CANADIAN UNDERGRADUATE JOURNAL OF
COGNITIVE SCIENCE

The Canadian Undergraduate Journal of Cognitive Science is an electronic and print journal published by the Cognitive Science Student Association at Simon Fraser University. Our aim is to provide a forum for students to share work among peers and gain valuable experience in the process of getting an academic paper published. As a publication, CUJCS provides a unique reference for students, showcases research by other undergraduate students, improves the contact and exchange of ideas between Canadian students and cognitive scientists alike, and illustrates the interdisciplinary work that is the hallmark of cognitive science. Although preference is given to Canadian students, contributions from students elsewhere are strongly encouraged.

Copyright Information

Submissions

Undergraduate students interested in contributing a paper to CUJCS should contact the editors for up-to-date submission guidelines:
cujcs-inquiries@sfu.ca

ISSN 1913-0651

Simon Fraser University
Cognitive Science Student Association
8888 University Drive
Burnaby, BC
V5A 1S6

Table of Contents

Why Animal Consciousness Falls Short: Language as a requirement for the highest level of conscious experience

Ben D. Amsel University of Calgary

It is argued that language is thus far a solely human endeavour, and that animal (non-human) forms of communication should not be classified as language. However, whether certain animals may possess the necessary neuroanatomical capabilities for language acquisition remains unresolved. It is also argued that certain intentional mental states are dependent on language, and have thus far been realized only in language-endowed humans. These mental states are implicated as constituents of the highest form of conscious experience on a proposed consciousness continuum, which includes humans and non-human animals alike. Language is seen as a product of cultural evolution, which differs in kind from Darwinian evolution, and can account for the substantial disparity between human and animal minds.

Imagine that tomorrow morning an all-encompassing airborne virus sweeps the earth, leaving no human unaffected. This virus alters the brain of every human in such a way that we can no longer produce, comprehend, or even imagine language. Would we then cease to have the full range of conscious experience? Assuming we continued to reproduce, would the subsequent generation of humans fail to develop consciousness? If so, it would seem that language is the insuperable barrier between human and animal minds (from herein, the human-animal distinction will be maintained for clarity, though it is acknowledged that humans are animals). The topic of animal consciousness has provoked considerable controversy, and continues to do so. Opinions have ranged from the outright relegation of animals to "unconscious automata" (most famously espoused by Descartes and T. H. Huxley), to the assertion that chimpanzees are much too close to humans—both in terms of phylogeny (genetic evolutionary history) and neuroanatomy—to deny them consciousness (Searle, 2002; Fouts & Jensvold, 2002).

In this paper I will argue that while certain thoughts are not reliant on language (and thus attainable by many animals), there do exist certain higher-order mental states that depend on language, and therefore on human minds. Furthermore, I will implicate these mental states as necessary components of consciousness, as we experience it in full. These suppositions assume that consciousness is not a binary on-or-off phenomenon, but is rather a towering vertical "consciousness continuum" of increasing awareness and mental complexity. According to my argument, language-proficient humans are the sole inhabitants of the highest level of this continuum. Importantly, this argument will not consign animals to the "unconscious automata" mentioned earlier, or even call into question our ethical responsibilities towards them. Rather than denying animal consciousness outright, I will argue that it simply holds a lesser place on this continuum.

Animal Communication and Human Language: Can We Bridge the Gap?

In this section, I will show that human-trained animals have not been successful in acquiring genuine language thus far. I will also attempt to show that the naturally occurring communication of animals cannot be classified as language. Finally, I will consider the possibility that certain animals may possess sufficient brains for language acquisition.

The Attempts and Failures of Animal Language Acquisition

The issue at hand is simply the extent to which animals have acquired language thus far. Rumbaugh and Savage-Rumbaugh (1998) headed one of the most successful animal language projects to date, and claimed that their two naturally trained chimps (i.e., like human infants they were constantly spoken to from birth rather than trained with conditioning techniques, as earlier projects had done) could understand limited past and future verb tenses as well as impute basic states of mind to other chimps—abilities found in the typical two- or three-year-old human. However, while these chimps may have learned some basic forms of language, they were vastly inferior to humans (even at age four) in understanding and producing grammatically sound language statements. Perhaps the best known chimpanzee language students (in this case, American sign-language or ASL) are Washoe and Nim Chimpsky, who were trained by the Gardners and Herbert Terrace respectively. After several years of intense language training, Terrace and colleagues blew the whistle (1979), as they discovered by watching videos of both chimps' signings that they could not be definitively extracted from mimetic gestures. That is, the chimps may have been copying the experimenters' hand signings without actually understanding them. Also, when Terrace brought in a sign language expert to observe, the expert reported that the chimps were essentially ignoring the structured grammatical code of ASL, and were signing almost at random. Furthermore, when Jane Goodall was asked to observe the chimps, she reported that she recognized almost all of the "learned" gestures from chimps she had observed *in the wild*.

Due to the length constraints of this paper, I will not address this issue further. At this point in time it is almost universally acknowledged—by would-be believers and sceptics alike—that animals have failed thus far to acquire language (Stamp-Dawkins, 1998; Macphail, 1998; Pinker, 1994; Dennett, 1991; Bloom, 1998; Searle, 2002).

Natural Animal Communication is Not Language

Following in Noam Chomsky's footsteps, many scientists have argued for a uniquely human language instinct that elevates language acquisition beyond a mere conditioned learning response, as B.F. Skinner so infamously posited. Backing this notion, linguists have isolated several unique features of language that are absent in natural animal communications such as calls, grunts, and echolocation (Lyons, 1981; Pinker, 1994). I will draw attention to three of these unique features to make my point. The first is the infinitive or combinatorial nature of language, which means that with a little creativity, humans can utter a completely original "first time in the Universe" statement. For example, I would bet against the existence—at any point in history—of the following sentence prior to this moment: "In a deliciously remote sub-arctic atoll, there lives a melancholy faction of slightly bulbous elephant seals who dejectedly sustain the two-hundred and thirty-seventh—no wait—two-hundred and thirty-eighth rainfall of the year." The second unique feature of language is its hierarchical nature. For example, essays are made of clauses, which are made of phrases, which are made of words, which are made of phonemes. That is, language is a rigorously structured system of *conferred* meaning, as opposed to a collection of *inherently* meaningful signals. This point becomes clearer in relation to the third unique feature of language: arbitrariness. Words are a combination of sounds used as arbitrary symbols to represent specified meanings. Different cultures ascribe different meanings to the same sounds, and thus, unlike animal communication, there is no one-to-one mapping of sound onto meaning (with the occasional exception of onomatopoeia). If we were forced to interpret meaning directly from sound, as animals surely do, we would run out of things to say rather quickly—there are only so many distinguishable sounds.

With these points in mind, it should be reasonably clear that natural animal communication is not language—in the true linguistic sense. But what about the possibility that a particularly bright chimpanzee, under the finest tutelage, may one day acquire language?

Do Chimpanzees Have What it Takes to Acquire Language and Does it Matter?

A logical starting point for this question is the evolutionary perspective. One must ask whether human and chimp brains have diverged *sufficiently* in the last five to eight million years since our common ancestor as to render chimp brains inadequate language processors. Not surprisingly, this question is far from resolved. Among others, Pinker (1994) and Lieberman (1993) have evidenced certain neuroanatomical disparities between the two species, such as the human possession of highly developed speech modules—Wernicke and Broca's areas being the most important. The usual Darwinian explanation for the evolution of these unique brain structures entails a process called "preadaptation", whereby "an organ originally constructed for one purpose . . . might be modified for some other and quite unique purpose" (Darwin, 1964 [1859], p. 190). In this case, the brain mechanisms that control syntax are thought to be derived from neural circuits originally involved in the *unrelated* automatic motor control system underlying vocalization (Lieberman, 1991). Furthermore, Pinker (1994) draws attention to brain imaging studies that have placed the locations of human language and chimp communication in different brain regions. When engaging in language, human brains are primarily engaged in the cerebral cortex, whereas chimpanzee calls, grunts, and shrieks are placed in phylogenetically older brain structures such as the brain stem and limbic system. Interestingly, it has also been established that when a human inadvertently calls out in anger or pain, it is these older structures rather than the more advanced cortical structures that are involved (Jurgen, 1992). These data certainly point to a unique human language instinct, but others have revealed contrasting neuroanatomical evidence.

Fouts and Jensvold (2002) report that recent neurological research has uncovered structures in chimp brains that are very similar to Wernicke's and Broca's areas, and furthermore, that these areas are not solely "language organs," but serve also to facilitate pattern recognition and manual movement in both humans and chimps. On a more general note, some have pointed out that the recent but feverish search for the neural correlate of consciousness (i.e., a certain biochemical state, arrangement of neuronal connections, combination of synaptic events, etc., that would exclusively produce consciousness; see: Chalmers, 2000) has not yet revealed any structures or processes necessary for consciousness that are present only in the human brain (Griffin & Speck, 2004; Searle, 2002).

After reviewing some of the evidence, we must indecisively conclude that some animals may possess the necessary neuroanatomical equipment for language acquisition. I have included this ultimately indefinite sub-section not because it is central to my argument, but because it will be central to my argument when it is eventually resolved (I agree with Dennett and Searle here in that there is no reason why science will not one day conquer these types of problems). For the time being however, all that matters is that animals have not thus far achieved any advanced sort of language capabilities. What also follows indirectly from the evidence discussed here is that while animals like chimpanzees may possess sufficient brains for eventual language proficiency, small-brained animals such as insects, bats, mice, etcetera, almost certainly do not, and almost certainly (this is now my supposition) will never attain consciousness as humans know it.

Knowing Versus *Really* Knowing

Donald Griffin, the father of cognitive ethology, takes a particularly liberal view on animal consciousness: "The goals of insects may be simple compared to the theories of philosophers. But this provides no firm basis for dogmatically ruling out conscious perception of simple meanings and desired goals" (2001, p. 259). What exactly does he mean by the *conscious* perception of a desired goal? Does he believe that an insect could possess a conscious intentional mental state (intentionality is the notion of the "aboutness" of thought, or the object that a thought refers to), wherein it thought *about* a nearby food source? I believe Griffin has confused the basic state of having and using a sensory system (which does not require awareness), with the more complicated state of having knowledge *about* a particular mental state, whether that mental state is intentional itself or sensory in nature (cf. Rosenthal's "creature consciousness" vs. "state consciousness," 2002). For example, an insect will zero in on a food source for the simple reason that its sensory system has detected it; that insect need not have an intentional state such that a) it recognizes the need to sustain itself with nourishment, and therefore must seek a food source (read: "I'm hungry"), or b) it recognizes that the food source detected by its sensory system is in fact the very place to provide nourishment. Of course, these are no more than speculations, and I cannot be sure that this insect lacks "conscious perception of simple meanings and desired goals," but I am merely paying heed to Occam's razor (a key principle of the scientific method: one should not increase, beyond what is necessary, the number of entities required to explain anything). This insect's food-seeking activities can be wholly explained by the Darwinian premise of differential fitness. Since the insect's sensory system regularly succeeds in locating a food source (and hence allows that insect to fully develop and reproduce, thus passing on its genes), it is due either to the natural selection of similarly efficient sensory systems in its predecessors, or a random genetic mutation that has suddenly conferred such an ability upon the insect. These basic tenets of evolution will be entirely sufficient in explaining this insect's behaviour. There is no need for the insertion of intentional mental states here, and the same can even be said of humans. When my body needs refuelling, my sensory system will eventually lead me to a food source, regardless of whether I have thoughts like "I am hungry," or "The fridge will do just fine for this task." As we shall see, it is only when humans engage in language-dependent higher-order thought that intentionality becomes necessary.

Of considerable import here is the controversial relationship between language and thought. Perhaps the most popular contemporary view is the development of human language as a communicative system based upon a "universal grammar"—an innate human module of the mind that houses purely mental (rather than linguistic) thought, often termed "mentalese" (Pinker, 1994; Chomsky, 1975). According to this theory, language serves as the communicatory vehicle of thought, which pre-exists language. That is, the framework for thinking is already present in our brains at birth—it is innate. While the directly opposing theory of radical "Whorfian" linguistic relativism (e.g., French-speaking people and English speaking people's thought is fundamentally different because their languages are different) has all but disappeared as a viable hypothesis, many thinkers continue to support the cognitive or constitutive role of language in some forms of thought (Dennett, 1991, 1995; Carruthers, 1996). According to this theory, some thoughts are not only expressed in natural language, but are also composed of natural language—these thoughts are natural language sentences. If this is true, there must be certain complex thoughts that are unrealizable (they would not exist) unless they are arrived at with language. This latter position will be backed herein, as I will attempt to provide evidence for certain language-dependent thoughts, which by my definition are only available to humans.

An Extension of Searle's Language - Dependent Intentional States

John Searle recognizes the existence of certain types of intentional thought that require language (he suggests five), and argues that they fall into two classes: "Either the state has conditions of satisfaction that are essentially linguistic or the mode of representing the conditions of satisfaction is essentially linguistic. Or, quite commonly, both" (Searle, 2002, p. 70). Examples of states that fall into both classes are the belief that "throw" is a transitive verb, or that "Saturday" and "Sunday" fall on the "weekend"—these beliefs are dependent on our arbitrary linguistic code, and employ purely normative (human-invented) concepts. The question is not whether these particular intentional states are attainable in animals (they are certainly not), but what are the consequences for consciousness? I do not want to (although one probably could) deny animals the full range of human consciousness based on the inability to have these mundane intentional thoughts. However, there are certain language-dependent states, thoughts, or beliefs that must stand as sentinels to the highest level of the consciousness continuum.

Of the five types of Searle's language-dependent thought, two types are of particular interest here:

1. *Intentional states that represent complex facts, where the complexity cannot be represented without language.*
2. *Intentional states that represent facts that are so remote in space and time from the animal's experience as to be unrepresentable without language (Searle, 2002, p. 69).*

Searle was good enough to think up these special cases, but he does not provide more than one brief example of each, and does not discuss the implications for animal or human consciousness. I now propose to suggest a few instances of these two types of intentional language-dependent states that will provide at least a start for the eventual construction of the highest level on the consciousness continuum.

1. *Upon waking from a dream-like state, experiencing the sudden realization that a particular belief or intention originating in that state is false.* For instance, if one has a dream in which one has gone bankrupt or won the lottery, this intentional belief will be rejected at some point during the transition from unconsciousness to consciousness. In the absence of physical stimuli, it is only the addition of language that can make a belief true or false. That is, a person does not have to get out of bed and perceive that nothing in

their home and garage has improved, and thus infer that they have not in fact won the lottery. Instead, they merely have an intentional thought like: "Now that I am conscious, I may disregard that particular belief, as I know that events originating in dreams have no intrinsic basis for truth." Of course by adulthood, such excessive rationalization is unnecessary, but at some point in one's development this sort of reasoning had to have occurred. Interestingly, researchers have been successful in teaching people to slide this point of rational reckoning further back in the direction of unconsciousness, with the culmination being lucid dreaming—where one can differentiate fact from fiction while dreaming (Purcell, Mullington, Moffitt, Hoffmann, & Pigeau, 1986).

2. *Having the belief that one will inevitably die in the future.* Are animals cognizant of their unavoidable end? Likely not. It has been shown in several studies (e.g., Betz & Poster, 1984; Mahon, Goldberg, & Washington, 1999) that humans do not come to fully understand the concept of death until age six at the very earliest—at which time human language capabilities (as well as all other cognitive capacities) reach far beyond those of the most adept language-trained chimps. Also, in a sample of mentally challenged adults, lack of concept of death and lack of language skills were positively correlated (McEvoy, Reid, & Geurin, 2002). Lastly, well-respected psychiatrist Robert Langs has studied the concept of death for several decades and has found that language acquisition is critical to the development of death anxiety—a consequence of the unique human ability to perceive death as an abstract concept (1997). As far as I know, there has been no evidence for the presence of an abstract death concept in animals. It seems highly unlikely that an animal could possess a mental state in relation to mortality or the transience of life without ever having learned about it, as humans do.

3. *Dissonance in consciousness as revealed or created by language.* Turning to what we can imagine as the summit of the consciousness continuum, many great writers have recognized (and certainly suffered from) the potentially pernicious effects of language on consciousness, wherein language unearths despairing or unnerving thoughts and forces one to confront them. To differing degrees, every individual must confront the hopes and fears that language creates in relation to nature, culture, and one's self. It should not be surprising that gifted writers often struggle with their extraordinary ability to manipulate language and the consequences for their own consciousness. We do not hesitate to grant the human mind (especially the artistic mind) the unique ability to fathom the wonderful, magical, or sublime, and therefore we must be prepared to accept its capacity for darker undertakings. In F. Scott Fitzgerald's semi-autobiographical first

novel *This Side of Paradise*, the protagonist continuously struggles with his reeling literary mind:

> *He had a sense of reality such as material things could never give him. His intellectual content seemed to submit passively to it, and it fitted like a glove everything that had ever preceded it in his life . . . He pitched onto the bed and rolled over on his face with a deadly fear that he was going mad. He wanted people, people, some one sane and stupid and good (2000, pp. 107-108).*

Dostoevsky expresses similar sentiments in *Notes From the Underground*, but ultimately sees language itself as the only way to stabilize the maelstrom of *language-induced disorder* that invades consciousness.

> *I swear, gentlemen, that to be too conscious is an illness—a real thoroughgoing illness. For man's everyday needs, it would have been quite enough to have the ordinary human consciousness, that is, half or a quarter of the amount which falls to the lot of a cultivated man . . . The more conscious I was of goodness and of all that was "sublime and beautiful," the more deeply I sank into my mire and the more ready I was to sink in it altogether . . . But I will explain it. I will get to the bottom of it! That is why I have taken up my pen (1945, p. 132).*

American literary scholar Paul Johnston also views language as both the cause and the saviour of the unpredictable and expanding nature of consciousness:

> *It is poetry's task not to lie about the world and make it more ordered than it is, nor to give in and simply repeat its disorder, but to fully contain the world's disorder within the order imposed by art, to transform chaos into beauty. In doing so our minds are strengthened. Our language capacity is strengthened. Our consciousness is strengthened, better able to contain the contradictions our consciousness inevitably creates (1998, p. 735).*

On a final literary note, it is an unfortunate but well-documented fact that the suicide rate among poets is much higher than among the normal population—a reality that many researchers have attributed not only to the type of people drawn naturally to poetry, but to the nature of composing poetry itself (Wiltsey Sterman & Pennebaker, 2001; Jamison, 1993; Silverman & Will, 1986). I am not denying that animals experience uncomfortable or distressing mental states, but the kind of mental discomfort discussed above falls into a class of language-dependent dysphoria. It is entirely reasonable to suppose that these mental states feel different—are of a different experiential kind than the dysphoric mental states of animals.

4. *The lack of "cognitive closure" in human puzzlement.* According to philosophers Jerry Fodor and Colin McGinn, cognitive closure exists in differing degrees in all living creatures, including humans. Basically, what is beyond the mental capacity of a mouse is attainable by a monkey, but what is attainable by a human is beyond a monkey. McGinn (1991) makes the accurate claim that a monkey could not understand the concept of an electron. However, this example does not translate into human comprehension. It is certainly true that there is much in the Universe currently beyond our comprehension, but the difference is this: we are aware of it. Not only do monkeys fail in fathoming an electron, they almost surely fail in fathoming period. The profound sense of awe and bafflement that is created by thinking about the size of the Universe, the age of the Universe, free will, and yes, even consciousness, must be absent in animals—though of course I cannot prove this. Put simply, while humans may not be able to comprehend these answers, we are certainly aware of the questions. On the possibility that humans will ever be able to answer these questions with the explanatory power of language, disagreement abounds. Fodor and McGinn think not, as they believe that dismissing cognitive closure in humans is simply the mark of humanistic arrogance. While this may have some truth, there is no reason to think that language itself is not up to these tasks. Dennett (1995) correctly recalls the uniquely human ability to "parse" or make sense of an infinite number of combinations in the realm of language—a skill discussed earlier in this paper. He then asks: "If we can understand all the sentences (in principle), couldn't we understand the ordered sets of sentences that best express the solutions to [these] problems?" (1995, p. 382). Precisely. The burden of explanation and understanding falls on us, not on language. There is a particular combination of words that best explains the nature of consciousness or the size of the Universe . . . we just haven't found it.

While these four instances of language-dependent mental states are only a start, I believe they are a representative sample of what kinds of mental states can gain access to the highest level on the consciousness continuum. In the next section, I will attempt to explain why language is so much more powerful than purely biological tools of the mind, and why it even dwarfs Darwinian evolution.

Beyond Darwin: Language and Cultural Evolution

One of the key principles of Darwin's theory of natural selection is the concept of incremental change. For the most part, major adaptations and improvements stemming from random genetic mutation that provide the impetus for sustained life on earth are visible only on a macroscopic time scale (i.e., several hundred thousand years). This includes the anatomy of the brain, and with only five to seven million years since the chimpanzee-human common ancestor, evolution could not have fashioned a brain so vastly improved as to explain the current difference between chimp and human knowledge. Our respective mental apparatuses are not so very different, but when ours are conferred with language, the products certainly are.

The Almighty Word

Think for a moment about all the knowledge you possess—everything that you have ever learned. Now hazard a guess as to what proportion of that knowledge came from purely first-hand experience or introspection, and what proportion came from hearing about it, reading about it, and subsequently thinking and writing about what you have learned. The very essence of science itself is based on public access to acquired knowledge. Consider the following analogy: The difference between computers in 1981 and 1983 was minimal; they relied upon the same types of processors, microchips, and hardware in general. Each computer could store a moderate amount of data, but this information was permanently imbedded in each single host. However, 1982 saw the birth of the Internet. Within a year, millions of individual computers were connected by a communication system and all that previously enslaved information was liberated. Suddenly every single computer had access to an almost infinite amount of information, and could acquire new information at a furious rate. Substitute brains for computers, language for the Internet, and a few hundred thousand years for 1982, and you have it.

The amount of knowledge available to the human race not only expanded exponentially with the birth of language, it does so every generation. As Dennett puts it: "When comparing the time scales of genetic and cultural evolution, it is useful to bear in mind that we today—every one of us—can easily understand many ideas that were simply unthinkable by the geniuses in our grandparents' generation" (1995, p. 377). Unthinkable in this case does not mean beyond the capacity of comprehension. The smartest physicists and astronomers of the 19th century had the same kind of brains as those in the 21st century. Yet, if told about an extremely matter-condensed region in space that produced a sufficiently powerful gravitational field as to draw in all matter, even light, they likely would not have believed or understood such a statement—dismissing it as some sort of imaginary black hole.

Getting back to animal consciousness, it does not matter how similar human and chimp brains are, because for now, language is the sole domain of humans. If language had not made its appearance in recent evolutionary history, this paper as well as every other attempt to shed light on consciousness would not exist. In fact, the entire problem and concept of consciousness would cease to exist. Surely the very acknowledgement of consciousness, which is entirely unique to humans, is a necessary permit for admittance to the highest level on the consciousness continuum.

Concluding Remarks

As one may have noticed, the argument put forth in this paper is not particularly counter-intuitive or radical. That is, I have not strictly argued for or against animal consciousness, but placed it lower on the continuum as a reflection of human language capabilities. I also avoided the declaration that animals do not possess at least the potential for language acquisition—this remains to be seen. In circumventing a binary approach to consciousness, we can do away with the misguided question: "Are animals conscious or not?" When one considers the full gamut, from insects to rodents, from birds to apes, and from infants to adults, it seems overly simplistic to impose a single divider of consciousness upon such a vast and diverse group of organisms. Though not directly discussed due to length constraints, it is acknowledged that recent empirical research into animal consciousness has revealed a number of previously unknown and duly impressive mental abilities and states in a variety of animals (e.g., metacognition and basic theories of mind) (e.g., Hampton, 2001; Smith, Shields, & Washburn, 2003; Riess, 1998). However, based on the nature of my argument, none of these abilities are sufficient to gain access to the *language-dependent* mental states that I have proposed in this paper. Like consciousness heavyweights Dennett, Pinker, and Searle, I believe consciousness is originally a product of biology that can and should be studied with the steady hand of science. However, as I have attempted to exhibit in this paper, consciousness in humans can reach beyond its biological roots. As we have seen, language has added a limitless dimension to human consciousness in the sense that it can represent or constitute an infinite number of thoughts—the majority of which have probably not occurred yet. The consciousness continuum I have spoken of is no doubt expanding, and we do not know its limits. As Albert Einstein said: "The knowledge about man is still in its infancy."

references

Betz, C. L., & Poster, E. C. (1984). Children's concepts of death: Implications for pediatric practice. *Nu rsing Clinics of North America*, 19, 341-349.

Bloom, P. (1998). Language and mental life. In S. Hameroff, A. Kaszniak, & A. Scott (Eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* (pp. 561-571). Cambridge, MA: MIT Press.

Carruthers, P. (1996). *Language, thought, and consciousness: An essay in philosophical psychology.* Cambridge, MA: Cambridge University Press.

Chalmers, D. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (pp. 17-40). Cambridge, MA: MIT Press.

Chomsky, N. (1972). *Language and the mind.* New York: Harcourt, Brace, & Jovanovich.

Chomsky, N. (1975). *Reflections on language.* New York: Pantheon.

Darwin, C. (1964). [1859]. Facsimile Edition. *On the origin of species.* Cambridge, MA: Harvard University Press.

Dennett, D. C. (1991). *Consciousness explained.* New York: Little Brown and Company.

Dennett, D. C. (1995). *Darwin's dangerous idea.* New York: Touchstone.

Dostoevsky, F. (1945). Notes from the underground. In *The Short Novels of Dostoevsky.* (pp. 129-222). New York: Dial Press.

Fitzgerald, F. S. (2000). *This side of paradise.* New York: Penguin Classics.

Fouts, R.S. & Jensvold, M.L. (2002). Armchair delusions versus empirical realities: a neurological model for the continuity of ape and human languaging. In M. Goodman, M. Moffat (Eds.), *Probing Human Origins* (pp. 87-101). Cambridge, MA: American Academy of Art and Sciences.

Griffin, D. R. (1991). *Animal minds: Beyond cognition to consciousness.* Chicago: University of Chicago Press.

Griffin, D. R. & Speck, G. B. (2004). New evidence of animal consciousness. *Animal Cognition*, 7, 5-18.

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences and of the United States of America*, 98, 5359-5362.

Jamison, K. (1993). *Touched with fire: Manic-depressive illness and the artistic temperament.* New York: Free Press.

Johnston, P. (1998). Poetry, mental models, and discordant consciousness. In S. Hameroff, A. Kaszniak, & A. Scott (Eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* (pp. 733-738). Cambridge, MA: MIT Press.

Jurgen, U. (1992). On the neurobiology of vocal communication. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (pp. 31-42). Cambridge, MA: Cambridge University Press.

Langs, R. (1997). *Death anxiety and clinical practice.* London: Karnac Books.

Lieberman, P. (1991). *Uniquely human: The evolution of speech, thought, and selfless behaviour.* Cambridge, MA: Harvard University Press.

Lieberman, P. (1993). On the evolution of language. *Current Anthropology*, 34, 165-170.

Lyons, J. (1981). *Language and linguistics: An introduction.* Cambridge, MA: Cambridge University Press.

Macphail, E.M. (1998). *The evolution of consciousness.* New York: Oxford University Press

Mahon, M. M., Goldberg, E.Z. & Washington, S.K. (1999). Concept of death in a sample of Israeli kibbutz children. *Death Studies*, 23, 43-59.

McGinn, C. (1991). *The problem of consciousness.* Oxford: Blackwell.

McEvoy, J., Reid, Y. & Guerin, S. (2002). Emotion recognition and concept of death in people with learning disabilities. *British Journal of developmental disabilities*, 48, 83-89.

Pinker, S. (1994). *The language instinct.* New

York: Morrow.
Purcell, S., Mullington, J., Moffitt, A., Hoffmann, R. & Pigeau, R. (1986). Dream self-reflectiveness as a learned cognitive skill. *Sleep*, 9, 423-437.
Reiss, D. (1998). Cognition and communication in dolphins: A question of consciousness. In S. Hameroff, A. Kaszniak, & A. Scott (Eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* (pp. 551-560). Cambridge, MA: MIT Press.
Rosenthal, D. (2002). Explaining consciousness. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 406-421). New York: Oxford University Press.
Rumbaugh, D. & Savage-Rumbaugh, S. (1998). Perspectives on consciousness, language, and other emergent processes in apes and humans. In S. Hameroff, A. Kaszniak, & A. Scott (Eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* (pp. 533-549). Cambridge, MA: MIT Press.
Searle, J. R. (2002). Consciousness and language. Cambridge, UK: Cambridge University Press.
Silverman, M. A. & Will, N. P. (1986). Sylvia Plath and the failure of emotional self-repair through poetry. *Psychoanalytic Quarterly*, 55, 99–129.
Stamp-Dawkins, M. (1998). *Through our eyes only? The search for animal consciousness.* New York: Oxford University Press.
Terrace, H. S., Petitto, L. A., Sanders, R. J. & Bever, T. J. (1979). Can an ape create a sentence? *Science*, 206, 891-902.
Wiltsey Sterman, S. & Pennebaker, J.W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63, 517-22.

Categorization in Problem Solving: Underpinning of cross-domain, expert level performance

David Baldwin Northeastern University

Problem-solving is a ubiquitous activity not limited to well-defined, formal tasks such as mathematics. Research, however, has yet to provide a set of examples beyond these formal domains indicating how domain affects problem-solving processes. This paper provides evidence that categorization in problem-solving, regardless of domain, operates in a relatively uniform fashion. By framing problem-solving and analogical transfer as instances of categorization and vice versa, it is possible to apply categorization research and its study of ill-defined domains to this task. The paper attempts to reconcile the differential characterization of knowledge acquisition in both areas of research by providing a new, complementary theory. Possible experimental tasks and directions for future research are also included.

1 INTRODUCTION

Problem-solving is a fundamental function of human cognition that spans a range of highly differing tasks. Little effort is required to observe simple instances of such behavior in normal, everyday settings. Some are banal and routine, yet share similar structure to less trivial tasks. Getting dressed in the morning, for example, requires a series of interchangeable steps subject to certain constraints—donning a belt before a pair of pants may lead to an awkward wardrobe. Each of these steps requires the recognition and classification of the relevant article, knowledge how they are manipulated and a diverse range of other high level cognitive functions.

Problem-solving, though, also encompasses far more interesting phenomena. More complex tasks include a notion of a goal and the means for achieving it—board games or literature analysis, for example, are instances of problem-solving. The board game uses a set of xed rules upon which more intricate strategies might be built. The analysis of a complex poem requires the realization of connections between the reader's unique set of literary and cultural knowledge to frame an interpretation of a particular piece of work. Beyond the obvious similarities, such as low level perception, it would appear as though problems might use a differing arrangement of primitive reasoning processes.

One commonality between complex domains is the differing levels of skill and expertise exhibited by problem solvers. As a novice, only the most simple of 'problems' can be performed. Before more complex issues can be attempted a certain level of proficiency is required. Experience and its development rely not only on the accumulation of previous information, but its subsequent integration and mental representation. Sophistication of thought in these activities is necessarily grounded on what is gleaned from previous experience and its abstraction.

The act of retrieving mental representation and putting it to use when faced with a task is largely a matter of categorization—identifying some item as an instance of one or more previously learned concepts. Ability to categorize in increasingly intricate ways has a direct correlation to skill. The analysis of a board game position requires determining, at a novice level, what moves are available before any evaluation may occur. As knowledge increases, a player may be able to characterize a position as one where a certain tactic is often found. A poem may be classified as containing some basic aspect, such as metaphor, or something far more complex, such as the amalgamation of two different classical styles. In both instances, the problem requires a novel instance of familiar domain to be categorized before further processing occurs. Simply put, it provides the critical first step in knowledge retrieval.

This paper discusses the use of categorization within problem-solving through the research of expert/novice differences. The remainder of this section presents definitional and theoretic foundations.

Section 2 examines traditional categorization and establishes it as a type of simplified problem-solving. It is then argued in section 3 that a key aspect of the problem-solving process—analogical transfer—connects simple categorization and complex problem-solving. Finally, theoretical issues are considered and accompanied by suggestions for further research in section 4.

1.1 What Constitutes Problem Solving? Two Central Classes

The typical interpretation of problem-solving is somewhat restrictive. It generally includes only those tasks that have concrete goals, subgoals and rules to achieve them. These types of tasks constitute the vast majority of those utilized in problem-solving research. However, they are but a small subset of human reasoning that ignores more common, everyday cognitive function. An overly-generous definition might include the most trivial of tasks (e.g. those in which the vast majority of the population is extremely proficient), such as carrying objects or observing color.

Such an interpretation, however, includes examples that may tell little about the effects of high level cognitive function. In order to obtain a general, yet focused, view of problem-solving, it will be examined both in and outside of strictly structured domains, but only in tasks that are particularly revealing of problem-solving behavior.

Problem-solving will be primarily observed in terms of ill- or well-defined tasks. These two classes provide a focused subset reliant on domain knowledge and non-trivial levels of expertise. Ill- and well-defined domains form the endpoints of a spectrum of rule formality. Well-defined domains usually encompass analytical tasks, such as chess or mathematics, which contain a generative set of in flexible rules upon which more complex properties are built (Chi, Feltovich & Glaser, 1981). In ill-defined areas, these principles are more ad-hoc and may be learned from experience without being explicitly demonstrated. Conflict resolution, for example, may have a flexible set of guidelines depending on various situational factors. Instead of atomic rules, experts in ill-defined domains must rely on increasingly skillful inductive inference to decompose and examine a particular task (Ormerod, 2003).

Despite the sizable body of problem-solving literature, previous research has yet to conclusively illuminate patterns in behavior as a function of domain class. Shin et al. has suggested that a variety of problems are dependent on knowledge structure and ability to justify reasoning (Shin, Jonassen & McGee, 2003). Other studies have shown evidence that ability across certain domain classes is independent (Brabeck & Wood, 1990). However, little work has been performed to explore what processes affect such commonalities. By showing the categorization functions similarly across a variety of tasks, strong evidence may be established for its generality in the problem-solving process.

1.2 Categorization's Theoretical Role in The Problem Solving Processes

Problem space theory provides a formal mechanism for analyzing problem-solving. Beginning at a starting state (the problem itself), the problem solver is faced with using a set of transition rules to reach new states. One or more goal states define an acceptable solution and the conclusion of the problem. Newell decomposed the problem-solving processes into a series of steps—an exploration of these states through planning (identifying salient subgoals and their properties) to find an optimal path to a goal (Newell & Simon, 1972). Expertise is shown with more efficient and successful foresight and execution, which is reliant on a number of factors such as abstraction of prior examples and chunking through memory (Richman, Gobet, Staszewski & Simon, 1994).

Further research by Newell and others observes that the difference between ill and well-defined problems lies in the definition of transition and goal states. These are often explicit in formal domains—mathematical formulae and chess moves both being transitions—but within ill-defined areas, even the problem itself may not be completely explicit. Problem-solving here consists not only of exploring the space, but identifying the salient transitions. Additionally, there may be several, or even an infinite number, of feasible goals (Ormerod, 2003).

Categorization plays a significant role in both domain areas. In ill-defined domains, categorization of the unknown states is crucial in monitoring progress and determining when the problem is complete. Formal domains provide transitions explicitly but, in complex tasks, planning cannot take place without abstraction—in a sufficiently large problem space the problem may be intractable without means of reduction. A myriad of chess studies have shown experts' ability to recall chess positions perfectly after only brief exposure. The accepted explanation for this phenomena lies in the chunking of common chess structures, rather than the individual pieces, into short term memory (de Groot & Gobet, 2003). These structures, however, are not static, and chunking is not merely a matter of perceptual recognition—the groupings are members of categories salient to chess players (Richman et. al, 1997).

Newell's model of problem space exploration is reliant on factors such as memory and general ability, but it is also highly dependent on contextual information. This context is obtained through categorization of the current state. It follows that categorization is one of the initial bootstrapping processes used when a problem is assessed. Categorization must be utilized again when determining subgoals, and recategorization may lead to further insight. Finally, monitoring progress and determining when a solution has been reached relies on this sort of abstraction. It can be seen that problem space theory most certainly utilizes several forms of categorization. As such, it is an excellent framework for the comparison of ill- and well-defined domains.

2 EXPERT/NOVICE DIFFERENCES IN CATEGORIZATION RESEARCH

A wide body of knowledge and its associations guide how such items are perceived and exemplified in reasoning. Categorization is studied chiefly in these simple contexts within ill-defined domains. This allows for relatively precise control over expert and novice reasoners, allowing novices to exhibit a set of default knowledge gleaned from everyday experience. Over several knowledge areas such as folk-biology and food, research has shown key differences between expert and novice categorization.

The majority of categorization research has explored the representation of concepts within a specific domain. Using comparisons between human behavior and similarity based models of category membership, research shows that most educated subjects tend to categorize using loosely based taxonomies (Medin, Lynch, Coley & Atran, 1997). Much of this work explores the nature of folk-biological categorization—general purpose knowledge about biological types such as plants and animals. Novices (usually subjects with relatively little classroom biological experience) show a consistent taxonomic representation of these concepts based on surface features (Heit & Bott, 2000). A typical biological taxonomy might include the path *mammal* → *dog* → *beagle*. This is intuitive and allows for relatively simple inference. For instance, subjects will categorize based on the amount of diversity between examples (Figure 1), and within base level categories (dog) based on example similarity (Figure 2).

the koala bear has property x
the whale has property x
therefore...
the monkey has property x

Figure 1: Diversity Based Induction
A strong generalization to the category of 'mammal'

the carp has property X
the bass has property X
therefore...
the salmon has property X

Figure 2: Similarity Based Induction
A strong generalization to the category of 'fish'

Expertise is the primary means of acquiring further knowledge and relational information. Ross & Murphy (1999) found that the use of food categories show several different modes of categorization[1]. The study showed that subjects are more likely to group based on taxonomic categories (e.g. fruit) than script (e.g. breakfast foods), but applied them consciously and differentially based on contextual cues. Ross & Murphy also demonstrate that experts, with their abundance of relations, are able to construct ad-hoc

1. Food is encountered daily by the general population. The interaction obtained from this constant exposure is sufficient to claim expertise.

categories ("foods that make good projectiles") with little difficulty. These differential applications apply to general contextual areas rather than specific item instances. One study (Shafto & Coley, 2003) found that fish experts' categorization of novel properties to show high use of taxonomic information. Associations between animals and novel diseases, however, received ecological characterization. Expert categorization, then, is based on a large number of relational associations between knowledge and their differential application based on contextual cues.

Furthermore, expertise is based primarily on domain specific knowledge rather than superior processing abilities. Several comparisons of two differing groups of experts have demonstrated that increasingly rapid categorization stems from additional knowledge and that such experience only transfers when the domains are highly similar (Tanaka & Taylor, 1991). Chess experts are able to perform chunking (which, as discussed in section 1.2, is a form of dynamic categorization) only on positions that might occur in a normal play—when presented with a randomly scattered board, performance is akin to novices (de Groot & Gobet, 2003). Intuitively, different forms of expertise in cross-domain tasks lead to unique solution paths (Ormerod, Fritz & Ridgway, 1999). Simply put, experts only have high levels of performance in a narrow field—the additional information does not automatically transfer to other, contextually different, tasks.

These results suggest that expertise not only increases knowledge and the relations between such knowledge, but is dependent on the context of the problem. While taxonomies are often seen as a default reasoning scheme in Western cultures (Medin et. al, 1997), accretion of knowledge highlights other, less superficial relations. How this distinction manifests itself in knowledge representation is less clear. A simple explanation suggests experts may simply have more information from which to categorize and infer information. The use of non-default relations, then, is largely a matter of their presence (Medin & Smith, 1984). New theories, however, suggest that these differences may be due to the salience of the relations that lead to specific category assignments. In other words, novices may have some knowledge of categorization methods employed by experts, but their use requires additional cognitive resources.

Initial work has shown, in simple cases, evidence of novices reasoning in the same fashion as experts when cues for non-default relations are made explicitly available (Shafto & Coley, 2003). It appears that the naive novice may have a variety of loosely established relations between categories. Any difficulty in determining non-default relations between items may be due to the default relations prominence. It is more likely, however, that novices simply do not recognize the relation without significant additional processing, even though the necessary relations are present. When given the relation explicitly, the processing has already been performed.

2.1 Categorization as Simplified Problem Solving

Routine tasks (e.g. a math problem involving simple addition) rarely require intensive processing—once the problem has reached a level of automaticity, little resources are spent exploring the task. Instead, it is recognized as a certain type of problem and the necessary actions are performed with relatively little further evaluation. Automaticity requires the knowledge of a previous problem and its solution, but the only processing performed is the classification of the problem and use of associated knowledge—the rest is primarily memory retrieval (Ross, 1996).

The above results follow many of the same experimental paradigms—simple induction, pairwise similarity ratings and grouping tasks—all of which probe how an item is categorized under certain constraints. In essence, this is a direct parallel of routine problem-solving. The execution of the actions to reach the desired goal are omitted—the category itself is the goal—but success or failure in a particular categorizational strategy is reliant on knowledge and the availability of relations.

Categorization then, can be seen as a first step in general problem-solving tasks, and the task in problems that have become rote or automatic.

The current body of categorization research offers little in terms of its connections to problem-solving. This is partially due to the ill-defined domains used through the literature. Because many categorizations are 'correct', it is difficult to quantifiably measure performance. It is also similarly difficult to construct problems in these areas. As such, each area's explanation of expert/novice differences appear to conflict - the notion of relational availability is highly different from explanations of expertise in problem-solving. These theories, however, represent two complementary processes rather than contradictory explanations. Categorization, as described by the literature above, is a type of simple problem-solving, augmented by a more broad restructuring of knowledge and its subsequent transfer.

3 CATEGORIZATION WITHIN PROBLEM SOLVING

3.1 The Role of Analogical Transfer

If basic, automatic problem-solving is primarily categorization and memory retrieval, more intense tasks are those which cannot be immediately classified - the problem space must be explored to reveal further insight. The process of analogical transfer provides the means to integrate information gleaned from this exploration and link it with prior knowledge to obtain a new representation of the problem. It is the process of aligning a source representation (prior knowledge) to the novel problem. This is performed by mapping specific features—aligning or rejecting certain similarities—and comparing their relevance (Holyoak & Thagard, 1995).

The classic "tumor" experiment illustrates this concept as well as the influence of transfer on ability. Subjects were given a problem involving the elimination of a cancerous tumor. Additionally, some subjects also received a story about the siege of a castle—a story containing a problem solved in analogous fashion. 10% of subjects given only the tumor problem solved the problem. However, 75% of those given the analogous story successfully derived a solution to the tumor problem (Gick & Holyoak, 1983). Subjects receiving knowledge which could be indirectly applied to a novel setting discovered a correspondence between non-obvious features in each. Activation of this knowledge during the testing phase allowed for fluid transfer of the story's solution to a novel problem.

This experiment is a excellent example of analogical transfer's reliance on categorization. In order to observe its relation, it is illustrative to postulate specific instances of this phenomena in the context of the above experiment. Following the initial reading of the problem, subjects attempt to classify the problem features. As subjects had little formal medical experience, this initial categorization might be something as ostensibly useless and abstract as 'medical diagnosis'. This, however, narrows the problem space significantly—already the problem solver has eliminated a wide range of possible solutions, making the task far more tractable. Since this classification does not immediately provide a means to a solution, exploration must occur—the subject examines the idiosyncrasies of the problem and its constraints. The relation of soldier and radiation beam, for example, can be categorized as an 'attacking force'. When each salient aspect is sufficiently classified, the (successful) problem solver has abstracted the problem away from its surface features, creating a generalized schema. The mapping to the previous solution then becomes apparent and the problem can be completed. Indeed, this process is one of continual categorization and abstraction.

Manipulations that induce expert-level behavior in categorization research, however, are far less successful in encouraging successful analogical transfer. The priming of relations that cause differential categorization behavior (see section 2) does not generalize to analogical transfer. Several laboratory studies demonstrate difficulty in encouraging novices to draw analogical inferences. Explicitly providing subjects with the abstract concept used in transfer is far less effective than having the subject explore

the mapping themselves (Gentner, Loewenstein & Thompson, 2004). Additionally, encouraging subjects to reason analogically hurts performance under some circumstances (Ormerod, 2003). Analogical transfer is reliant on the quality of relations and knowledge used to draw mappings—explicit direction under non-trivial settings, therefore, does not produce expert level reasoning. Providing novices with assistance without the underlying contextual information (built from training or practice) is not sufficient to aid transfer. While analogical transfer is reliant on categorization, thorough knowledge in relation to the complexity of the problem is required to facilitate its use.

The process of analogical transfer can be seen as mechanism to bridge an unfamiliar problem to previous knowledge. By categorizing aspects and structure of problems, prior knowledge can be applied to novel tasks to further meld them into a previously seen problem. It is clear that categorization and analogical transfer interact to aid difficult problem-solving. The next section explores this interaction through two influential studies.

3.2 CATEGORIZATION IN CLASSIC PROBLEM SOLVING STUDIES

3.2.1 Reframing Chi et al

The direct study of problem categorization remains fairly rare, but the research of Chi et al. remains one of the quintessential reference points in this area. Using physics problems, the study showed the differential effects of experience on categorization in formal domains. Undergraduate novices with relatively little knowledge were compared with physics graduate students in the groupings of four different problems. These stimuli were specifically controlled on two orthogonal dimensions, the objects used in the problem (e.g. an inclined plane) and the underlying principle necessary to derive a correct solution. Subjects were then asked to group sets of similar problems. Experts overwhelmingly derived similarity based on fundamental principles while novices attended to the distracting surface features (Chi et. al, 1981).

This classic result, despite similarity in format to ill-defined categorization experiments, elicits vastly different explanations. In the example of physics and other similarly academic domains, expertise is seen to embody a radical change in conceptual and domain specific knowledge (Pretz, Naples & Sternberg, 2003). When learning occurs, relations are reformulated as practice reveals flaws in current reasoning. These theories also place primary emphasis on knowledge itself, rather than the ability to manipulate and recognize relations.

Much like the studies discussed in section 3.1, it would appear that explicitly providing these principles to novice reasoners (who lack the necessary experience of their application) would do little to induce differential categorization. One would presume that analogical transfer is essential in this categorization, as the deeper structure must be gleaned from ostensibly different problems.

Ross (1996) showed that interaction with a problem facilitates discover of these deeper relations. Yet, no manipulation of the problem is performed in Chi et al. as to aid such a distinction. One possible explanation is that the problems were sufficiently simple to be rote to the particular experts—the underlying principles used in such problems, though the inculcation of formal schooling, have become especially salient categorization information.

This has two implications. First, novices[2] who are asked to manipulate the problem may categorize similarly to experts. This may not even require knowledge of the underlying principle as in Ross (1996)—exploration may provide the means to discover sufficient relations. This crude analogical mapping may not be sufficient to solve the problem, but might alter categorization.

Second, experts may group as novices in more difficult problems when asked to categorize without manipulating the problem (in a speeded task, for example). Such a result would be highly consistent with a variety of categorization studies in ill-defined domains such as music (see Coley, Shafto, Stepanova & Barra, 2005, for a review).

It is possible, then, that the difference between Chi et al. and simple categorizational studies is mainly one of complexity. The relative difference in skill between experts and novices is far greater in Chi et al than canonical categorization studies. Simple manipulations in the latter highlight already present (but less salient) relations (Shafto, Coley, & Baldwin, submitted). This is especially true in domains such as folk-biology or food where novices have a comparatively larger knowledge base through general experience. Ostensibly more difficult tasks, such as physics problems, should be far less sensitive to the simple manipulations seen in categorization experiments—subjects lack the prior knowledge to make use of additional information. Problem complexity, then, rather than domain, is largely responsible for these apparent differences.

3.2.2 Increasing Problem Difficulty

Problem-solving research has provided several explanations of how task

2. Novice subjects here assume a certain level necessary elementary mathematical and physics background. Without such a foundation, it may be extraordinarily difficult to aid deep categorization in laboratory setting.

structure influences difficulty. Kotovsky, Hayes & Simon (1985) examined three increasingly veiled isomorphic instances of the Towers of Hanoi problem. Isomorphic problems contain a mapping between elements of the problem and, in this case, solving the problem required the same sequence of moves. In this experiment, the 'towers' were replaced by items such as 'acrobats' and the instructions were changed accordingly. When given a random instance of the problem, subjects found the deliberately obfuscated variations significantly harder regardless of the fact that the underlying principle was known.

Isomorphic problems are especially useful in studying analogical transfer and this particular study is revealing of its reliance on categorization. The task of solving a particular isomorphic instance is largely a matter of the subjects' ability to categorize it as the Towers of Hanoi.[3] This is not automatic without the sort of expertise seen in Chi et al.—although subjects were exposed to the original puzzle before testing, they are considered to be operating at the novice level (Kotovsky et al., 1985). Deep processing—exploration and transfer—is required to obtain a mapping and distracting features block transfer through incorrect mappings. The time differences in isomorphic instances may be attributed on the need to search through an increasing number deliberately misleading features.

In the Tower of Hanoi (and its isomorphic variants) categorization interacts heavily with manipulation of state—exploration aids recognition and increasing availability of the problem schema. The success or failure in mapping a particular feature to a previously seen pattern leads to recategorization. This classification may lead to a small change, perhaps illuminating another aspect to be explored, or may relate to the original problem/solution sufficiently to make the mapping apparent. While the latter (often described as conceptual 'slippage') has long been a topic of discussion in analogy research (Hofstadter, 1995; Holyoak & Thagard, 1995), the emphasis here is that such shifts occur due to recategorization following a fruitful path of exploration. Difficult problems prevent these explorations and may cause misclassification, increasing the resources needed to ultimately arrive at an acceptable solution.

4 IMPLICATIONS

4.1 The Gap Between Ill and Well Defined Domains

The preceding sections have surveyed the sizable bodies of research con-

3. Resolving the problem from 'first principles' may be possible as well. Analogical transfer is considered primary here as it is an excellent example of reuse of knowledge through categorization. Arguments as to which is more intelligent or creative, while not irrelevant, is beyond the scope of the current work.

cerning ill and well-defined tasks and established categorization as a fundamental aspect of problem-solving. There remains, however, a rather large disparity in theories of performance and expertise in each area. Categorization theories performed in ill-defined areas rely on the ideas of instance typicality and similarity metrics. Studies of formal domains cite a pervasive change in knowledge structure. These appear to be mutually exclusive, but it possible this is an artifact of the problems used in each domain. In this section, an alternate theory is proposed, integrating both explanations in a two stage process.

Initial categorization, the recognition of the problem as an instance of one or more categories, is influenced heavily by relations between knowledge. The process filters the breadth of relations based on the salient features of the problem, allowing them to be used in further processing. This is much akin to semantic activation. Strong relationships between knowledge reveal the most important aspects the problem and these are highlighted for use in further exploration. If this categorization is sufficient to solve the problem, the procedure to do so need only be recalled.

When categorization does not facilitate an instant solution, this initial processes greatly reduces the problem space.[4] However, further exploration is necessary - examining or manipulating certain aspects of the task in order to obtain an analogical mapping. Whereas the initial categorization focuses on breadth, analogical transfer provides a mechanism for more fine-grained analysis. When this mapping is explored, it facilitates a new categorization of problem state. These processes, then, are both cyclical and complementary, a continual reduction of problem until it can be categorized in a sufficiently concrete way that previous knowledge provides a solution.

Section 3 shows that problem-solving in well-defined domains is well suited to such a theory. Research in ill-defined domains has, to a great extent, however, ignored this second aspect of processing—analogical transfer and subsequent exploration. When grouping a pair of animals together, simplistic relations define most performance differences. These can be manipulated easily, showing their lack of complexity. More difficult categorizations requiring some sort of analogical transfer should be far less susceptible to such changes.

It is hypothesized that this theory holds across both ill- and well-defined domains—it is abstracted from knowledge of the transitions and goals and has strong foundations in the categorizational reasoning experiments where most ill-defined problems are found. Highly complex ill-defined

4. It is important to note that the novice, however, may reduce the problem to unimportant features, making transfer of the problem more difficult.

domains, however, hold the final piece of this puzzle. Future research must show that analogical transfer processes are used in ill-defined tasks and that relational availability is not sufficient for difficult tasks.

4.2 Suggestions for Future Empirical Research

It is clear that categorization studies must go beyond their current scope and include not only the singular act of categorization, but its use in tasks that are considerably more difficult. The study of this process must use tasks that cause subjects to interact with a problem space as in Ross (1996). By modifying both the typicality of stimuli and distracting features that might obfuscate the discovery of relational information, it may be possible to replicate Kotovsky's (1985) work with problem isomorphisms under ill-defined domains. If such a replication can be achieved, it will provide extremely strong support for the ideas presented here.

This paper focuses on analogical transfer as a key connection between initial categorization and the transformation of previous knowledge. If categorization is to be fully integrated into theories of problem-solving, it also must be fully related to other theoretical problem-solving methods (means-end analysis and case-based reasoning, to name a few). These considerations are discussed here mainly in terms of the abstract notion of exploration that occurs between categorization.

5 CONCLUSION

The nature of differences in high level cognitive processes when solving ill and well-defined problems remains an open research question. Although comparative research in this area is severely lacking, this paper presents categorization as a process in problem-solving that is independent of domain. It aligns problem-solving literature, studied primarily in well-defined domains, to categorization research in ill-defined areas through reinterpretation of the research of each area. By framing categorization as simple problem-solving while simultaneously revealing problem-solving's dependence on categorization through analogical transfer, it is possible to compare expert/novice behavior in both domain classes.

The reexamination of several key studies casts some doubt on the differential characterization of ill and well-defined domains seen in theories of conceptual acquisition and expertise. The use of categorization appears to be fundamentally similar in both research areas. While often mistaken for differences in thought, the primary disparities between the study of these two domains are complexity and procedure. Problem-solving studies allow subjects to categorize, manipulate, explore and then use an enhanced representation to repeat this process. Categorization research attempts to determine the nature of knowledge structure by exploring only categorization under simple circumstances. To account for these factors, this paper has presented an integrative hypothesis. It that postulates that basic categorization is an initial and rapid reduction of the problem space. Analogical transfer then guides a fine-grained reduction and exploration. Further empirical exploration of problem complexity in both types of problems through replication of cross-domain research is needed to investigate the theory's feasibility.

If skilled categorization is, indeed, domain independent, current theories of task differences may be questioned. It is possible that categorization is one of only a few fundamental processes involved in solving both well and ill-defined tasks. It is equally plausible, as demonstrated throughout the paper, that the differences in ill and well-defined domains are illusory. The latter possibility would contradict traditional approaches to problem-solving and further undermine the differential characterization of cross-domain concept acquisition. Thorough examination of this possibility across a variety of mental processes would allow ill and well-defined domains to be characterized not by an abstract and somewhat arbitrary notion of 'definedness', but by their reliance on specific and concrete cognitive processes.

references

Brabeck, M. M. & Wood, P. K. (1990). Cross-sectional and longitudinal evidence for differences between well-structured and ill-structured problem-solving abilities. In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, T. A. Grotzer, & J. D. Sinnott (Eds.), *Models and methods in the study of adolescent and adult thought.* New York: Praeger.

Chi, M. T. H., Feltovich, P. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.

Coley, J. D., Shafto, P., Stepanova, O. & Barraff, E. (2005). Knowledge and category-based induction. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & and P. W. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin.* Washington, DC: APA.

Gentner, D., Loewenstein, J. & Thompson, L. (2004). Analogical encoding: Facilitating knowledge transfer and integration. *Proceedings of the Twenty-sixth* Brabeck, M.M.

Brabeck, M. M. & Wood, P. K. (1990). Cross-sectional and longitudinal evidence for differences between well-structured and ill-structured problem-solving abilities. In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, T. A. Grotzer, & J. D. Sinnott (Eds.), *Models and methods in the study of adolescent and adult thought.* New York: Praeger.

Chi, M. T. H., Feltovich, P. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.

Coley, J. D., Shafto, P., Stepanova, O. & Barraff, E. (2005). Knowledge and category-based induction. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & and P. W. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin.* Washington,

DC: APA.

Gentner, D., Loewenstein, J. & Thompson, L. (2004). Analogical encoding: Facilitating knowledge transfer and integration. *Proceedings of the Twenty-sixth Annual Meeting of the Cognitive Science Society.*

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.

de Groot, A. & Gobet, F. (2003). Perception and memory in chess. *Heuristics of the professional eye.* Assen: Van Gorcum.

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *The psychology of learning and motivation* (pp. 163-199). Academic Press.

Hofstadter, D. R. (1995). *Fluid concepts and creative analogies.* NY: BasicBooks.

Holyoak, K J. & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Kotovsky, K., Hayes, J. R. & Simon, H.A. (1985). Why are some problems hard? Evidence from the Tower of Hanoi. *Cognitive Psychology*, 17, 284-294.

Lopez, A., Atran, S., Coley, J. D. & Medin, D. L. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, 32, 251-295.

Medin, D. L., Lynch, E. B., Coley, J. D. & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, 32, 49-96.

Medin, D. L. & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, 35, 113-138.

Newell A. & Simon, H. A. (1972). *Human problem-solving.* Englewood Cliffs, NJ: Prentice-Hall.

Ormerod, T.C. (2003). Planning and ill-defined problems. In R. Morris & G. Ward (Eds.), *The cognitive psychology of planning. London*: Psychology Press.

Ormerod, T. C., Fritz, C., & Ridgway, J. (1999). From deep to superficial categorization with increasing expertise. *Proceedings of the Twenty-first Conference of the Cognitive Science Society*, Vancouver, 502-506.

Pretz, J. E., Naples, A. J. & Sternberg, R. J. (2003). *Recognizing, defining and representing problems: The psychology of problem-solving.* Cambridge: Cambridge University Press.

Richman, H. B., Gobet, F., Staszewski, J. J. & Simon, H. A. (1994). Perceptual and memory processes in the acquisition of expert performance: The EPAM model. In K. A. Ericsson (Ed.), *The road to excellence.* Mahwah, NJ: Lawrence Erlbaum.

Ross, B. H. (1996). Category learning as problem-solving. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation* (Vol. 35). New York: Academic.

Ross, B. H. & Murphy, G. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, 38, 495-553.

Shafto, P. & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 641-649.

Shafto, P., Coley, J.D. & Baldwin, D. Availability and Novice Reasoners. Memory & Cognition. Submitted.

Shin, N., Jonassen, D. H. & McGee, S. (2003). Predictors of well-structured and illstructured problem-solving in an astronomy simulation. *Journal of Research in Science Teaching.*

Tanaka, J. B. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23, 457-482.

Damasio Without Error: On uniting emotion and reason while separating the neural and the biochemical

Govind Persad Stanford University

Antonio Damasio's extensive popular work on neuroscience and its philosophical and social implications has argued—in intelligent, consistent, and poetically elegant prose—that much modern thought about consciousness and mental processes unthinkingly accepts a sort of Cartesian dualism. Such a dualism separates mental processes from physical ones; further, within the realm of the physical, it separates the brain from the body. According to Damasio, these dualistic constructions fail to do justice to the complex and interrelated nature of consciousness. Despite all his careful work, however, Damasio does not entirely escape the Cartesian error of creating dichotomies where none genuinely exist, nor does he always pick the right dichotomies to challenge. I will reject Damasio's division of reason and emotion, while upholding a scientifically defensible version of the separation between brain and body against his criticisms. I will then analyze the philosophical implications of my suggested alterations to Damasio's view.

Given that Damasio's project has been to overturn scientifically unsupported views such as Cartesian dualism, he can and should extend his project by undermining the last such remaining dichotomy, that of emotion and reason. Currently, Damasio's work always respects that dichotomy,[1] even as it depicts the many ways in which emotion and reason interact. Damasio's underlying project in fact ought to lead him to endorse a merger of emotion and reason, rather than continuing to see the two as separate. In contrast, Damasio's attempt to break down the dichotomy of brain and body fails because some kinds of events, those which occur in substantial aggregations of neural tissue such as the human brain, cannot happen in other parts of the body because of physical limitations and differences. This empirical fact supports a genuine dichotomy, buttressed by evidence from the biological and physical sciences, between neural processes and standard biochemical ones. I propose to separate the processes Damasio discusses into two classes: not reasons and emotions, but biochemical processes and neural processes. Only neural processes are fast enough to be involved in practical decision-making on the time scale that Damasio considers.

The replacement of this last bastion of the Cartesian error, the separation of emotion and reason, with an scientifically grounded theory that embraces both emotion and reason while differentiating between neural and biochemical activity has helpful implications for current philosophical theories that attempt to explain emotion's role in the project of practical decision-making and reasoning. Fusing emotion and reason can make such a theory more plausible by eliminating the claim that emotion must act separately from reason in decisionmaking. Further, this new concept, supported as it is by empirical fact, can ground ethical values better than either emotion or reason can alone.

The Unjustified Dichotomy: Reason and Emotion

Damasio implicitly introduces the separation of emotion and reason at the opening of his discussion of consciousness in *The Feeling of What Happens*:

> *[I]t is easy to envision how consciousness is likely to have opened the way in human evolution to a new order of creations not possible without it: conscience, religion, social and political organizations, the arts, sciences, and technology. Perhaps even more compellingly, consciousness is the critical biological function that allows us to know sorrow or know joy, to know suffering or know pleasure, to sense embarrassment or pride, to grieve for lost love or lost life. (1999: 4)*

Here, Damasio realizes that consciousness encompasses both qualities we traditionally think of as reason-based, such as the ability to bring about

1. Consider the title of his bestseller: "Emotion, Reason, and the Human Brain."

technological advances, and those we think of as emotional, such as the sensitivity that enables knowledge of pleasure and pain. Damasio also realizes that, because of this, "consciousness and emotion are *not* separable" (1999: 16). Despite this realization, Damasio does not make the further inference that reason and emotion are not separable. Rather, he chooses, from the outset, to separate these two central aspects of consciousness. Damasio describes reason as a cognitive function (1999: 18) that interprets the processes of the brain.

Despite accepting this separation of reason and emotion at the start of his work, Damasio by no means discounts emotion's importance. Damasio discusses the history of emotion and its dismissal by the Cartesian philosophical tradition: "Philosophy, notwithstanding David Hume and the tradition that originates with him, has not trusted emotion and has largely relegated it to the dismissible realms of animal and flesh" (1999: 40). Here, Damasio grasps the problems inherent in the elimination of emotion, criticizing the separation of emotion and reason that has characterized 20th-century science:

> *Throughout most of the twentieth century, emotion was not trusted in the laboratory. Emotion was too subjective, it was said. Emotion was too elusive and vague. Emotion was at the opposite end from reason, easily the finest human ability, and reason was presumed to be entirely independent from emotion. This was a perverse twist on the Romantic view of humanity. Romantics placed emotion in the body and reason in the brain. Twentieth-century science left out the body, moved emotion back into the brain, but relegated it to the lower neural strata associated with ancestors who no one worshiped. In the end, not only was emotion not even rational, even studying it was probably not rational. (1999: 39).*

Damasio's neuroscientific investigations have challenged much of this older view about emotions. As he states, "work from my laboratory has shown that emotion is integral to the processes of reasoning and decision making, for worse and for better" (1999: 41). Thus, according to Damasio, emotion constitutes a central part of reasoning: the loss of emotions entails the loss of the ability to make rational decisions, at least within the personal domain (1999: 41).

Nonetheless, the persistence of the division between reason and emotion remains clear when Damasio discusses certain neurological disorders: an effect of these disorders, he states, is that "the delicate mechanism of reasoning is no longer affected, nonconsciously and on occasion even consciously, by signals hailing from the neural machinery that underlies

emotion" (1999: 41). In Damasio's picture, even before the onset of disease, reasoning's "delicate mechanism" is conceptually and neurally distinct from emotion's "machinery," and requires external signals from emotion in order to respond to emotional concerns. This already seems a retreat from his earlier assertion that emotion is *integral* to reasoning—he now seems to be making the weaker assertion that emotion *affects* reasoning. Damasio continues this careful separation in his analysis of his results:

> *I suggested that certain levels of emotion processing probably point us to the sector of the decision-making space where our reason can operate most efficiently. I did not suggest, however, that emotions are a substitute for reason or that emotions decide for us. (1999: 42)*

Damasio's rebuttal of the anti-emotional doctrine is both substantiated and incisive, up to the point that he runs into the inviolate barrier between reason and emotion. Part of the reason for this failure to take the next step may be the overwhelming weight of bias against emotion in the history of philosophy and psychology. Once Damasio believes that he has established the salience of emotion, he may feel obliged to rest there to avoid overextending his case. However, as de Sousa suggests, emotion and reason seem connected in an important way that defies easy explanation:

> *In a more pervasive and less easily definable way, the capacity to experience emotion seems to be indispensable to the conduct of a rational life over time. Antonio Damasio (1994) has amassed an impressive body of neurological evidence suggesting that emotions do, indeed, have this sort of function in everyday reasoning. (1)*

Damasio's evidence therefore suggests that he ought to press his point further and break down the dualism of reason and emotion.

Damasio's failure to realize and demolish the barrier between reason and emotion rests on an equivocation on the meaning of the concept of reason itself. Damasio defines reason in *Descartes' Error* as follows: "I generally use reason as the ability to think and make inferences in an orderly, logical manner" (1994: 269). Yet, on this view, the neurologically damaged persons whom Damasio describes as having lost reason as a result of the destruction of their emotional abilities would in fact retain reason, since "Their ability to tackle the logic of a problem remains intact" (1999: 41).

How, then, should we speak of these persons, given that they have reason under definitions that exclude emotion, but not under those which include emotion as a requirement? It seems that the term "reason" is the problem here. This is a problem that Damasio himself realizes: "In English, the word 'reason' has long had, and still has, a large number and a wide variety of senses and uses, related to each other in ways that are often complicated and often not clear" (1994: 270). Damasio at once says:

> *It is perhaps accurate to say that the purpose of reasoning is deciding, and that the essence of deciding is selecting a response option, that is, choosing a nonverbal action, a word, a sentence, or some combination thereof, among the many possible at the moment, in connection with a given situation. (1994: 165)*

On this definition, it seems like, insofar as emotion is involved in making choices among options, it is a part of reasoning. However, Damasio argues that reasoning must also involve a "logical strategy" (1994: 166), which seems to exclude emotion from immediate involvement and render it a bystander or, at best, a mere supporting process (1994: 166).

Given the confused state of the term "reason," and the unclear relationship between emotion and each of the different definitions, perhaps the very subdivision of reason from emotion is contributing to the problem. A new concept could solve this problem by uniting the two attitudes under one banner, but what new, clearer concept could incorporate the contributions of emotion while replacing the current unclear concept of reason? I am enticed by Damasio's idea of "higher reason" (1999: 55), which incorporates both emotional and logical responses into a holistic conception of practical reason. However, giving the new concept a moniker that incorporates only "reason" and not "emotion" perpetuates the linguistic problem rather than solving it. Therefore, the new concept, since it ought to bring together both reason and emotion as equal partners, should involve either both terms or neither term. The first option—a term like "reasoned emotion" or "emotional reason"—seems both to be vacuous and to perpetuate the dichotomy. Meanwhile, the separation of reason and emotion is so entrenched in common vocabulary that it is difficult to find a term that does not seem to reinforce the dichotomy from one side or the other: "rationality," "thinking," and the like seem to sway towards reason, while "feeling" or "intuition" seem to bias the term towards emotion.

Given the difficulty of finding neutral language, as well as a desire to ground the new term in empirical evidence, I propose the concept of *practical cerebration* as a term that encompasses the aspects of both emotion and reason that are useful in decision-making. This unusual term has the advantage of avoiding the language either of emotion or of reason, in

favor of purely empirical description of the process's nature and purpose.[2] Cerebration implies a biological location for the process: the process in question is something only performable in a specific neural tissue, rather than anywhere in the body as a whole. While this seems a problem for other parts of Damasio's theory, the distinction between neural processes and ordinary body processes is a defensible one. All processes involved in practical decision-making, whether they were previously called emotional or logical, share one feature: they are processes that take place in neural tissue. In the next section, I defend this neural-biochemical distinction against Damasio's challenge.

The Not-So-Somatic Marker and the Nerual-Chemical Distinction

Armed with this new concept of practical cerebration, I propose to investigate Damasio's somatic-marker hypothesis, which attempts to show that emotional responses affect and inform reasoning. While Damasio's conclusion that emotion affects reasoning is sound, Damasio's separation of reason and emotion leads him to an unjustified conclusion: that emotion, unlike reason, involves slow biochemical processes in the body. In contrast, I believe that some emotions necessarily involve the body, while others can be duplicated just as well using purely neural means. Only the latter sorts of emotions are involved in decisionmaking.

Damasio proposes the somatic marker theory in *Descartes' Error* as follows:

> *When the bad outcome connected with a given response option comes into mind, however fleetingly, you experience an unpleasant gut feeling. Because the feeling is about the body, I gave the phenomenon the technical term somatic state ("soma" is Greek for body), and because it "marks" an image, I called it a marker. Note again that I use somatic in the most general sense (that which pertains to the body) and I include both visceral and nonvisceral sensation when I refer to somatic markers. (1994: 173)*

The way that Damasio sets up the theory leaves the possibility entirely open that the "somatic" marker is not a feeling in the body, but merely a feeling about a body that may or may not actually be connected. Damasio explicitly constructs somatic markers in neural terms (1994: 177). Their

2. Interestingly, cerebration does not necessarily imply consciousness of the deliberative process. The OED defines it as: Brain-action. First used by Dr. W. B. Carpenter in the phrase unconscious cerebration, to express that action of the brain which, though unaccompanied by consciousness, produces results which might have been produced by thought. My use of it does not imply unconscious cerebration, though that possibility is interesting.

main center is the prefrontal cortex, which is part of cerebral tissue; so far, their somatic aspect remains quite unclear. However, Damasio's claim that somatic markers are actually in the body is an extremely tenuous one. He realizes the possibility that the somatic marker could work without body involvement:

> *In this alternative mechanism, the body is bypassed and the prefrontal cortices and amygdala merely tell the somatosensory cortex to organize itself in the explicit activity pattern that it would have assumed had the body been placed in the desired state and signaled upward accordingly. The somatosensory cortex works as if it were receiving signals about a particular body state, and although the "as if" activity pattern cannot be the same as the activity pattern generated by a real body state, it may still influence decision making. (1994: 184)*

Damasio's conclusion, however, is the opposite of what the facts show: the cortex certainly influences decision making, and the body *may* but does not certainly do so. Damasio states that "To what extent we depend on such "as if" symbols rather than the real thing is an important empirical question" (1994: 184); given the speculative nature of his evidence, the onus is on Damasio to prove that the somatic marker ever genuinely works via the body.

Note, first, that the speed of practical decisionmaking is too fast to be achieved by non-neural biological processes. Even neural processes outside the brain may be too slow: the top speed of nerve impulses is 100m/s, while impulses conveying thought travel at approximately 20-30 m/s(REF), which is arguably not fast enough for choices that occur in a matter of milliseconds.(REF?) Furthermore, the secretory and other endocrine processes that would need to occur in the body to actually produce the necessary somatic state (such as a bad gut feeling) would slow the turnaround time much more than this, likely taking a matter of minutes. Yet, Damasio stipulates that the reasoning process will be accelerated by the somatic marker: "Imagine that you are asked to say yes or no quickly, in the middle of other distracting business" (1994: 174). If we are using the somatic marker to accelerate our reasoning, why not use our fastest reasoning tool—our brain? It seems that, if the "as if" pathway exists, there is no parsimonious reason not to use this quick pathway rather than going through the slower body loop. If my criticism is true, we can make a defensible distinction between brain and body: not the folk-biological one of the Cartesian tradition, but rather the modern distinction between cerebral and biochemical pathways, a distinction borne out by in-depth

biology rather than superficial external resemblances or differences. The term "somatic marker" is, on this reading, a misnomer. The marker is never truly somatic; in order for it to be effective in decision-making, it must always be "as if" somatic.

A corollary to this rejection of the stronger reading (that the somatic marker actually involves the body) of the somatic-marker hypothesis, is that the dichotomy of reason and emotion that Damasio sets up, where reason is purely in the brain while emotion involves the body as well, is dismantled. Once emotion and reason prove not to be entirely different methods of responding to stimuli, but to be merely different aspects of the same neurophysiological system, the idea that emotion acts in both body and brain, while reason acts solely in the brain, must likewise be given up. Rather, quick-acting aspects of practical cognition—such as the somatic-marker response that Damasio describes as aiding in decision-making—act at a primarily neural level, and thus primarily in the brain, while other responses, such as the sadness or depression that may result from harm to the body or from altered chemical states such as intoxication are acting at a much slower, biochemical level. While such states can help or hinder the brain's overall performance, they cannot offer particular assistance in decision-making on a short time-scale.

As further evidence for this view, consider Damasio's analysis of patients with locked-in syndrome. Damasio notes that locked-in patients seem to feel certain emotions differently than patients who are in full command of their bodies. He suggests that this change occurs because such patients do not have access to the body's mechanisms of representing emotions. It follows from Damasio's theory that if these patients really cannot access emotions through the body, their decision-making ability ought, according to the somatic-marker hypothesis, to be affected in some way. However, this ability remains, for all intents and purposes, unharmed: consider these patients' ability to produce not only entirely coherent but also breathtakingly beautiful and emotional creative works, such as Jean-Dominique Bauby's *The Diving Bell and the Butterfly*, and to maintain strong and healthy relationships with their loved ones. Their limitations are in the body, not in the decision-making process. This argument becomes even stronger when we consider the lack of overall effects on the reasoning abilities of persons who have parts of their bodies damaged or amputated. It does not seem that patients who have undergone gastric surgery that removes much of the stomach have their "gut feelings" altered in any way, even though the part of the body the somatic marker points to has disappeared. Rather, the brain simply creates a phantom of the body part that is missing.

Based on this evidence, it seems that the body, while it might conceivably be required for whatever overall responses are lost in locked-in syndrome, such as the terror of feeling paralyzed, is not necessary for the emotional responses that are essential to decision-making. In the case of locked-in patients the "as if body loop" seems able to represent all the emotions relevant to decisions, while the actual body loop is only involved with feelings, such as abject terror, that are not strictly necessary for decision-making. It is plausible that the "as if body loop" functions this way in all persons, and not merely as an adaptation in locked-in or otherwise disabled patients. Note, as further evidence along this line of thought, that the persons Damasio shows to lack the somatic-marker response are those with frontal lobe (cerebral) damage (1994: 208), not those with damage to nerves going to the body or damage to actual body parts. If the existence of the actual body, or, as Damasio says, the "body-proper," deeply affects our decision-making abilities, we would expect the latter group to be far more impaired than they are.

At this point, one could argue that I have merely replaced Damasio's dichotomy of reason and emotion with another dichotomy, one that separates neural processes from biochemical processes. As Damasio argues,

> *While the modern scientific coupling of brain and mind is most welcome, it does not do away with the dualistic split between mind and body. It simply shifts the position of the split. In the most popular and current of the modern views, the mind and brain go together, on one side, and the body (that is, the entire organism minus the brain) goes on the other side (2003: 190).*

How, if at all, is my view an improvement on the above? I am willing to grant the objection that I, like the scientists Damasio criticizes, have replaced Descartes' dichotomy with another. What I would say in response is that my goal is not to eliminate all dichotomies, but rather to eliminate empirically unjustified ones. Cerebral processes, whether classified as reason or emotion under Damasio's system, all share an empirically defensible resemblance. Likewise, biochemical (endocrine) processes also share such a resemblance. What Damasio ought to do is modernize rather than groundlessly uproot the Cartesian distinction of brain and body, by recognizing the different speeds possible for neural and biochemical processes. It is not the distinction between emotion and reason that is important—a distinction that lacks a biological, empirical basis. Nor is the arbitrary distinction between brain and body justified: the brain is simply an organ located at the top of the body. The secretory activity of the hypothalamus, for example, is an brain event, but it is not involved in

decision-making, because it is too slow a process to work on the time-scale necessary. In contrast to these two unfounded dichotomies—emotion/reason and brain/body—the distinction between cerebral computations and biochemical processes rests soundly on biological and physical facts.

One could give the further argument, "But aren't cerebral processes just biochemical processes?" Here, the interlocutor could accuse me of just committing Descartes' brain/body fallacy all over. How can processes carried out in the brain, which is made of the same sort of physical stuff as any other organ, be qualitatively different? Where does this magic notion of a cerebral process come from? However, there is a real, empirical difference here that is substantiated by experimental evidence. For example, some of Damasio's suggestions, such as the concept of the binding problem, support this distinction:

> *In terms of an overall mental picture it is likely that binding requires some form of time-locking of neural activities that occur in separate but interconnected brain regions. There is little doubt that the integrated and unified scene that characterizes the conscious mind will require massive local and global signaling of populations of neurons across multiple brain regions. (1999: 335).*

The aggregation of neurons together, as previously discussed, can create computational speed and power that is not possible for slower or more physically separated biochemical processes. The brain, unlike other organs, is specialized to carry out *purely* neural processes.

Despite this argument, Damasio's central conclusion, that emotion is necessary for effective decision-making, is still strongly supported and correct. So, what difference does my empirical dispute with Damasio about where emotions occur make? Our disagreement is relevant because it highlights the problem with the reason-emotion divide. If these two attitudes—reason and emotion—were not separated, there would be no reason to situate one in both body and brain and the other in the brain alone. As it is, what distinguishes emotion from reason on Damasio's account is that emotion, unlike reason, involves the body. If we no longer have to maintain the emotion/reason dichotomy, the motivation for the unwieldy *actual* body loop will go away, and emotion, insofar as it affects decision-making, can be accepted as part of the larger brain process of practical cerebration.

Philosophical Implications

Dismantling the dichotomy of reason and emotion suggests important implications for philosophy. In particular, the problem of explaining how emotions can affect moral sentiments and practical reason no longer exists, since emotions now simply *are part of* the practical reason in question. However, it substitutes a new challenge: what is the relation between practical cerebration and moral deliberation? My modification of Damasio's view, and the new concept it produces, suggests relevant and interesting tracks for philosophers to explore. By taking practical cerebration, rather than either reason or emotion alone, as a grounding for moral claims and claims of agency, the virtues of Kantian rationalism and Humean moral sentimentalism can be combined.

I will briefly consider the help that an empirically grounded redefinition of reason and emotion can lend to a theory like that advanced by Karen Jones in her "Emotion, Weakness of Will, and the Normative Conception of Agency." Since Jones' theory is explicitly inspired by empirical work (180), even citing a case very similar to Damasio's (1999: 43) David, new empirical evidence from a theory like Damasio's is certainly relevant. Jones advances the view that "our emotions sometimes key us to the presence of real and important reason-giving considerations" (181). She sees this fact as having implications for rational agency and decision-making:

> *In this paper, I want to explore the implications of the fact that emotions show varying degrees of integration with our conscious agency—from none at all to quite substantial -- for our understanding of our rationality (181).*

Jones then considers the implications of this emotion-considering view of rationality for a naturalist conception of rationality. She also realizes that "naturalist inspired reflection on what we know about how the emotions contribute to the practical rationality of finite creatures like us might lead us to reconceive norms of practical rationality—or so it would seem" (183). This reconception of rationality might possibly turn out to look something like the idea of "practical cerebration" proposed above as a reconceptualization of Damasio's theory.

Jones then reconstructs Nomy Arpaly's theory of emotion's role in decision-making. According to Arpaly, "well-functioning mechanisms capable of latching on to reasons can be sufficient for rationality" (Jones 188), even if these mechanisms are not logical or "all-things-considered" judgments. Jones, while she sees the appeal of Arpaly's account, is troubled by the implication that reason and emotion sit on the same level. In Jones' view, there must be something over and above mere emotional response to my actions in order for me to count as an agent:

> *Nor can I view my reasoning self—that part of me that engages in conscious deliberation about what to do or what to believe—as simply one additional epistemic mechanism operating side-by-side with other mechanisms such as perceptual or emotional ones. To think of myself in this way is not to think of myself as an agent at all. It is to give up thinking of myself as rationally guiding my actions via reason. (189)*

Because of this worry, she is obliged to include an unwieldy condition in which the dictates of emotion are overseen by some internal arbiter, a "reflective self-monitoring" system within the agent, that decides whether the reasons pointed to by emotion are genuinely the right reasons to act upon. The new concept of practical cerebration could help greatly here, by forestalling the worry that reason must operate concurrently with emotion but remain independent. Rather, the "reasoning self" above would reason using emotional as well as logical mechanisms. In a sense, the logical arbiter would be built into practical cerebration, rather than having to evaluate the reasons-responsiveness of actions from an external perspective. Jones would not have to view her concept of reason as coeval with emotion, but rather would be able to adopt a new definition of reason that directly involves emotion in the deliberative process. The actions resulting from deliberation would be "cerebrationally guided," rather than solely rationally guided.

While I do not see my response to Damasio as an explicitly feminist one, the project of practical cerebration also has the advantage of being responsive to some feminist concerns about the normativity of rationality. If, as some feminists (e.g. Gilligan 1982) claim, emotions are a greater part of women's lives for social or biological reasons, and are simultaneously devalued in our moral theory as motivations for action, women's voices are thereby placed at a disadvantage. In contrast, by placing both emotion and reason within a larger conception of practical cerebration, women's motivations are placed on an equal footing. Agents—male or female—who did not cerebrate effectively would be poor decision-makers, but I see no reason, even on a view like Gilligan's, that this new model, unlike the old one that separated reason and emotion and privileged the former over the latter, would tend to single out one gender as making unjustified decisions.

On a final note, the ability for practical cerebration suggests a concept, analogous to Kant's concept of rational nature in the *Groundwork* (20-21), that can serve as a ground for agency and moral responsibility. However, unlike Kant's concept, this new concept rests on a quality that is more uniquely human than either rationalism or pure sentiment is alone. Computers can calculate more effectively and animals suffer just as saliently, but neither (at least currently) is capable of integrating both reason and emotion into a coherent whole as humans can through the process of practical cerebration. This combination of input from non-logical sources, such as intuition, with input from logical computations is what makes humans such effective practical decision-makers. Thus, practical cerebration not only places the ideas of justified decision-making on firm empirical ground, but suggests a shared and uniquely human ability that gives moral worth to the ability to make decisions based on these reasons. A human being becomes an agent when he or she acts based on the concerns of practical cerebration.

Such an ethic would go beyond the rough-edged sociobiological speculation that Damasio employs in *Looking for Spinoza*, where he attempts to bring out the ethical implications of his theory of mind (2003: 172-75). Rather, it would reconcile Kant's attempt (20-21) to create an ethics that exalts the human property of rationality with the criticisms of those, like Hume (ch. 72), who have proposed the importance of emotion in morality. My ethic, in its goal of creating an empirically defensible theory of decision-making, would also add to current strands of ethical thought two aspects of Hume's theory: his moral sentimentalism and his associationist rather than separative theory of mind (ch. 40), which would give emotion importance and agree with bringing reason and emotion together insofar as they both occur within the brain. Such a theory fulfills the project of explaining the role of emotion in decision-making that Damasio has attempted to carry out, without committing itself to the separation of emotion and reason that undermines this ultimate goal.

references

Bauby, J. (1997). *The Diving Bell and the Butterfly: A Memoir of Life in Death.* Trans. Jeremy Leggatt. New York: Knopf.

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York: Harcourt Brace & Co.

Damasio, A. (1994). Descartes' Error: Emotion, Reason, and the Human Brain. New York: G.P. Putnam's Sons.

Damasio, A. (2003). Looking for Spinoza: Joy, Sorrow, and the Feeling Brain. New York: Harcourt.

de Sousa, R. "Emotion." The Stanford Encyclopedia of Philosophy. Ed. Edward N. Zalta. <http://plato.stanford.edu/archives/spr2003/entries/emotion/>

Gilligan, C. (1982). In A Different Voice: Psychological Theory and Women's Development. Cambridge: Harvard UP.

Hume, D. A Treatise of Human Nature. <http://etext.library.adelaide.edu.au/h/hume/david/h92t/>

Kant, I. (1997). *Groundwork for the Metaphysics of Morals.* Ed. and trans. Mary Gregor. Cambridge: Cambridge UP.

Jones, K. (2003). Emotion, weakness of will, and the normative conception of agency. In A. Hatzimoyisis, (Ed.), *Philosophy and the Emotions.* Cambridge: Cambridge UP.

Emergence in the Mind: On reductionism and the multiple realizability of mental states

Adrien Tough Simon Fraser University

Does the multiple realizability of mental states preclude their reduction to physical states? If a large rock were to fall on my new Martian friend's appendage just as we were making first contact, I would almost certainly believe it experienced pain as a result. It is often thought that mental states such as pain can be realizable in cognitive beings who are formed from a variety of different physical structures. In fact, it is this very property of mental states that is supposed to preclude their reduction to particular physical states, but the claim has been controversial and the debate rages on.

Jaegwon Kim (2002) rejects the multiple realizability of mental states on the grounds that the categories appearing in scientific laws cannot have disjunctive sets of causal powers. In other words, if different physical microstructures, each with a different set of causal powers, form the basis for some category then the subsequent causal powers of that category are just the disjunction of the separate causal powers of each different microstructure, and furthermore, such categories cannot be legitimate nomic kinds. He further argues for a reduction of the mental to the physical by restricting the identification of mental states to the activation of particular—possibly species-based—physical *structure-types.* I intend, however, to dispute the claim that the causal powers of different physical microstructures are necessarily disjunctive and then argue that this allows for genuine, multiply realizable nomic kinds. I will do so by arguing that the various structures' heterogeneous *micro*-causal powers *converge* on the same/similar *emergent macro*-causal powers. First, though, we must examine Kim's arguments and the claims of, as Kim refers to it, "The Multiple Realization Thesis" (p. 135) in general.

Simply put, The Multiple Realization Thesis comes from imagining that we can attribute various mental states, such as pain, to a variety of different types of creatures, from—trivially—other humans, to dogs, octopi, and even—non-trivially—Martians and possibly also robots some day. What this amounts to is the claim that mental states can be "realized" or "instantiated" by a radically heterogeneous group of physical microstructures, and so it follows that they can't be *identical* to any one of them, in accordance with Kripke's (2002, pp. 330-332) account of identity. Thus, what we have is a quick and painless rebuttal of the type-type identity theory of the mind, which *equates* mental states with brain states. For, if the psychological kind, M, and physical kind, P, were, in Putnam's words, *nomologically coextensive* (as quoted in Kim, 2002, p. 136, original italics)—that is, it is a matter of law that any system instantiates M at time t iff it instantiates P at t—then only those systems with the physical kind P could instantiate M. For example, if M were pain and P were C-fibres, then only creatures with C-fibres could feel pain and too bad for my injured Martian friend with his D-fibres firing; he doesn't have C-fibres so he can't be in pain! So, if we want to be able to attribute the same kinds of mental properties and events to creatures made from different physical structures, we have to abandon a type-type identity approach and find something better.

Kim's (2002) argument hinges upon three main premises. The first is what he calls the "Physical Realization Thesis" (p. 141), which actually consists of two ideas: one about events and the other about properties. In effect, it says that mental events only occur when an appropriately physical event is also present; in addition, it says that mental properties are a result of, and can be explained by, the lawful, causal connections between their physical realization bases. The second premise is the "Principle of Causal Individuation of Kinds" (p. 143), which says that sharing a property or being a nomic kind of event or object should be determined on the basis of particulars sharing similar causal powers. And, the third premise is "The Causal Inheritance Principle" (p. 143), which states that a supervening property, say the mental property M, occurring at time t, has as its causal powers all and only those causal powers present in its subvening realization base, say the physical base P, also occurring at t. This last is necessary for Kim because he claims that to deny it is to accept '*emergent* causal powers' that would violate 'the causal closure of the physical domain' (p.

143, original italics), something Kim claims all good physicalists should demand of their explanations. I will argue that a good physicalist can, and indeed *should*, accept emergent causal powers—in a way that *does not* violate the causal closure of the physical domain—and then extend this by analogy to the mental-physical debate.

But first we must continue to lay out Kim's (2002) position, which involves a discussion about jade. Jade is actually a kind composed of two chemically distinct substances, jadeite and nephrite, thanks to an historical confusion. Jadeite always has some green colour in it, whereas nephrite usually won't. The relevance of this is in a thought experiment that Kim poses: say that without knowing it you had only ever seen jade of the jadeite variety and you were considering the potential law, (L): All jade is green. You might think it's a law, but you would be wrong, since some jade (i.e. nephrite) is not green. We can't incrementally strengthen our belief in L on the basis of new observations—which Kim claims is supposed to be the 'hallmark of lawlikeness' (p. 140)—because instances of nephrite would soon disprove L and more instances of jadeite miss a key part of the category. The disjunctive nature of the kind, then, is responsible for our failure to accurately *project* L onto *jade*.[1] Kim sums it up nicely with the following passage:

> *...causal powers involve laws, and laws are regularities that are projectible. Thus, if pain (or jade) is not a kind over which inductive projections can be made, it cannot enter into laws, and therefore cannot qualify as a kind; and this disqualifies it as a scientific kind. (Kim, 2002, p. 144)*

And so we have our justification of premise two: only if a kind is composed of particulars sharing similar causal powers can lawful inductive projections legitimately be made over it. Jade, however, is composed of two heterogeneous sub-kinds because of their distinct chemical microstructures, which Kim claims must therefore be 'heterogeneous as causal powers' (p. 143). Thus, jade is a disjunctive kind over which lawful projec-

1. Ned Block (1997, p. 113) has a useful summary of Kim's (2002) concept of projectibility. If a theory, (T), entails that some data, (D), will be observed, then the observation of D confirms T. Supposing we introduce an arbitray claim, (A), then T OR A entails D, and thus the observation of D confirms not only T, but also A. But A was arbitrary, so we've just demonstrated that anything confirms anything. Therefore, we need to introduce the concept of being well-confirmed, such that the data, D, must confirm all of the generalization in order to well-confirm it. And projectibility depends on a generalization being well-confirmed. The observations of jadeite, while confirming the potential law, "All jade is green," don't well-confirm it because they don't confirm all of it (e.g. they don't confirm, "All nephrite is green"). Our observations must well-confirm a theory for it to be legitimately projectible to future observations.

tions cannot be made and so it cannot be a nomic/scientific kind.

The analogy with mental properties should by now be apparent: if pain is multiply realizable, each of the heterogeneous physical realization bases has distinct causal powers (e.g. human brains vs. octopus brains, microchips, etc.). Given premise three, pain then realizes a heterogeneous disjunction of causal powers, and from premise two we have that it therefore *must be* a heterogeneously disjunctive kind. Thus, pain-theories aren't projectible and so pain isn't a *nomic* kind. Kim concludes that, since identifying mental kinds with disjunctive physical kinds undoes their projectibility, we are left with a species or *structure-type restricted* type-type reductionism, and so psychology should be shattered into as many disciplines as there are nomic-kinds of physical realization bases. I intend, however, to dispute Kim's claim that the causal powers of different physical microstructures are *necessarily* heterogeneous and I will do so by arguing that their various *micro*-causal powers converge on the same *macro*-causal powers through the process of *emergence*.

I would like to briefly note what Ned Block (1997, p. 120) calls "The Disney Principle," which facetiously says that in worlds with scientific laws similar to ours—and in contrast to Disney movies—teacups can't think; that is, the laws of our reality impose certain constraints on what can realize thinking. Hooker (2004, p. 472) neatly captures Block's (1997) point, saying that a kind's various realization bases can't be so radically heterogeneous as to have no physical properties in common at all. Hooker (2004) considers this a step in the direction of his own analysis, and it would seem to be a worthwhile idea to keep in mind as we explore Rueger's (2000) and Hooker's (2004) theories—Hooker's being an expansion and refinement of Batterman's (2000) argument—which will constitute the bulk of the rest of this essay.

Hooker (2004) argues that an *emergent* system is one that exerts top-down constraints on the behaviour of its constituent elements. His discussion centers on a largely non-technical account of asymptotics, which is the study of mathematical domains as some parameter, **p**, approaches a critical point (in Hooker's case, zero). Rueger's (2000) discussion looks at asymptotics for some system as **p** approaches some *general* critical value—call it k—with k not necessarily equal to zero. As **p** moves into the asymptotic domain, new features may appear in the system (this is *singular asymptotics*), and specifically what we are interested in are new *emergent* features that form top-down constraints on the system's behaviour. (It is these features that I will be arguing constitute the convergent macro-causal powers of different physical microstructures.) So, say we have two theories, T_a and T_b, then what we're exploring is whether $[\mathrm{Lim}_{p\to k}(T_a)] = T_b$. Two examples of inter-theory characterizations that many people may be passingly familiar with are quantum theory as Planck's constant approaches zero and special relativity as 1/c approaches zero both yielding Newtonian mechanics (quantum theory roughly so). Hooker (2004) characterizes three domains: the non-asymptotic (**p** far from k), the asymptotic (**p** near k), and the limit (**p** = k). The interesting ones for our consideration are the asymptotic domain and the limit, and here Hooker makes use of an enlightening example involving the phase transition of iron from molten to solid form.

Essentially, the phase transition of iron occurs in a temperature dependent system where the iron molecules, when entering the asymptotic domain from a molten temperature, form crystalline-like cluster fluctuations. These intermediate crystalline macrostructures, while existent, have the causal power to top-down constrain the microstructural behaviour of their constituent iron molecules. Furthermore, at the limit **p** = k, all of the molecules stabilise on a collective crystalline macrostructure and the phase transition is complete. The result is an iron bar that, in philosophy, is said to supervene on its molecules, and yet in a dynamical-analysis sense also strongly top-down constrains the behaviour of the physico-chemical microstructure "...through the formation of a new macro-scale force *constituted in the chemical bonds formed*," (Hooker, 2004, p. 455, italics mine). It also changes the 'force form' of the dynamical equations governing the microstructural behaviour and is able to retain its relationship of top-down constraint under external perturbations via energy dissipation. Hooker says that "...the dynamical cohesion created by such constraints is the ultimate foundation of all physical system identity... for it *determines the substantive system boundary*, and with properties also dynamically individuated... the character of the new individual is constituted by its capacity to do new work" (p. 461, italics mine).

Rueger (2000) describes a slightly different idea. He explains that *regular* asymptotic limits of T_a 'go smoothly' over into T_b, and he gives special relativity transitioning to Newtonian mechanics as an example of this. *Singular* asymptotics—the primary focus of Batterman's (2000) essay and also a central focus of Hooker (2004)—according to Rueger (2000), however, are discontinuous in their transition to the limit. Rueger starts at the limit, k, and considers small variations in **p** around this value. If the equations of the perturbed system can still be topologically mapped back to the limit equation then they are 'structurally stable' (p. 473); that is, the two behaviours are essentially similar if the systems are topologically equivalent under small changes in **p**. However, systems with *singular asymptotes* are topologically inequivalent under perturbations of **p** around k, resulting in qualitatively different behaviours for the system. The molten-to-solid phase transition of iron is just such a case: as we move through the asymptotic domain to the limit, the solid-iron system displays characteristically novel behaviours from the molten iron that preceded it.

The significance of this becomes clear as we note Rueger's (2000) look at *families* of systems that all have similar instabilities under similar perturbations of their parameters (p. 478); that is, the systems display structurally stable individual instabilities as *families of systems*. So, say we have another metal undergoing a molten-to-solid phase transition. If it displays the same form of behavioural change (that is, it transforms to a new topology equivalent to iron's new topology), then we have just found what Batterman (2000) calls a *universality class* that shares some behaviour/property across the physico-chemically distinct microstructures (p. 123). As a result, perturbations in the subvenient microstructures are filtered out by the new top-down constraints and become largely irrelevant to the systems' behaviours; Batterman notes that it is the *collective* properties of the microstructural constituents that dominate the systems' behaviours. One last point is that each member of the universality class may have a different critical point, k, but still behave similarly (i.e. still be topologically equivalent).

There are two important questions yet to answer: (1) why can't we just describe this 'emergent' behaviour near the limit in terms of the *microstructure's* behaviour—that is, in terms of T_a—without resorting to talking about T_b directly; and (2) what does this have to do with mental states reducing to physical states? As regards (1), let's first be clear that the macroscopic property/state is *dynamically determined* by the microstructure it is composed of. However, Hooker (2004) contrasts this with *logical* determination, in which the macroscopic state/property is logically derivable from the states/properties of its constituents. He distinguishes between two types of reduction: reduction$_2$, which scientists speak of when they say T_a reduces$_2$ to T_b when $[\mathrm{Lim}_{p\text{->}k}(T_a)] = T_b$, and reduction$_1$, which philosophers use when they say that T_b (an ontologically coarse theory such as Newtonian physics) has been reduced$_1$ to T_a (an ontologically broader theory). In this case, Hooker notes that the 'formal mathematical transform schema' (p. 459) always holds, whether with regular or singular asymptotics. However, he claims that there is a failure of type 1 reduction, since we fail to achieve a dynamical basis for asserting *ontological* identity when using $[\mathrm{Lim}_{p\text{->}k}(T_a)] = T_b$ as a schema in the philosophical sense. This is because "...the iron bar is a new macro-scale level with respect to its molecular constituents *because it has its own characteristic dynamical interaction form*," (p. 462, italics mine); the top-down constraints eliminate micro-structural freedoms and render a

collective macroscopic behaviour that is insensitive to many of the micro-constituents' individual states/properties. Thus we may legitimately speak, ontologically, of macroscopic objects, properties, and events when Hooker's dynamical criteria of emergence have been met.

The answer above also addresses question (2): we may *legitimately speak* of ontologically genuine, multiply realized mental states insofar as they are *emergent* elements of a dynamical singular asymptotics. As Batterman (2000) notes in passing, and as also claimed in J. Satinover's book (2001), there is a certain kind of network model with biological plausibility (Hopfield networks, though Batterman doesn't note their name) that is formally similar to the models of lattice systems and spin glasses. These models are like those used to represent fluid microstructures and Batterman (2000, p. 133) states that the mathematics of such systems involves singular asymptotic analysis of the sort we have just been discussing. Thus, we also have a plausible biological basis for the forms of intelligence *that we know of*—given that these singular asymptotics generate emergent new levels of top-down constraints on their realization bases—and therefore, a possible formal dynamical representation of the mental for many, maybe even all, forms of intelligence. In addition, Hooker (2004) notes that such emergent systems can form the constituents for *even larger* systems, thus creating "...complex histories of emergent phenomena marked by strong dynamical fixation of historical constraints where dynamical form may change as a system—including even just its initial conditions—changes," (p. 476). It is with this in mind that Hooker cautions against hoping for general laws to permeate the special sciences; such laws may be *incredibly* computationally inaccessible to us, taking more time to discover than we have time in the universe.

Finally, then, does the multiple realizability of mental states preclude their reduction to physical states? In light of Block's (1997) Disney Principle, we should expect that other forms of intelligence at least share something in common with us. And, as we have seen, what we have is an account of genuinely emergent macro-causal powers that should be acceptable to the physicalist, as there is no violation of the causal closure of the physical domain since they are constituted, dynamically, in their micro-constituents. In one respect, this account in fact *demands* type 2 reduction of the mental to the physical, but only insofar as it asserts a mathematically formal schema for the emergent universality classes of physical realization bases for the various mental states. This variety of reduction, in essence, asserts the possibility of 'in principle' discoverable mathematical descriptions of all emergent states/properties, from iron crystals to humans and their societies, all of which are dynamically determined by their constituent elements. Hooker (2004) notes, however, that all of these systems are strongly historically dependent and so we may have difficulty discovering the laws governing them. From the ontological perspective, emergence of the mental from the physical denies type 1 reduction as a result of emergent top-down constraints (i.e. macro-causal powers) with legitimate claims to ontological validity and logical irreducibility to the ontologies of their microstructures. And, given that such genuinely emergent macro-causal powers converge on topologically equivalent domains, they can even be considered legitimate nomic kinds. Thus, Kim (2002) was mistaken to claim a shattering of the special sciences, for while the situation is more involved than we had previously anticipated, there is indeed a plausible justification for the multiple realizability of mental states, with their various physical bases converging on the same causal powers. At last, then, my Martian friend and I can *both feel pain* again.

references

Batterman, R. W. (2000). Multiple Realisability and Universality. *The British Journal for the Philosophy of Science, 51*, 115-145.

Block, N. (1997). Anti-Reductionism Slaps Back. *Noûs: Supplement: Philosophical Perspectives, 11, Mind, Causation, and World, 31*, 107-132.

Hooker, C. A. (2004). Asymptotics, Reduction and Emergence. *The British Journal for the Philosophy of Science, 55*, 435-479.

Kim, J. (2002). Multiple Realization and the Metaphysics of Reduction. In D. J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 135–149). New York: Oxford University Press.

Kripke, S. (2002) Naming and Necessity. In D. J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 329-334). New York: Oxford University Press.

Rueger, A. (2000). Robust Supervenience and Emergence. *Philosophy of Science, 67*, 466-489.

Satinover, J. (2001). *The Quantum Brain.* New York: John Wiley & Sons.