way to include that which benefits or could benefit a mind or that which that mind cares about. These definitions are not perfect. For example, one reasonably could argue that an act can only be good or evil if it is volitional or intended or at least the act of a conscious being. One could argue that only acts that have consequences are good or evil, or one could argue that thoughts and intentions, even in the absence of enactment, can be good or evil. However, no single conception can fully account for the numerous, often inconsistent ways that the terms "good" and "evil" are used. For present purposes, these definitions will suffice.

Because human acts and ideas operate over symbolic representations, we are capable of acts and ideas that no other animal could conceive of, let alone enact. No animal could conceive of killing all individuals who believe X or look like Y. We are capable of wanting to destroy all individuals of one group because those individuals are not truly individuals for us. They are symbols. They stand for something that we find abhorrent and want to eradicate. Since all individuals of this class stand for the same thing, they are variants of the same symbol. What we are doing when we view individuals as variants of the same symbol is this: we are tokenizing them. They are not individuals but tokens that stand for something evil that must be destroyed. Since we cannot eradicate a concept, such as that which they stand for, say the concept of "greed" or "woman" or "evil" or "deceit," we opt to eliminate the symbols. Thus, one symbolic root of human evil lies in tokenizing individuals and treating them as symbols,

when they are in fact individuals with whom we might have shared a friendship under other circumstances.

Beyond merely stopping undesirable behaviors, the way to confront this source of human evil is to detokenize people in the minds of the people who tokenize them. This is more easily said than done, but it can be done. It requires a change of mind and heart. It seems that the human brain tends to minimize effort whenever possible by dealing with people, objects, and events at the level of tokens of preexisting categories or symbols that stand for preexisting concepts. For example, the upshot of many recent experiments involving change blindness is that people do not notice that a person has been replaced by another person unless that person switches categories, especially the categories of gender, race, and age, and unless they paid attention to the person and encoded their individual characteristics rather than just encoding them as the instance or token of a preexisting category. Great spiritual leaders of the past have warned against operating within this zombie-like state of tokenization, in part because it can lead to acts of stupidity and evil. They have emphasized the need to transform one's mind by cultivating compassion and love for individuals and by paying attention to the particularities of events.

Why does the brain tokenize? Minimization of effort is part of the explanation, but not all of it. There may be parts of the brain, such as certain nuclei of the amygdala, that are designed to operate at a level of tokenization in order to rapidly detect imminent threats. This brain structure may have a high rate of false alarms, but it exists because it is better to make a false alarm for certain kinds of threat than it is to make an incorrect rejection, which can result in being eaten or otherwise killed. Indeed, it is probable that the amygdala is hardwired to carry out certain kinds of stereotyping and certain types of stereotypical false alarms. This is why people across cultures tend to mistakenly see people or animals in their peripheral vision, when it was in fact only a human-like shape. It is also why people rarely experience a false alarm on things that could not be an imminent threat. It would be a strange mind indeed that often mistakenly saw a distant mountain range or constellation out of the corner of its eye. There are still other reasons why the brain tokenizes. We are taught to tokenize. We mimic and learn from those around us, and if they are operating at this level, we will probably learn to do so too, unless we are especially sensitive individuals. Another reason appears to be that tokenization and stereotyping allow us to ignore a great deal of irrelevant detail. As such, tokenization is probably a necessary cognitive strategy to avoid sensory and attentional overload.

Another symbolic source of human evil lies not in tokenization but in sadism. No animal could take sadistic pleasure in generating psychological torment in another mind, because this requires a theory of mind, which even chimpanzees appear to lack. As described earlier, the emergence of a theory of mind only became possible with the emergence of symbolic and abstract thought. The sadist internally models the mental state of his victim and draws pleasure from the power he exerts over his victim. Whereas a nonhuman animal could conceivably enjoy being in a dominant position, and may even gain pleasure from hurting another animal, it cannot gain pleasure from the psychological torment of its victim, because psychological torment is invisible and cannot be represented by nonsymbolic minds.

Interestingly, psychopathy, which can be thought of as a lack of conscience and empathy, may emerge in part from an inability to conceptualize the pain of another as if it were one's own. As such, psychopaths may be the opposite of sadists. Whereas a sadist savors the pain in another's body and mind that he is internally modeling in exquisite detail, the psychopath is incapable of this internal modeling, so operates as if other people do not feel pain as he sets about fulfilling his selfish goals. The end result, from the victim's point of view, may be similar, however. While a sadist is no doubt fully responsible for the pain he inflicts, a case could be made that a person who truly lacks a conscience is not morally responsible. But given our operational definition of evil, from the point of view of the victim, there is no difference.

Yet another source of human evil lies in conceptualizing oneself as not really doing the acts that one is doing or conceptualizing oneself as not responsible for the acts that one is committing. One is just obeying orders. One's acts are not the product of one's own volition. Hannah Arendt has used the term "the banality of evil" to describe this source of human evil. Since no one feels responsible or in charge, no one stands up to stop the evil. A more modern example would be a corporate president who knowingly engages in illegal logging in a tropical rainforest, because he has to feed his kids, and anyway, if he did not do it, someone else would.

It is commonly said that the root of all evil is money or, perhaps, greed or selfishness. However, other animals are greedy and selfish about things like food, territory, and sex. What is different about human greed/selfishness is that it is symbolic. Rather than merely wanting to maximize our share of territory, sex, and food, we want to maximize the multitude of things symbolically linked in our minds with these things. Whereas an animal might horde nuts, we might stuff our résumé with honors and

prestige. Whereas an animal might try to increase its attractiveness to a potential mate by building a quality nest, we might attempt to accomplish the same by having the right career, or right opinions, or right friends. Whereas an animal might try to maximize its position in the social hierarchy, we are driven by a comparable urge for dominance but express it as a maximization of power, prestige, or honor. In general, human symbolic cognition operates on top of numerous desires and emotions that we share in common with our nonsymbolic primate relatives, including, for example, lust, fear, aggression, greed, deceit, social scheming, territoriality, jealousy, rage, and a desire for revenge. The potential for evil emerges in the symbolic enactment and fulfillment of these and other nonsymbolic urges.

Although there are several more sources of human evil that could be mentioned, the final one to be considered here is culture. Because humans are symbolic, they can have symbolic culture, in which information, behaviors, and attitudes are passed from generation to generation through symbolic modes of transmission. A culture itself can become a source of evil when it fosters information, behaviors, or attitudes that lead to evil intentions or actions. Although modern American capitalism has generated many goods for many people, and although there have been far worse cultures in the past, American culture is the dominant culture now, so a few words on its potential for generating evil are appropriate. There are certain aspects of American capitalism that can be characterized as harmful to individuals. There is a tendency in America to believe that markets are the best mechanism for making decisions. However, market mechanisms are only a way to decide prices. There are decisions that markets cannot make. Markets cannot decide what is good or beautiful or right, for example. Because of the emphasis on market mechanisms, these other things sometimes get short shrift as the majority of people pursue profit. Billboards, development, and resource extraction, processing, and use undermine beauty, health, and the environment. More insidious, the attitudes of market capitalism come to pervade many aspects of life that in the past would not have been regarded as a domain of buying and selling. There is a distinct danger in seeing oneself as an object to be sold. Once this mentality is in place, it is a short leap to wanting to improve the "product" by getting bodily implants or surgery or by continually trying to look better on paper. Seeing oneself or others as commodities to be easily replaced when used up is to see them as means to ends rather than individuals who are ends in themselves. Such a mentality, fostered by a culture, can lead to various immoral acts, such as using people to get ahead or knowingly polluting.

Understanding the Multiple Symbolic Roots of Human Goodness

If symbolic cognition affords humans the possibility of being and doing evil, it also affords them the possibility of being and doing good. Just as there are many ways to harm a mind or that which it cares about, there are many ways to benefit a mind or that which it cares about. Most of the ways of being or accomplishing good are the opposite of the above ways of being or accomplishing evil. One can be good by being kind, just as one can be evil by being unkind, the extreme of which is sadism. By its very nature, kindness occurs best when one is not tokenizing the recipient of one's kindness, since only then can one tailor one's kind acts to the particular needs of an individual. Indeed, only if one is attending to the changing needs of the recipient of one's goodness can one continually tailor one's actions to those changing needs. It requires internal modeling of the mind of the recipient. Kindness is therefore not even possible for animals, because they lack a theory of mind. Animals can be affectionate, but they cannot be kind.

It thus appears that doing and being good requires the cultivation of a certain kind of mentality, not just a certain kind of behavior, since good behaviors follow from the mentality and less so vice versa. This mentality is one that is kind and compassionate rather than cruel, attentive rather than inattentive, and focused on the individual rather than on the class or token. Just as evil can emerge in the symbolic realization of desires and urges that we have in common with animals typically centered on aggression, dominance, territoriality, and sex, good can emerge in the symbolic realization of other desires and urges that we also have in common with animals and which are not themselves symbolic. Among these would be urges and desires for affection, love, community, compassion, nurturing, protecting, parenting, commitment, and bonding. However, it would be simplistic to say that evil emerges from the symbolic enactment of one set of desires and good emerges from another, mutually exclusive set. In fact, good or evil can arise from the symbolic enactment of any one of these urges and desires. For example, overprotectiveness can harm the recipients of this excess of love by making them dependent and weak. A lack of aggression can lead to harm, as one fails to protect oneself or others whom one loves. Thus, affection can be expressed in an immoral way, and aggression can be expressed in a moral way. Morality should therefore not be equated with the enactment of prosocial desires. Immorality should not be confused with the enactment of destructive or "base" desires.

All evil and all good derive from the fact that our minds are symbolic. Even though being symbolic has cursed us with so much past and present

evil and so much potential for more evil, being symbolic has blessed us with the ability to choose what is right and good. Our symbolic mind equips us to actively fight those possessed of evil mentalities. The only salvation from human evil lies in human goodness. No one will rescue us from our own minds. There is hope for us because human minds can be changed. Humans absorb and create culture and can be taught to act and think differently than they now do. This is especially true of children. However, you cannot force a change in mentality by threatening, coercing, or killing people. You can change behavior with violence or the threat of violence, but this lasts only as long as the threat is in place. A change in mentality, with its consequent changes in behavior, is enduring and only emerges as one symbol structure comes to replace another in a person's mind. The tried and true ways to accomplish such a shift are education, persuasion, and inculcation. Mentalities can change for the better in more than just incremental ways if the correct approach is taken. The only mind that we have complete access to is our own. It is here that there is the greatest hope of a transformation. But it takes a great deal of effort.

Notes

1. Information processing in animals has undergone several revolutions, where the type of information processed after the revolution differed in kind, not just degree, from the type that preceded it. Although there may be more, at least four revolutions stand out as apparent. (1) Reflexive information processing: The earliest information processing probably involved little more than detection of ambient chemical or electromagnetic energy. This detection triggered behaviors in an automatic, reflexive, and stereotyped fashion. A given input had a single possible output. Even after the evolution of multicellular organisms with neurons, reflex probably remained the dominant mode of information processing in animals. (2) Nonreflexive information processing: The second revolution occurred when nervous systems were able to represent more than one possible course of action in response to a given input. This mode of information processing required the development of memory buffers where possible courses of action and their consequences could be modeled, stored, and compared. This class of information processing reaches its highest complexity in the higher mammals, where complex scenarios can be played out before any action is taken. (3) Analogical and symbolic information processing: This revolution occurred when links could be made between classes of information and types of information processing that were previously modularly encapsulated. (4) Syntactic processing occurred when operations could be carried out upon symbols in a generative and nested fashion. There are example where humans process input and

generate output in each of the four ways, whereas other animals generally lack the capacity for (3) and (4).

2. Other authors have written about the existence of domain-general mental capabilities that do not meet the criteria of modularity (i.e., fast, dedicated, mandatory, impenetrable, encapsulated; Karmiloff-Smith, 1995; Fodor, 1983, 2001; Gardner, 1993) which may emerge from the interplay of domain-specific types of modular representations and processes (Spelke, 2000). Fodor himself ends his 1983 book unable to account for slow, general-purpose reasoning and cognition in terms of his modularity hypothesis. Nonetheless, some authors, particularly champions of evolutionary psychology (e.g., Duchaine, Cosmides, & Tooby, 2001; Pinker, 1999) appear to focus primarily on the explanatory power of the modularity concept without emphasizing the prevalence of domain-general operators in human or animal cognition. Attention is an example of a domain-general operator, as is reason. The defining qualities of nonmodular processes are the exact opposite of those that define modularity. They are slow, not automatic, not dedicated, not mandatory, accessible to more than one module, and not encapsulated.

3. We say that attention binds the joint contents of the object file into a single bound representation, as if attention were a separate thing from the binding process itself. However, attention may just *be* binding. We will continue to use the word "attention" as shorthand, as if attention "does" binding, but really it might be more accurate to say that "binding takes place," rather than "attention binds."

4. Indeed, modular processing in general, which can be characterized as operand dedicated, encapsulated, parallel, and impenetrable to other modular or domain-general processes (Fodor, 1983), would help an animal survive precisely because it is rapid, and, being automatic and free from cross-talk interference from other operators or operands, less prone to error. Most types of recognition or response need to be done as rapidly as possible in a world where one can lose prey or become prey in an instant. Puzzling over the interpretation of a tiger or mistaking a shadow for a solid surface could result in death. The puzzle is not the existence of modularity in animals but the presence of relatively slow domain-general processes. There are presumably benefits to operations that amount to a slow, nonstereotyped consideration of options whose outcome is unpredictable. For example, always following the locally best option can lead to one's demise or getting stuck in a local minimum. One path may look pleasant and inviting, and another grim, but if the former leads to a tiger, and the latter to food, then the local characteristics are misleading. It is precisely this sort of internal generation and weighing of possible outcomes that likely led to the existence of domain-general operators. Rapidity and automaticity are a boon in the processing of local input, but when input must be integrated over multiple modules and over space and time, or when processing is open-ended (how long into the future should one consider before a path looks good?), then rapidity and automaticity are potentially fatal. Because domain-general

operators are potentially open-ended, they give rise to the danger of taking too long before a path of action is chosen. Building domain-general operations into a cognitive architecture also requires building in automatic resets, interrupts, and self-terminations of these open-ended processes. Such automatic resets may include forced reprioritization mechanisms such as exogenously driven attentional capture, hunger, pain, or fear, which would bring an animal "back to reality," and prevent it from getting stuck in an open-ended process.

5. There are two types of inferences that must be distinguished here. There are those that occur prior to consciousness and which are integrated into our conscious experience. Then there are those that are based upon the facts that present themselves in our conscious experience and which therefore follow consciousness in time. The former are largely generated by perceptual modular systems, such as the visual system, and are therefore automatic, rapid, and cognitively impenetrable. The latter are typically not perceptual but cognitive operations that are based upon facts. For example, when I go outside, I see that the ground is wet and infer that it must have rained. Seeing the ground as wet is an example of a preconscious or preperceptual inference. What I perceive has that inference built into it. Consciousness comes precompiled in the sense that I cannot choose not to see the ground as wet. I cannot choose to see the world as a collection of colors that have not been interpreted into materials, spatial layout, reflectances, shadow, lighting, and so forth. The world is presented to me as such by the visual module. When I infer that it must have rained, however, I am making a cognitive inference which is open-ended and slow. I can make countless inferences from the fact that the ground looks wet.

5.1 A Just-So Story for Symbolic Thought? Comment on Tse

Michael R. Dietrich

In this comment, I will consider Peter Tse's essay from the perspective of evolutionary biology. More precisely, I will focus on the nature of Tse's evolutionary claims and their justification.

During the debates over sociobiology in the 1970s, Stephen Jay Gould and Richard Lewontin characterized putative evolutionary explanations that were merely consistent with some of the principles of evolutionary biology as stories (later as just-so stories).[1] These evolutionary stories usually appealed to the power of natural selection and so were seen as an expression of adaptationism. The problem was that while these evolutionary stories may seem plausible and were often very imaginative, they were rarely justified evolutionary explanations. If evolutionary biology were reduced to a process of endless adaptationist conjectures, it would indeed be in danger of rendering itself an unfalsifiable morass. This does not mean that evolutionary storytelling has no value for evolutionary biology. As John Tooby and Leda Cosmides point out, the value of evolutionary just-so stories lies in their predictive utility, not their explanatory power. In their words, "modern selectionist theories are used to generate rich and specific prior predictions about new design features and mechanisms that no one would have thought to look [for] in the absence of these theories . . . "[2] With this history in mind, I will now turn to the evolutionary narrative offered by Tse.

In his essay on the origins of symbolic thought, Peter Tse has offered a compelling account of how cross-module binding could produce a wide range of cognitive consequences. He is clear that he does not want to get "too deeply into any 'engineering' solutions," (p. ••) and focuses much of his discussion on developing a cognitive theory that itself is presumed to not be tightly bound to any neuronal mechanisms (p. ••). Nevertheless, Tse does offer "some speculation" on genetic mutations that produced "the rewiring of neuronal circuitry that permitted cross-modular binding"

(p. ••). He even offers a prediction that "there are corticocortical, corticothalamic, and thalamothalmic connections between neurons within different modules in humans that are not present in chimpanzees or other nonhuman apes" (p. ••). As an evolutionary explanation, Tse's account of the origins of cross-module binding is a just-so story. It is consistent with the principles of evolution by natural selection, even if it is acknowledged as speculation. The question remains, however, is this a fruitful speculation from the perspective of evolutionary biology?

Tse recognizes that we cannot know the consequences of cross-module binding in our ancestors. In other words, we cannot know how natural selection or sexual selection acted on the initial effects of cross-module binding. Evolutionary conjectures concerning the origins of cross-module binding are, therefore, unavoidably speculative (p. ••). Yet, the presumption is that because "human behavior is rich with evidence of cross-modular binding" (p. ••), human anatomy will show evidence of cross-module binding, and that those anatomical structures will have been positively selected for for tens of thousands of years and that what we consider to be moral behavior is one of the many products of that process of selection. Note that no evidence of anatomical features is provided, and no specific connections are made been neuronal structures and their genetic foundation. Moreover, even if we accept the evolutionary story, it does not entail that morality itself was the object of natural selection, then or now. Many of the other features of this scenario could be the actual targets of selection, and morality could simply be along for the ride, so to speak. The connection between morality and evolution is another aspect of the story, but it is a conjecture. The plausibility and unifying power of the story should not be considered to justify it from a scientific point of view.

Evolutionary explanations lay in the engineering details. Indeed, recently many of these details are being worked out in studies linking brain development and evolution and in the effects of specific genes, such as FOXP2.[3] Tse's top-down theorizing may be useful and is certainly interesting, but without more work from the bottom up it will not qualify as an evolutionary explanation. Tse has provided an evolutionary conjecture. An interesting conjecture, but without more biological detail the cross-binding hypothesis cannot be evaluated as an evolutionary explanation. Instead Tse's paper is indicative of how much science remains to be done to link proposed "cognitive architectures" to genetically grounded, evolving biological systems.

Notes

1. Gould and Lewontin (1979).
2. John Tooby and Leda Cosmides (1997, July 7).
3. Gilbert, Dobyns, and Lahn (2005); Marcus and Fisher (2003).

5.2 Morality and the Capacity for Symbolic Cognition: Comment on Tse

Kathleen Wallace

In this comment I will briefly summarize Tse's theory of mind but otherwise take it for granted. The discussion of the empirical research and its implications for philosophy of mind and cognitive science I leave to others qualified in those areas. My focus will be on the claims made about morality.

Tse suggests that the roots of human good and evil lie in symbolic processing, which is accounted for by the breakdown of strict modularity in the mind. Cross-modular binding accounts for the emergence of abstraction and analogical, metaphorical, and symbolic thought. Moral judgment involves prohibitions, norms, principles, laws, and the concepts of good and bad. It is symbolic because it requires the ability to abstractly represent an act or object, that is, the ability (1) to conceptualize an act or object in terms of abstract categories (e.g., "theft," "property") and (2) to place an act in an abstract category of good or evil, right or wrong (Tse, p. ••). Cross-modular binding and the emergence of symbolic processing also account for the extent of possible human evil (or good) through the process of tokenization, through the capacity to represent another mind, and through the cultural operation of symbols. On tokenization, Tse says,

> Because human acts and ideas operate over symbolic representations, we are capable of acts and ideas that no other animal could conceive of, let alone enact. . . . We are capable of wanting to destroy all individuals of one group because those individuals are not truly individuals for us. . . . They stand for something that we find abhorrent and want to eradicate . . . they are variants of the same symbol . . . We are tokenizing them. They are not individuals but tokens that stand for something evil that must be destroyed. (p. ••)

Tse explains sadism as rooted in the ability to have a theory of mind about another individual and suggests that the capacity to internally model

the mental state of another person depends on the emergence of symbolic and abstract thought (pp. ••–••). Culture, or the ability to communicate symbolically, also allows human beings to abstractly represent, for example, to commodify, individuals, *and* to extend that commodification across many individuals and perpetuate it across many generations (Tse, p. ••).

That morality requires the ability to symbolize and to abstractly represent objects, persons, events, and actions would not distinguish morality from many other distinctive human functionings—art and aesthetic judgment, scientific reasoning, and rule-governed play, to name a few—all of which equally involve capacities for abstraction, generalization, representation, tokenization, and modeling. For moral theorists, a claim that morality involves a capacity for abstract or symbolic thought, in at least some respect, seems relatively uncontroversial. Even emotivist moral theorists who think that moral judgments are simply expressions of emotional states could allow that their range is extendable by the possibility of symbolic cognition without giving up the claim that the judgments are, at bottom, emotional states. Particularists might object to characterizing good and bad acts in terms of abstract categories because that presumes too much generalization.[1] However, the claim that recognition of an act as theft or an object as property, as well as the recognition of "good" and "bad" as moral tags, would still seem to involve symbolic cognition, and particularists admit these concepts as necessary for moral judgment. Moreover, symbolic cognition could be invoked as part of an explanation of both a particularist's and a generalist's claims. Regarding particularism: if symbols are flexible and changeable in meaning and if moral judgments involve concepts that are symbolic in nature, then the flexibility of symbolic cognition might be part of the explanation of moral judgments that do not depend on principles.[2] Regarding universalism: if categorization and generalization depend on symbolic processing, then such cognition might be part of the explanation of how moral judgments do rely on principles. In either case, a capacity for symbolic cognition by itself does not appear to favor one or the other theory of moral judgment.

If symbolic processing is distinctive of human beings and is a necessary feature of morality, and if animals lack the cross-modular binding required for symbolic processing, then the theory would give a naturalistic explanation for the exclusion of animals from having moral duties.[3] There are some who suggest that the exclusion of animals from being moral agents is arbitrary.[4] It might be noted that Tse's claim turns not on animals versus human beings but on the possession of the capacity for symbolic cognition as a necessary condition for making moral distinctions.[5] If there were

animals or any other beings that have a similar capacity, then they, too, would have to be regarded as capable of morality. However, even if animals are not capable of making moral distinctions due to the lack of a capacity for symbolic cognition, that would not entail that they couldn't or shouldn't be included in the sphere of beings for which a moral agent should have moral concern; nor would it entail that they couldn't have rights.[6]

There appears to be some slippage in Tse's account of how the capacity for symbolic cognition contributes to morality between the claim that symbolization and categorization is necessary for the possibility of moral categories at all and the claim that symbolization, particularly in the form of tokenization, is the root of immoral acts. Tse suggests that an act becomes immoral for us "because it comes to symbolize or stand for other similar acts and thereby becomes a member of the abstract category "bad," "wrong," or "evil" (p. ••). But good/right involves, he says, a kind and compassionate mentality that is "focused on the individual rather than on the class or token" (Tse, p. ••). The suggestion seems to be that tokenization itself is the basis for evil or a morally bad attitude towards others, whereas a morally good attitude toward or treatment of others involves treating them compassionately as individuals.[7] However, this is not entirely consistent with other features of the theory. For instance, if categorization is a necessary feature *of morality*, then a morally good act should be good insofar as it symbolizes or stands for other similar acts and thereby becomes a member of the abstract category "good" or "right."[8] Moreover, recognizing a human being as a member of a rights-bearing class would seem to be an instance of tokenization, but one which recognizes or creates a positive moral category. Presumably such "tokenization" would be involved in recognition of a duty to respect the rights of persons, even when a particular person is someone for whom one has little or no feelings of kindness or compassion. Ignoring individual characteristics (tokenization per se) by itself is not evil, for sometimes ignoring them, or at least not taking them as *the* basis for moral choice, is the morally right thing to do. Tokenization by itself is morally neutral; whether it is morally objectionable, or morally desirable, will depend on other features of the attitude, action or behavior and not just that it is an instance of tokenization.[9] If there are some *specific* types of tokenization that distinctively contribute to immoral actions, further study is needed to determine that. If there are conditions under which tokenization leads to evil behavior, a further question would be whether they count, noncircularly, as ones that constitute the evilness of tokenization.

Tse suggests that the capacity for compassion focused on an individual requires the ability to model another mind and, therefore, involves symbolic cognition. First, compassion: Tse's characterization seems to assume that it is, in and of itself, morally good, and that is, at the least, debatable. Compassion neither necessarily leads to moral behavior nor is necessarily a moral attitude. One can also feel compassion focused on an individual and yet be morally obligated to act against what compassion might recommend. For instance, in serving on a jury, one might feel compassion for the defendant yet be morally obligated to honestly render a guilty verdict. Of course, compassion may be moral and may be important to the development of moral sensitivity and judgment. But compassion per se is neither necessary nor sufficient for the moral goodness of an action or outlook, nor does it always contribute to a particular action's being moral even when it is present.

Second, "focusing on the individual": this is no more likely to be morally good than treating someone as a member of an abstract class. There may be times when it is morally preferable to regard someone as a member of an abstract class, such as "rights holder." Moreover, "focusing on the individual" could just as easily constitute morally reprehensible behavior. For example, Tse suggests that sadism involves deriving pleasure from the psychological torment of the particular victim. In modeling the mental state of a victim (Tse, p. ••), the sadist is "focused on the individual."

The conjunction of compassion and focus on the individual is not necessary or sufficient for an action to be judged morally good. The juror mentioned earlier may be morally obligated to disregard her experience of such a conjunction and to act on the basis of legal principles and a duty to evaluate evidence.

In the discussion of a contrastive case to sadism, namely, psychopathy, Tse suggests that a psychopath lacks the capacity for internal modeling and thus may not be morally responsible for her or his actions, even though from the point of view of the victim, the sadist and the psychopath are equally morally objectionable (Tse, p. ••). However, if symbolic cognition is what distinguishes human cognitive capacity from that of other animals and explains why human actions and behavior are judged in moral terms, then on this account it seems odd to say that a psychopath's behavior is *evil* or *morally* objectionable and an animal's infliction of pain on a human being by mauling the human being is not. On Tse's account, each (the psychopath and the animal) would not be morally responsible because each lacks the appropriate capacity for symbolic cognition. The argument might be that other human beings who do have the appropriate capacity

for symbolic cognition regard the psychopath and the animal differently and that's why the one is morally objectionable and the other isn't. But the question is, why do other human beings regard them differently? It would seem odd to say that, on the one hand, the capacity for symbolic cognition in the agent is the root of moral evil (as in the sadism case) but, on the other, the lack of symbolic cognition in the moral agent (as in the psychopathy case) is the root of moral evil (and only when that lack occurs in human beings, not in animals). What's odd here is not the idea that there could be multiple explanations for morally objectionable behavior. Rather, the same feature (lack of symbolic cognition in some respect) is invoked to simultaneously explain why a human being's behavior is morally objectionable and an animal's is not moral at all. However, the feature can't just change its reason-giving force randomly or as it suits an ad hoc explanation.

There is another problem with the notion of symbolic cognition and modeling that I would like to note.[10] By modeling, Tse seems to mean the ability to represent another being's mind. This is allegedly symbolic because representation is assumed to involve some abstraction and not just a picture, if you will, of the other being's mental state. Even so, modeling seems to be different from the symbolic cognition that involves generating flexible symbols that can alter their meaning (and that therefore involves cross-modular binding). Even if it were true that animals lacked the latter, that wouldn't entail that they lacked some capacity for modeling another being's mind (even if it turned out to be true that their ability to do so were more limited than human beings' is). Thus, an independent argument for why animals couldn't be compassionate or sadistic appears to be needed.

I have noted some problems with Tse's association of evil with tokenization and good with kindness and compassion that is focused on the individual, in particular, (1) a capacity to attribute or model complex mental states is equally present in sadistic and nonsadistic agents; (2) even if some kinds of immoral behavior involve "tokenization," so, too, do some kinds of morally good behavior: for example, recognizing a person as a "rights holder," a member of an abstract category. Unless it can be shown that some particular neurological process always and without exception yields the judgment that it is bad or that it is good, it is just not clear how moral judgments themselves could simply be derived from neurological processes such as tokenization. And, even if such invariance obtained, the description of the neurological features themselves does not contain the additional moral predicate. The moral or evaluative predicates may coincide

with or supervene on the descriptive features, but their meaning is not captured by the neurological description.[11] If some particular kind of tokenization always resulted in a morally bad action, then that tokenization would be a contingently necessary part of the explanation for why a particular person acted in a morally condemnatory way. However, to say that a particular action is a particular species of tokenization does not seem to forestall the question, why is that particular species of tokenization bad? Is it bad because it causes pain or harm, or rather unjustified pain or harm? What makes the pain unjustified? Is some species of tokenization good because it causes pleasure or benefit, or rather pleasure or benefit that doesn't compromise other rights and duties? Or, does the badness or goodness depend on or invoke some other standard that has been developed through the exercise of substantive practical reason? That some action is or involves tokenization or a particular species of tokenization does not seem to be sufficient as an explanation or justification of its moral badness or goodness.

If Tse's claim is that a particular action can be understood simply as the expression of some symbolic processing, or of processing that has gone awry in a particular way, it also raises the question, are human beings responsible for their actions? Without any second-order reflection on and choice made with respect to possible actions contemplated or urges felt, it is not clear what responsibility would amount to. Guilt and innocence are a function of intentionality and choice, as well as of the nature of the action performed. In order to explain moral judgments of guilt and innocence, one would need to be able to explain the difference between the person who has sadistic urges but chooses to refrain from acting on them and the person who has such urges and chooses to act on them. It won't be sufficient to say (1) that the person who acted on them just had stronger urges or (2) that the person's symbolic processing rendered the person unable to comprehend the moral reprehensibility of his or her action. In the first case, one might say that the person was guilty because she ought to have had better control over her urges, but that would be somewhat different from saying that the particular expression of symbolic processing is bad or evil (even if one made that judgment as well). Or, if the person cannot exercise control over irresistible urges, then one might say that the person isn't guilty at all. While it is true that having sadistic urges or desires is not a good thing, it's not clear how symbolic processing per se is what makes them problematic. In the second case (some symbolic processing renders someone unable to comprehend the moral wrongness of his or her action), one might say that the person is simply incapable of morality.[12]

That would be an especially odd conclusion to come to if the crux of the theory is that symbolic processing is the basis for the possibility of moral judgment.

If morality necessarily involves symbolic cognition, then Tse's approach would seem to rule out any view which regarded morality in strictly non-cognitive terms—for example, a view which regarded "good" as simply an actual desire or "pro" response to some object, act, or person. Such views are implausible on other grounds; actual responses can be morally neutral, morally reprehensible (even when a "pro" response for or to an individual),[13] and morally commendable independently of the actual response. In other words, actual responses do not necessarily track moral categories. However, Tse comes very close to advocating just such a view:

> [T]he morality that one embraces and attributes to God or the universe, or that one formalizes in an "objective" body of enduring law, is an external projection of private intuitions and desires. . . . Human moralities, in short, are rooted in irrational urges and clothe themselves in rational justifications from which they claim to derive in order to disguise this. The reason that we have conflicting moralities and conceptions of good and evil is that we have conflicting desires and feelings about the various domains of choice and activity that moralities seek to govern. What all human moralities have in common is that they are symbolic. (Tse, p. ••)

Thus, the view appears to be a dispositionalist or response-dependent view of moral distinctions: symbolic processing provides for abstract representation of moral concepts, but moral distinctions are rooted in urges, feelings, and desires which depend on individual dispositions ("taste and personality") or on those in conjunction with cultural dispositions:[14]

> A fundamentalist Christian will typically reject people who embrace the body in this way [as standing for life, the divine or love] as decadent or evil. Similarly, a person who embraces the body will view the Christian suppression of the body as wrong or evil. Who is right? The answer is that no one group has succeeded in imposing its morality upon all other people. . . . The reason that no one has succeeded is that one's morality, like one's religion, and the God that one worships, is chosen based on what feels good, comfortable and right. One's moral system largely follows one's taste and personality. (Tse, p. ••)

In this passage, Tse appears to be committed to the denial of independent formulation and justification of moral standards. If that is so, then, going back to the sadist, what Tse should say is not that the sadist's behavior *is* evil but rather that some (maybe even most) individuals or cultures regard it as evil; that even if most individuals regard the sadist's behavior

as evil, that simply reflects the urges and feelings that most people happen to have about the sadist's behavior.

Tse defines "evil" and "good" as follows:

> Here "evil" will be operationally defined relative to a subject's mind to include that which harms or could harm that mind or that which that mind cares about. "Good" will be defined in a similar way to include that which benefits or could benefit a mind or that which that mind cares about. (p. ••)

On this definition, the sadist's pleasure would be "good" relative to the sadist and "evil" relative to the victim, and presumably most others would see causing pain to another as harmful to what they cared about and thus would regard the sadist's behavior as wrong or evil. However, these are descriptive claims, not prescriptive ones about how people ought or ought not to behave. The actual dispositions and responses of individuals or even of most people could be morally arbitrary at any given time or in any given context. If a dispositionalist or response-dependent theory is meant to be a theory about ethical or moral distinctions per se, then there has to be a way of sorting out actual responses from moral responses.[15] Otherwise, we simply have the sadist who experiences the infliction of pain as something good and the nonsadist who experiences it as bad, and no basis for evaluating the legitimacy of such judgments and behaviors.

As moral psychology, the theory may have much to contribute to the understanding of moral judgment and agency. However, Tse seems to be staking out a stronger claim. When he asserts that morality just is a collection of moralities, and that there is no rational adjudication of one "morality" over another, he seems to be expressing the view that personal and cultural dispositions (made possible because of symbolic processing) produce moral distinctions and to the extent that they differ, then to that extent will moral distinctions differ. Such a view is unsatisfying to the philosopher, who is interested in justification, and, I suspect, to most ordinary people, who want to know which responses are in fact morally justified and which are not; who want to know how they ought to behave, not only how and why they and others do behave as they do.

That ethical knowledge and moral standards are guided by discovery about the nature of human capacities and dispositions seems right. However, that doesn't mean that ethical standards can be explained as simply the expression of particular psychological states and dispositions. For example, waterboarding[16] is wrong because human beings are so constituted that they experience pain, coercion, humiliation, and an undermining of the capacity to function freely and as moral agents when so

treated. A moral judgment that such treatment is wrong involves the recognition that infliction of pain, humiliation, and loss of such capacity is bad, and that without sufficient justifying reasons for doing so, it is wrong to so inflict pain. It may be true that a nonsymbolic mind would not invent waterboarding. However, the infliction of pain on a victim is wrong not because the agent has the capacity for symbolic thought but because there is not a sufficient reason to justify the deliberate infliction of pain.[17] To take a contrasting case: if a physician is justified in deliberately inflicting pain on a patient, it is because presumably the reason for doing so (e.g., the hope of cure) is sufficient to justify the behavior and also because the patient has presumably consented to and is voluntarily undergoing the treatment.

Tse seems to want to allow for difference in moral judgment, to wit, his suggestion that morality is a collection of "moralities." However, that there could be legitimate difference in moral judgment does not entail that any difference is legitimate. For example, people may disagree over whether withholding medical treatment from a terminally ill patient is justified or not because they assign different weights to the value of suffering, extension of life, and so on.[18] What makes the disagreement *legitimate* is not that people have different urges and dispositions but that there are moral standards, principles, or rules that allow us to identify what differences and choices are morally legitimate.

Tse's account is helpful with respect to understanding psychological states, capacities, and dispositions that may influence the development of moral attitudes and behaviors. This account may be helpful in developing appropriate methods of moral education and persuasion. Tse raises the issue of moral education when he suggests that racist, sexist, and other discriminatory behavior might be altered by cultivating detokenization (Tse, p. ••). I have suggested some difficulties with the account of tokenization as evil. However, if there is a kind of tokenization that does operate with racist, sexist, and discriminatory behavior, then the theory may be helpful in developing methods in moral education that aim to disrupt the processes which lead to such attitudes. However, the solution may not be only to "detokenize," that is, to recognize and love persons as individuals.[19] Rather, an appropriate moral attitude may involve the cultivation of the right kind of "tokenization," for example, the capacity to recognize the equal humanity of all persons, even ones whom one doesn't and can't know or love as individuals.

Notes

1. For a more subtle and nuanced view of moral particularism than anything I shall discuss here, see Dancy (2004). Also, see Hooker and Little (2000).

2. Tse's own comments about the plurality of "moralities" would be consistent with this possibility (see below, p. ••).

3. Tse also claims that animals are not capable of the same degree of evil (or of good), of cruelty (or of kindness), of which human beings are capable.

4. See, for example, Hearne (1986).

5. At times it seems as if Tse is also claiming that it is sufficient for morality, but, as we shall see, it's not clear that that claim is supported.

6. Peter Singer and others argue for recognizing animals as having interests that are due the same moral concern as is due human beings and their interests. Tom Regan argues that animals have inherent value equal to that of human beings and therefore have rights and may not be treated merely as a resource for human needs and interests. Neither of these arguments depends on animals' having the capacity for symbolic cognition. And, both these claims are different from the claim (Hearne's) that animals themselves make moral distinctions or function as moral agents. (See, e.g., Singer, 1977; Regan, 1985a, 1985b.)

7. This interpretation might favor the particularist at least with regard to morally good acts, since they would depend on individual feelings and understandings rather than general or generalizable moral principles.

8. This interpretation might tend to favor the generalist over the particularist.

9. Tse recognizes that tokenization is by itself morally neutral when he comments that the brain may be "hardwired" to tokenize and that it may even be functionally desirable to do so from the point of view of survival (detection of dangers) and minimization of effort (Tse, p. ••).

10. Thanks to Walter Sinnott-Armstrong for raising this question and helping me to sharpen my thinking on it.

11. For a discussion of some of the issues involved here, see, for example, Blackburn (1993).

12. This is the tack that Tse takes in discussing the psychopath—although in that case, Tse suggests that the problem is that the psychopath lacks a particular type of symbolic cognition, namely, the capacity to model another mind.

13. For example, the juror, mentioned earlier, who allowed his or her compassion for the defendant (a "pro" response to the individual) to override an honest evaluation of the evidence might be morally reprehensible, particularly if the compassion

is itself a form of bias. For example, suppose the compassion is a response to a defendant with whom the juror personally identifies on the basis of sex, or skin color, or ethnicity, to the disregard of the law and the victim of the crime.

14. For philosophical discussions of dispositionalist or response-dependent views of morality, see, for example, Brower (1993); Enoch (2005); Johnston (1989). Actually, it is not clear to me whether Tse is expressing a dispositionalist view (that moral judgments are expressions of positive or negative psychological states) or a projectivist view (that moral properties are merely projections of our sentiments or urges onto the objects of those sentiments or urges). Even though he uses the term "projection" in the preceding quotation, the substance of his position seems closer to what philosophers would characterize as a dispositionalist view.

15. This is the purpose of idealization in idealized dispositionalist theories such as Lewis's. See Lewis (1989).

16. A form of torture in which the victim is strapped to a board with the feet higher than the head and is subjected to water treatments designed to simulate drowning. On November 18, 2005, ABC News reported that former CIA agents claimed the CIA had engaged in a modern form of waterboarding, along with five other "Enhanced Interrogation Techniques," against suspected members of al Qaeda. http://abcnews.go.com/WNT/Investigation/story?id=1322866 site visited 12/17/05.

17. I don't know whether the cat that plays with a mouse before killing it lacks the ability to model the mouse's experience, but if the cat is not morally responsible it is because it lacks the ability to conceptualize (symbolize) the actions as wrong and to engage in justificatory reflection about its action.

18. Reading the work of and discussions with Bernard Gert have been helpful in formulating this point. See Gert (2005) and Gert (forthcoming).

19. Tse puts it as "the need to transform one's mind by cultivating compassion and love for individuals and by paying attention to the particularities of events" (p. ••).

5.3 Reply to Dietrich and Wallace

Peter Ulric Tse

Michael Dietrich comments only on the first part of the essay, and Kathleen Wallace focuses only on the second part of the essay concerning morality. So I will address them in sequence.

Dietrich correctly points out that I am offering a hypothesis or conjecture about the changes in cognitive architecture that may have given rise to symbolic thought. I agree that a comprehensive theory will tackle the engineering details that may have realized this change in cognitive architecture, including changes in neural interconnectivity, as well as potential genetic underpinnings. However, none of this is remotely possible at this stage of the game. What we can do now is come up with hypotheses that will help guide us in our search for better facts and deeper theories, including theories that make specific predictions about changes in neural circuitry, and the genetic changes that caused those changes in neural circuitry. I intentionally avoided any discussion at this engineering level, because I have no idea what the changes in neuronal circuitry or genetics may have been that led to the proposed changes in cognitive architecture. Neither does anyone else in my field. In neuroscience, we do not even truly understand how a minicolumn works, which may be a basic unit of information-processing machinery in the cortex. We are simply not at a stage where we can describe potential changes in neural circuitry underlying object file formation.

Hypothesizing is a part of science. I state the hypothesis that there was a change in the cognitive architecture underlying object file formation such that operators and operands become free of previously modular constraints. This makes concrete predictions, which can be tested. As of now, we have only very vague ideas in neuroscience, concerning the neural underpinnings of object files. But we will find out within a decade or two, I hope. And when we do, I predict that the neural underpinnings of object file formation will look quite different in humans and other animals, even

chimps. This might eventually require doing things that are currently impracticable, like running chimps as awake and behaving subjects in attentional tasks during functional magnetic resonance imaging experiments or running humans in single-unit work, but these things are not impossible in principle. Thus, I am stating an admittedly speculative hypothesis about the roots of human symbolic thought, but it is science, because it is at least in principle falsifiable, or will be, when we understand how object files are processed, represented, and instantiated in brain circuitry.

The term "just-so" story is really not applicable here. Rudyard Kipling came up with this term to refer to fanciful stories that explain outcomes with no reference to evolution or fact. For example, Kipling writes that a leopard got its spots because an Ethiopian left fingerprints all over its pelt. This criticism is an attempt to paint the present hypothesis as fanciful or groundless. Kipling would never have seriously claimed that his story about the leopard's spots was true. I am hypothesizing that the object file hypothesis describes something that is potentially true. It is falsifiable, so it is science. My hypothesis ties together numerous behavioral observations and observations about human cognition, and it aims to make a statement about what may really have happened in fact. This is not meant to be pure fiction like the leopard's spots. Maybe my ideas are wrong. Facts will eventually help us decide if there is anything to them. But this is not just meant to be a fanciful story. It is meant to be a hypothesis about what really happened in our ancestors' brains and minds that led their minds to become symbolic.

It seems to me that many of Kathleen Wallace's criticisms stem from a desire to find a conception of morality, good and evil, that is prescriptive. Wallace finds that my view "is unsatisfying to the philosopher, who is interested in justification, and, I suspect, to most ordinary people, who want to know which responses are in fact morally justified and which are not; who want to know how they ought to behave, not only how and why they and others do behave as they do" (p. ••). But I purposely defined good and evil in a specific way, precisely in order to avoid getting into how people ought to behave:

> Here "evil" will be operationally defined relative to a subject's mind to include that which harms or could harm that mind or that which that mind cares about. "Good" will be defined in a similar way to include that which benefits or could benefit a mind or that which that mind cares about. (p. ••)

My claims about morality are unabashedly only about how and why people behave and think as they do, not about how people should act. I

am being descriptive, not prescriptive or normative. I am saying how normative ideas such as "ought," "appropriate," or "evil" came into existence, not what people should do or not do. I defined good and evil in this way because I did not want to get into a discussion of what is universally good or evil intention or action for all minds, or universally good or evil, right or wrong, for any particular circumstance. I do not deny that such universalist prescriptive arguments can be made and that they can be useful and instructive. History is full of such universalist prescriptions, often conflicting with one another. I just do not want to address these questions here, since they will lead me too far from my central point, which is that the possibility of morality, good, and evil results from symbolic cognition. Instead, I focus on what is good or evil from the point of view of a single person, regardless of universalist prescriptions, while realizing the limitations of such a narrow definition. The obvious danger here is moral relativism, namely, that what seems good to a sadist, say torture, will comprise evil for her victim. In order to counter this danger, it is necessary to engage in universalist arguments that will prescribe what is right and wrong in general. I have nothing against such reasoning. It is just too far afield from the point of this essay.

I agree that tokenization, compassion, or focusing on the individual do not necessarily determine the goodness or badness of an act or intention. I do not mean to say that tokenization is necessarily a root of evil. I agree that it is in itself neutral and can lead to good or evil acts or intentions. My point is simply that tokenization is only possible for a symbolic mind. It can be one root of evil or good, and it only exists in symbolic minds where a token can be taken to stand for something more general than itself.