OUP UNCORRECTED PROOF – FIRSTPROOFS, Tue Dec 18 2018, NEWGEN

9

Building Inner Tools

ROBERT L. GOLDSTONE

1. Building Inner Tools

Humans show a striking penchant for creating tools to benefit our own thought processes. Andy Clark (2003, 2008) has convincingly argued that the tools that we as humans recruit become integrated parts of an extended cognitive system that includes us as just one component. By extending cognition beyond our brains, Clark presents an "embiggened" perspective on what it means to be a cognizer and a person more generally. This perspectival shift runs counter to some recent forms of argumentation that in effect work to minimize personhood. For example, arguments for lack of personal culpability can take the form of "It wasn't *my* fault. It was the fault of my ______" to be filled in, perhaps, by "upbringing," "genes," "neurochemistry," "diet," or "improperly functioning amygdala." Instead, Clark (see also Dennett 1989) offers the opposite line of argumentation, according to which we consist not only of our amygdalae and hippocampi but also potentially our glasses, notebooks, friends, supporting technologies, and culture.

I find this vision of human nature to be empowering and transformative. At the same time, I find a critical plank in this argument to support a very different launch trajectory which I would like to pursue here. This plank is the stance that one should remain neutral about how functionally defined systems can be implemented by mixtures of biological and nonbiological components. The specific version of this stance that Clark and Chalmers (1998, 8) adopt is known as the Parity Principle: "If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process." The motivation for this principle is to avoid the biochauvinistic prejudice of accepting something as implementing cognition only when it is accomplished by a brain. Such prejudice begs the question of what kinds of things are capable of cognition by defining out of existence, for example, human-machine partnerships. By adopting the Parity Principle, we keep our mind open about how open are

Colombo300818ATUS_MU.indd 113 18-Dec-18 10:30:08 PM

minds—about how cognitive systems might be implemented. Clark (2005, 2n3) has likened the Parity Principle to a "veil of metabolic ignorance," analogous to Rawls's (2001) "veil of ignorance," to establish an impartial position from where we can determine governing principles rather than simply assert self-serving biases. Moreover, one also can apply one's intuitions and conceptual frameworks for understanding other people's minds to understanding the cognitive potential of other systems, such as technology-enabled humans or groups of people (Theiner, Allen, and Goldstone 2010).

Once one adopts the neutral stance of trying not to prematurely rule out instantiations of functional systems, a perspective approaching the reverse of Clark's "supersizing" naturally arises. Just as we can recruit our understanding of human minds to help us make sense of cyborg, robot, and group cognition, so we might be able to use our understanding of external systems to make sense of human cognition. To this end, it is possible to propose a:

> Reverse Parity Principle (RPP): a brain component should be considered to be part of a *distributed* cognitive system if we would accept it as being part of a distributed system if it were nonbiological.

The parity and reverse parity principles are flip sides of the same boundary-erasing coin. They share the position that what matters is the nature of a system understood as parts and their interactions. What does not matter is to precisely ascertain what is part of the central processor of the system and what is external to that central core. If a system has been identified that has integrity, then it is the entire system that must be analyzed as a whole to understand its function. For example, if a set of components forms a homeostatic negative feedback loop, and this loop is core to the functionality of the system, then all of the components that compose the loop are equally function-providing contributors, regardless of where they fall on a biological/nonbiological divide.

One motivation for the RPP is that we often know more about the external transduction of information in a system compared to information processing within a brain. Modern neuroimaging techniques have greatly improved our understanding of within-brain information flow, but information flow external to the brain is still typically easier to analyze because it does not require expensive laboratory equipment and often leaves a paper trail, lasting inscriptions, or digital traces. The primary motivation, however, is to invite a perspectival shift that is potentially as illuminating as the extended mind shift. Whereas extended mind arguments invite us to think of ourselves as broader, wider, and larger than we might otherwise, the RPP invites us to consider the strangeness and otherness that lie within what would normally be considered to be "us."

In developing implications of the RPP, my approach will be as follows. First, I will argue that cognitive systems can be identified that span multiple levels, from within-individual modules to individuals to groups of individuals. Four examples of these cognitive systems will be presented, with the aim of showing that postulating them can be inductively powerful despite their cutting across traditional scientific disciplines. Having laid out the case for the same kind of cognitive system being implemented with different kinds of components, the second step is to consider particular organizational principles that are prominent when considering distributed systems and can productively be applied within a single human brain. These principles are specialization, tool creation, and indirect levers. When applied to single brains, these principles suggest that our brains are made up of modularized tools that are adaptively tuned to our needs even though their inner workings can only be imperfectly and indirectly shaped. A third and final section describes specific ways in which we modify our own internal modules, via selective reinforcement of quasi-random variation, strategic training, and emulation. These mind hacks can lead to eventually highly refined cognitive systems that are aptly understood as created mental tools even though they reside inside one's head.

2. Inductively Powerful Systems

Both forward and reverse parity principles stem from the notion that there should not be one privileged level of analysis when thinking about cognitive systems. Systems with integrity exist that are wider or narrower than a single central nervous system. What it means for a system to have integrity is that its components are working together as a unified whole to achieve its relevant functionality. Systems arise at multiple levels for multiple purposes, which is to say that whether a particular set of components forms an integral system is typically a matter of perspective. To a neuroscience expert, the hippocampus, or even the dentate gyrus subcomponent within the hippocampus, forms an important system, but for many other purposes, it is just one component of a large system for encoding and retrieving memories, and for still other purposes, such as describing Otto's memory (Clark and Chalmers 1998), it may be, by percentage, an even smaller component of a system that includes notebooks, iPhones, and reminder ribbons tied around fingers.

This is not to say that all perspectives are equally useful. The utility of a perspective can be assessed by the inductive power of the categories that it brings to mind. A category is inductively powerful if valid, task-relevant predictions can be ascribed to something, given that it has been placed in the category (Anderson 1991). The category **llama** is inductively powerful because once it has been ascertained that something fits in it, one can make many potentially useful inferences about its size, number of legs, ability to carry heavy loads up mountains, and proclivity to spit. Common taxonomic categories like **llama**, **sofa**, **chemistry**, and **charcoal** are typically useful categories to ascribe because the objects contained within them tend to share many life-relevant properties in common (Rosch and Mervis 1975). Other ad hoc categories such as **things to remove from a burning house, things that could be stood on to reach a lightbulb,** and **things that float** do not share many properties and are generated only in a time of momentary need (Barsalou 1983). An ability to think about ad hoc categories may save a passenger on a sinking ship but are not generally inductively powerful. Their lack of lasting inductive power is even suggested by their categories' names being phrases rather than single words. Given the propensity for routine language to become progressively more efficient, if these categories had been generically useful, then they probably would have eventually become tokened by single short words.

Categories related to systems occupy an interesting intermediary case between taxonomic and ad hoc categories. Like ad hoc categories, they bring together entities that at first sight seem to have very little in common, but like taxonomic categories, they can also be the source of rich inductive inferences. One of the most simple and common systems concepts is a *positive feedback system* characterized by a system in which increases to something cause still further increases to it. Examples of this category include a microphone being placed too close to an attached amplified speaker, global warming (ice melting into water causes more sunlight to be absorbed into the earth, which causes more ice to melt), citations to an article leading to more citations, and children pestering their parents to buy them a doll that other children in their school already own (Goldstone and Wilensky 2008).

Once one knows that a system falls in the category of a positive feedback system many inferences are possible, such as the likelihood of a runaway growth process, a possible tipping point, and characteristic growth dynamics. To be sure, llamas are more like each other than positive feedback systems are like each other (though see Goodman [1972] for inherent difficulties in cashing out that claim). Still, the observer armed with the concept of a positive feedback system may be able to anticipate the future behavior of a system and nip vexing growth dynamics in the bud, which can make for a powerful induction indeed. In fact, one could well argue that education ought to be primarily in the business of teaching students system concepts like this—other examples being **diffusion**, **autocatalysis, negative feedback loop**, **lateral inhibition**, **annealing**, and **resonance**—precisely because of their cross-discipline inductive power. People might be presumed to invent **cat** and **table** concepts on their own from unsupervised exploration of our modern world, but might miss out on the common **diffusion limited aggregation** system that underlies lungs, cities, and electrical discharges unless they are explicitly given instruction concerning this system's growth dynamics (see also Goldstone and Theiner 2017).

The reason for stressing the inductive legitimacy of systems concepts is to support claims that an important cognitive system might be alternatively implemented within a single brain, across several brains, or in a brain-environment cyborg assemblage. Arguing against this possibility, Adams and Aizawa (2001, 61), reflecting on the myriad forms that memory-augmenting technologies take, ask, "What are the chances of there being interesting regularities that cover humans interacting with all these sorts of things? Slim to none, we speculate." By contrast, Clark and Chalmers (1998, 14) judge that "by using the 'belief' notion in a wider way, it picks out something more akin to a natural kind. The notion becomes deeper and more unified, and is more useful in explanation." While preferring the notion of **inductive power** to **natural kind** because inductive power does not carry unnecessary baggage associated with identifying kinds of things that are intrinsically "out there," I fundamentally agree with Clark and Chalmers that systems concepts are frequently useful, even though, indeed often *because*, they are applicable to many situations that seem to be dissimilar. In fact, a productive general strategy for figuring out what is core to a cognitive system is to purposefully consider as widely ranging cases of it as possible. In the same spirit that a linguist interested in what is common to all human languages would be better served by studying Swahili, Dutch, and Chinese rather than French, Italian, and Spanish, the best cure for overly parochial interpretations of concepts like **beliefs**, **memories**, and **concepts** is to study them in all of their diverse forms.

A skeptic might respond that some systems concepts in science, such as diffusion, positive feedback, and autocatalysis, do have inductive power that transcends traditional disciplines, but for *cognitive* systems in particular, the standard unit of one central nervous system is the only instantiation. This response seems unlikely given the general requirements faced by any system that has an information-based niche and the specific successes of functional accounts of adaptive behavior that apply to both artificial and natural intelligences. For now, I will simply describe four candidate cognitive systems that have been usefully employed to understand systems wider than, narrower than, and simply different from single central nervous systems.

1. Competitive specialization. A group of originally homogeneous, undifferentiated units are sequentially presented with resources (see figure 9.1). The unit which is closest or more similar to the resource moves or adapts toward the resource rapidly, while the other units adapt more slowly. The end result is a spontaneous, self-organized division of labor among the units in which different units are specialized for different resources, and resources that are close or similar will tend to be handled by the same unit. This learning algorithm has applications for allocating members of a team to spatially distributed resources, unsupervised pattern learning in neuroscience, and autonomous machine learning (Rumelhart and Zipser 1985).
2. Diffusion-based decision-making. Information continuously accumulates and a response is produced when a threshold amount of evidence has been achieved for that response. The diffusion model can be shown to be optimal in that it achieves the fastest mean decision time for a given accuracy (Bogacz et al. 2006) and has been recruited to understand both human perceptual judgments (Forstmann,

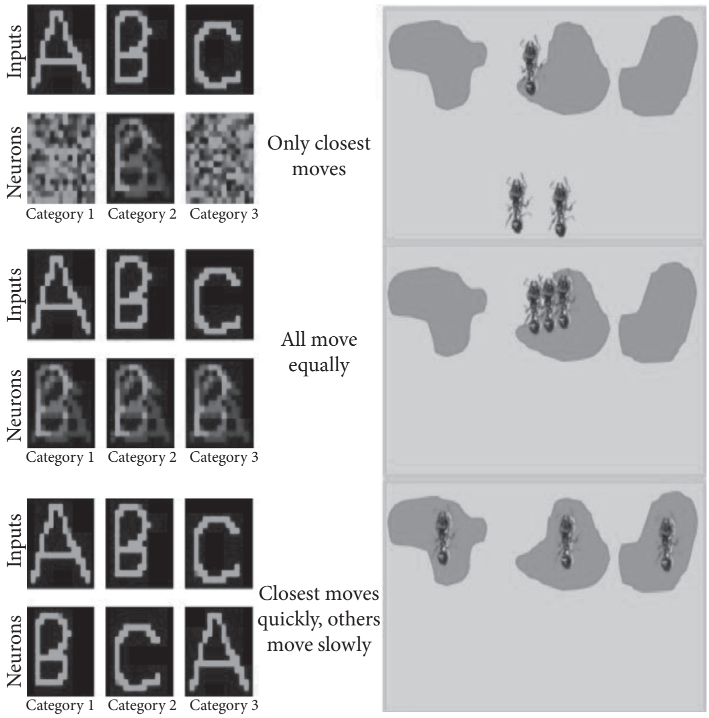

Figure 9.1 Two examples of competitive specialization, adapted from Goldstone and Sakamoto (2003). For the case on the left, three neurons are adapting to cover three input patterns that are randomly and repeatedly selected. For the case on the right, three ants are moving to cover three resource patches. When only the closest agent moves toward a resource (top panels), the agent will inadequately cover all three resources, while other agents do not adapt at all. If all agents move equally quickly (middle panels), they will end up being equally influenced by all of the resources and will come up with the same inadequate solution. If the closest agent to a selected resource adapts quickly while the other two adapt slowly (bottom panels), then they will spontaneously self-organize so that each agent is specialized for one of the inputs, efficiently covering the entire space of resources.

Ratcliff, and Wagenmakers 2016) and how ant colonies and beehives choose new nesting sites (Marshall et al. 2011).

3. Temporal difference learning. Learning can be based not only on a rare external reward at, say, Time_{100}, but also on the degree to which a prediction at Time_{10} about that later reward matches a prediction at Time_{20}. When predictions disagree, then later predictions can critique and modify the earlier predictions in a bootstrapping process. This self-critique is supplied by the neurotransmitter dopamine in nervous systems, which provides an internal reward signal that can shape animals' actions without requiring an external reward (Schultz, Dayan,

and Montague 1997). In artificial intelligent systems, this form of learning can dramatically increase the efficiency of learning in systems for playing games, maze solving, and dynamic control tasks (Sutton and Barto 1998).

4. Decision-making by similarity-based sampling. One way to make generally appropriate decisions in a given situation is to sample previously remembered situations, determine what was or would have been the correct decision in those situations, and make the corresponding decision in the given situation, with "recommended" decisions weighted by the similarity of the remembered and current situation. This notion of making decisions by sampling memory serves as the basis for financial decisions (Stewart, Chater, and Brown 2006), human categorization judgments (Nosofsky 1984), and machine learning models of induction such as Support Vector Machines and Bayesian inference (Fung and Mangasarian 2001).

While chosen semi-arbitrarily, these examples suffice to show that very different components can implement the same cognitive function. While none of these identified systems implements a complete mind, they all represent the kind of algorithms that cognitive scientists spend most of their time thinking about and for which tangible gains in understanding have been made. They represent the kinds of systems that are likely to underpin cognition in us and other agents. For this reason, I remain optimistic that inductively powerful cognitive systems can be identified that apply across many different instantiating components.

3. Piecing Together People

Thus far, only a generic defense of the RPP has been ventured. Like the Parity Principle, one of the main benefits of the RPP is to help us see familiar phenomena in new ways. What are the specific notions that we can usefully borrow from clearly distributed systems to understanding how our own brain function works? I will focus on three.

3.1 Specialization

Divide-and-conquer is a valuable strategy for information-processing systems. We understand well that a good team will divide its task into well-differentiated parts and that success depends on division of labor. This division of labor clearly happens within a brain too and creates, to some degree, fragmented rather than unified minds. Division of labor means that no part will have everything in mind.

It might seem impossible or exceptional for a system to become more richly internally structured over time, but in fact it is rather common as the competitive specialization example above suggests. Environments tend to be "clumpy," with objects distributed in clusters rather than strewn randomly or homogeneously. This is true for spatial clustering, and is even more true for objects' nonspatial dimensions. Dimensions tend to be strongly correlated. Thanks to the **bird** clump of the world, nesting in trees, singing, flying, and laying eggs are all correlated with each other. As long as an environment is naturally "clumpy," a distributed system of agents can adaptively divide this world into its clumps and develop agents specialized for different clumps.

A striking example of this acquired specialization within human brains is the acquired functional specialization of relatively small parts of the brain for recognizing words and mathematics. The visual word form area in the left fusiform gyrus, for instance, is implicated in reading (Dehaene and Cohen 2011). Visual number forms are processed in the inferior temporal gyrus and anterior to the temporo-occipital incisure (Shum et al. 2013), and the intraparietal sulcus and prefrontal cortex have been identified for their contributions to mathematical cognition (Amalric and Dehaene 2016). Written language and mathematical reasoning have probably existed for less than six thousand years, with examples dating back to the Sumerians in Mesopotamia. These inventions are decidedly recent with respect to biological evolutionary time. Brain specializations for words and numbers are possible because (1) people solve new cognitive tasks by reusing brain regions that evolved for other purposes (Anderson 2015; (2) the brain is sufficiently malleable that its parts become tuned, over development, to new requirements; and (3) cultural artifacts have adapted over relatively brief recorded history to improve the efficiency with which they are processed by relatively slowly evolving brains (Changizi and Shimojo 2005). Combining the latter two points: we adapt to culturally relevant objects while at the same time they are adapting to us (see also Clark 2003).

3.2 Tool Creation

It is clear that people create external, nonbiological tools to help them cope with their world. Clark (2003, 2008) has convincingly argued that humans are remarkable in their inclination to create tools, create tools for creating tools, and so on. This perspective on people as consummate tool builders applies to internal tools as well—that is, to brain changes that allow us to do our cognitive tasks better.

The phenomenon of categorical perception (CP) is a good example of how perception comes to better support cognition by leading us to perceive our world in terms of the categories we have formed (Goldstone and Hendrickson 2010). By CP, our perceptions are warped such that differences between objects that belong in different categories are accentuated, and differences between objects that fall into the same category are de-emphasized (Harnad 1997). CP transforms relatively linear sensory signals into relatively nonlinear internal representations. This transformation is important because it promotes the crucial cognitive function of treating distinguishable stimuli as the same thing. Once different examples of a phoneme /d/

, different cats, or different chairs are treated as the same kind of thing, then irrelevant variations are de-emphasized and connections can be made between things that have disparate superficial appearances. While we might have expected these connections to be made only at deeper, cognitive levels, turning over some of the work in creating equivalence classes to perceptual systems frees up executive control functioning for other tasks and leads to the fast and efficient detection of categories. A general strategy that the brain uses to become more efficient is to shift, using a well-understood basal ganglia circuit, tasks that once required executive function, frontal neural circuits to more posterior circuits that subserve habits (Poldrack et al. 2005). The brain performs its own "internal outsourcing," turning strategic action-outcome reasoning into automatic stimulus-response behavior.

There are other striking examples of visual processes being tailored to an organism's tasks. In a case considered by Clark (2008), Milner and Goodale (1995) have proposed interacting but distinct visual pathways involved in visual identification/recognition of an object and reaching for that object. One confirmed prediction of the idea that perceptual processes are tuned to the currently relevant actions is that visual systems engaged in identification/recognition versus reaching should not always show the same pattern of sensitivity to illusions (Bruno and Franz 2009). While verbally reported size judgments for a central circle are heavily influenced by the sizes of surrounding circles, grasps for the central circle are relatively unaffected. This is a plausible pattern if one assumes that successful grasps depend on metric calculations based on the target itself, whereas explicit perceptual judgments about that target can sometimes benefit from determining its relation to other objects.

In these examples, and many others, it is compelling to think of our own brain's perception and action modules as tools to be shaped to our needs. They are particularly important tools because of their immediacy, indispensability, and gatekeeper status—all other tools that we want to use have to go through these. An important apparent difference between these modules and external tools like hammers, beds, computers, and Google Search is that the internal modules seem not to be intentionally crafted, but shaped only by automatic, unwilled processes. I will return later to this apparent difference when considering the ways we hack our own brains.

3.3 Indirect Levers

In the massively distributed system known as economics, a market-based instrument encourages behavior through market signals rather than through explicit directives (Stavins 2003). Market-based instruments are frequently contrasted with "command and control" regulations that simply specify a target to be achieved, frequently detailing the technologies and method that must be used to reach the target. Examples of market-based instruments are 10 cent deposits on aluminum cans, cap-and-trade programs to control sulfur dioxide emissions, and rebates for water-efficient washing machines.

The insight from these indirect levers is that you can change a system's behavior even if you cannot or do not want to directly rig it up yourself. This insight may seem inapplicable to brains and minds. One might have thought that our own mind is the easiest thing to change because each of us is in extremely close contact with it, some going as far as to say that it *is* us. Furthermore, sometimes we can change our minds easily. If somebody trustworthy tells us "A baby llama is called a cria," we can absorb this fact with only one telling and potentially change our mind permanently.

Very often, though, changing a mind is frustratingly difficult, and is not made much easier just because one happens to be the mind's owner. This irreducible fact is rather inexplicable from the viewpoint that we are each a unified self with access to the mental components that confer intelligence on us. The fact, however, will be readily attested to by anyone who has tried to quit a bad habit, learn linear algebra, or play guitar well. We can make headway on these goals, but it is usually a long slog, and we often make more progress by applying mental market-based instruments rather than command-control regulations.

4. Hacking One's Own Mind

The suggested perspectival shift of RPP is for people to see their own internal mental operations as the result of the interactions among a distributed network of partially decomposable modules that are specialized and adaptive but not directly controlled by us. In fact, it is not even clear what remains of a central "us" that could possibly be doing the controlling. In the previous section, the last two perspectival borrowings from distributed systems, "tool creation" and "indirect levers," suggest mental modules that are liminally positioned in terms of their opacity to strategic manipulation. A tool that cannot be crafted according to the wielder's need is not much of a tool, but perception-action modules are classically considered to be black boxes incapable of being penetrated by cognitive entities such as goals, beliefs, and desires (Fodor 1983).

In fact, modules may be gray or murky boxes, but they are rarely completely black boxes. People devote considerable time and energy to hacking their own perception-action modules. Consider, for example, the ingenious efforts of athletes and musicians to create novel training methods for improving their own performance. Soccer players have been known to train themselves in a multiple-object tracking task in order to improve their global situation awareness on the field (Faubert and Sidebottom 2012). Musicians give themselves ear training exercises that specifically allow them to better discriminate troublesome intervals. Olympic kayakers monitor the force of individual strokes with sensor arrays to provide instantaneous feedback on the effects of slight modifications of their paddling technique.

Short of auto-neurosurgery, people do not have the ability to completely re-wire their perceptual systems to give themselves new perceptual capabilities, such as seeing infrared light or hearing a 30 kHz pitch. However, we routinely and strategically modify human perceptual systems by giving ourselves and our students targeted training. Very different training is required of music students to master discriminations between absolute pitches (e.g., A vs. A#) versus relative intervals (e.g., minor vs. major thirds) (Hannon and Trainor 2007), and students regularly avail themselves of training methods suited to their musical goals. As a result of musical training, musicians relative to nonmusicians acquire different responses to both music and speech in their brain stems. Musicians also show different patterns of sound emissions from their ears that is consistent with the theory that they have improved auditory processing that stems from top-down feedback from the brainstem to the most peripheral site of auditory processing: the cochlea (Moreno and Bidelman 2014). Dr. Susan Barry (2010) lacked binocular stereoscopic depth perception but was able to strategically train herself to have this ability by presenting to herself colored beads at varying distances and forcing her eyes to jointly fixate on them. It is worth noting for this last example that binocular depth perception is one of the human perceptual abilities with the strongest empirical claims for having status as a neurophysiologically and functionally genuine module (Nakayama 2005).

These examples show that training leads to changes in impressively peripheral sites of perceptual modules. People use various strategies to hack their own perceptual modules to make them better tools. Three of these strategies follow.

4.1 Selective Reinforcement of Quasi-Random Variations

Even if an agent does not have any access to the internal machinations of one of his or her neural modules, as long as that module produces natural or coaxed randomness, the blind flailings of the module can be selectively reinforced to reward favorable random changes (Goldstone, Landy, and Brunel 2011). Selective reinforcement is at the heart of the behavioral shaping techniques of the animal learning theorists. This method is not terribly fast, but with patience and repetition it can still produce remarkable results, such as allowing pigeons to correctly categorize mammograms of benign versus malignant breast tumors 85 percent of the time when tested individually and 99 percent when aggregated in a flock-sourced approach (Levenson et al. 2015).

4.2 Strategic Training

Individuals who are either introspective or well-read in the psychology of learning frequently expedite changes to their perception-action modules by strategically modifying their training. Examples of these strategies include giving oneself spaced rather than massed practice when learning anatomical structures in medical school, practicing a guitar song at quarter speed or a piano song separately with left and right hands, purposefully exposing oneself to different speakers when trying to learn a difficult speech sound discrimination in Mandarin, viewing deliberate caricatures of two highly confusable mushroom species before examining actual samples so as to orient oneself to their diagnostic features, and parents placing paintings on the walls of a baby's room when they want the baby to later have an easier time identifying the paintings. These strategies are powerful because they can be applied across many domains and they are still more directed than simply waiting for random flailing to produce a behavior worth reinforcing.

4.3 Emulation

An often efficient way to learn a new skill is to emulate an expert. Students of drumming will frequently emulate their teacher, either in synchronized duet playing or in call-and-response fashion. Teachers in many disciplines often find it more instructive to show than tell. Words are not completely useless, but teachers of dance, skiing, and even computer programming frequently have the experience that words are too narrow a bottleneck for conveying to students their nuanced message. Instead, teachable moments can reliably occur if students are made to move their bodies in the correct way and then reflect on what that feels like.

These three strategies are ordered in terms of their increasing directedness, and emulation is typically possible only after a long period of random flailing and reinforcement has transpired because emulation requires learners to already have significant control over their modules. Perhaps the most potent reward for achieving expertise in a domain is to be able to express oneself in a more nuanced, articulate manner within that domain. The reward for playing a game well is to get the chance to play more of the game better. Experts can perceive and produce nuances and complexities that they completely missed as novices. Originally coarse and poorly controlled hacks give way to more pointed and precise hacks. The hacking of one's own brain is the epitome of pulling oneself up by one's bootstraps. The progression of skill from infant to expert consists in the iterative creation of increasingly refined levers, channels, and expressive vocabularies that depend upon and hone the previous iteration. Very akin to how Clark (2003) has argued for the striking advances that humanity achieves by building tools for building tools that build tools, the eventual, highly refined organization of our perception-action brain modules would be impossible without hundreds of iterations of using existing neural organization to establish the next, incrementally refined organization.

Hacking one's own mind is, then, a potent case of tool creation even though and because it occurs inside one's head. Although this hacking is initially coarse and is never perfectly targeted, it is nonetheless a case of the willful, strategic training of perception to improve its operation for highly specific purposes. For example, as Barry (2010) trained herself to have depth perception, she *wanted* to have depth perception, she *believed* that her training regime would serve to give her this ability, and the training itself was systematically related to her developing the ability. If this does not count as a rational process, then many canonical cases of human rational inference will probably fail to count as well.

5. Conclusions

The RPP invites consideration of the ways in which the components of our brain can be understood as akin to components of an external distributed system. This perspective helps us understand why changing our minds is frequently a long, arduous, and incremental process. The visual word form area in the left fusiform gyrus is, by this account, a tool par excellence for helping us decipher written words. The inevitable answer to the question "But is it a tool that we use or is it simply part of us?" is simply "That is a matter of perspective." As described in the first section, there are multiple scales of systems that can be postulated, some which do not include the visual word form area as part of a cognitive system but rather treat it as an external tool, some which do treat it as part of a single system, and some which include it as well as notebooks, iPhones, and colleagues. The choice between these systems/perspectives should be based on the inductive power of the perspective, which will, in turn, be influenced by the kinds of properties one is interested in inferring.

The core conceptual advantage of both the forward and reverse parity principles is to dissuade us from believing that there is a single privileged level for understanding cognition—one that includes the brain, the whole brain, and nothing but the brain. The parity principle has the empowering effect of literally enlarging ourselves to include some of the reliable and trustworthy environmental entities that make our thoughts possible. The path to human betterment by the extended mind account lies in developing technological and infrastructural devices that expand our capabilities. Does the reverse parity principle therefore diminish us? This is certainly a possibility, but there is also a kind of fascination that comes with appreciating the otherness that lies within us. It also suggests an alternative (but not incompatible) path to human betterment by using science and technology to improve the training of our internal brain components. At a first pass, we depend on components for perceiving and acting on the world that we can control only imperfectly and after considerable experience. At a second pass, there is very possibly no central "we" that depends on peripheral components. There are only interacting components that imperfectly control one another—tools that create and shape other tools without needing a separate, original user at all.

References

Adams, F., and K. Aizawa. 2001. "The Bounds of Cognition." *Philosophical Psychology 14*(1): 43–64.

Amalric, M., and S. Dehaene. 2016. "Origins of the Brain Networks for Advanced Mathematics in Expert Mathematicians." *Proceedings of the National Academy of Sciences 113*: 4909–4917.

Anderson, J. R. 1991. "The Adaptive Nature of Human Categorization." *Psychological Review 98*: 409–429.

Anderson, M. L. 2015. *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.

Barry, S. R. 2010. *Fixing My Gaze: A Scientist's Journey into Seeing in Three Dimensions*. New York: Basic Books.

Barsalou, L. W. 1983. "Ad Hoc Categories." *Memory and Cognition 11*: 211–227.

Bogacz, R., E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen. 2006. "The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced Choice Tasks." *Psychological Review 113*: 700–765.

Bruno, N., and V. H. Franz. 2009. "When Is Grasping Affected by the Müller-Lyer Illusion? A Quantitative Review." *Neuropsychologia 47*: 1421–1433.

Changizi, M. A., and S. Shimojo. 2005. "Character Complexity and Redundancy in Writing Systems over Human History." *Proceedings of the Royal Society London B 272*: 267–275.

Clark, A. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. New York: Oxford University Press.

Clark, A. 2005. "Intrinsic content, active memory and the extended mind." *Analysis 65*(285): 1–11.

Clark, A. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.

Clark, A., and D. Chalmers. 1998. "The Extended Mind." *Analysis 58*: 7–19.

Dennett, D. 1989. "The Origins of Selves." *Cogito 3*: 163–173.

Dehaene, S., and L. Cohen. 2011. "The Unique Role of the Visual Word Form Area in Reading." *Trends in Cognitive Sciences 15*(6): 254–262.

Faubert, J., and L. Sidebottom. 2012. "Perceptual-Cognitive Training of Athletes." *Journal of Clinical Sports Psychology 6*: 85–102.

Fodor, J. A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Forstmann, B. U., R. Ratcliff, and E.-J. Wagenmakers. 2016. "Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions." *Annual Review of Psychology 67*: 641–666.

Fung, G., and O. L. Mangasarian. 2001. "Proximal support vector machine classifiers." In *Proceedings KDD-2001: Knowledge discovery and data mining*, edited by F. Provost and R. Srikant, 77–86. San Francisco, CA, New York: Association for Computing Machinery.

Goldstone, R. L., and A. T. Hendrickson. 2010. "Categorical Perception." *Wiley Interdisciplinary Reviews: Cognitive Science 1*: 69–78.

Goldstone, R. L., D. Landy, and L. Brunel. 2011. "Improving Perception to Make Distant Connections Closer." *Frontiers in Perception Science 2*(385): 1–10.

Goldstone, R. L., and Y. Sakamoto. 2003. "The Transfer of Abstract Principles Governing Complex Adaptive Systems. *Cognitive Psychology* 46: 414–466.

Goldstone, R. L., and G. Theiner. 2017. "The Multiple, Interacting Levels of Cognitive Systems (MILCS) Perspective on Group Cognition." *Philosophical Psychology 30*(3): 334–368.

Goldstone, R. L., and U. Wilensky. 2008. "Promoting Transfer by Grounding Complex Systems Principles." *Journal of the Learning Sciences 17*: 465–516.

Goodman, N. 1972. "Seven Strictures on Similarity." In *Problems and Projects*, edited by N. Goodman, 437–447. New York: Bobbs-Merrill.

Hannon, E. E., and L. J. Trainor. 2007. "Music Acquisition: Effects of Enculturation and Formal Training on Development." *Trends in Cognitive Sciences 11*: 466–470.

Harnad, S. 1997. *Categorical Perception*. Cambridge, UK: Cambridge University Press.

Levenson, R. M., E. A. Krupinski, V. M. Navarro, and E. A. Wasserman. 2015. "Pigeons (*Columba livia*) as Trainable Observers of Pathology and Radiology Breast Cancer Images." *PLOS ONE 10*(11): e0141357.

Marshall, J. A. R., R. Bogacz, A. Dornhaus, R. Plaque, T. Kovacs, and N. R. Franks. 2011. "On Optimal Decision-Making in Brains and Social Insect Colonies." *Journal of the Royal Society Interface 6*: 1065–1074.

Milner, A. D., and M. A. Goodale. 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.

Moreno, S., and G. M. Bidelman. 2014. "Examining Neural Plasticity and Cognitive Benefit through the Unique Lens of Musical Training." *Hearing Research 308*: 84–97.

Nakayama, K. 2005. "Modularity in Perception, Its Relation to Cognition and Knowledge." In Blackwell Handbook of Sensation and Perception, edited by E. B. Goldstein, 737–759. Malden, MA: Blackwell.

Nosofsky, R. M. 1984. "Choice, Similarity, and the Context Theory of Classification." *Journal of Experimental Psychology: Learning, Memory, and Cognition 10*: 104–114.

Poldrack, R. A., et al. 2005. "The Neural Correlates of Motor Skill Automaticity." Journal of Neuroscience *25*: 5356–5364.

Rawls, J. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Belknap Press.

Rosch, E., and C. B. Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology 7*: 573–605.

Rumelhart, D. E., and D. Zipser. 1985. "Feature Discovery by Competitive Learning." *Cognitive Science 9*: 75–112.

Schultz, W., P. Dayan, and P. R. Montague. 1997. "A Neural Substrate of Prediction and Reward." *Science 275*(5306): 1593–1599.

Shum, J., D. Hermes, B. L. Foster, M. Dastjerdi, V. Rangarajan, J. Winawer, K. J. Miller, and J. Parvizi. 2013. "A Brain Area for Visual Numerals." *Journal of Neuroscience 33*: 6709–6715.

Stavins, R. N. 2003. "Experience with Market-Based Environmental Policy Instruments." In *Handbook of Environmental Economics*, edited by K. G. Mäler and J. Vincent, 355–435. Amsterdam: Elsevier Science.

Stewart, N., N. Chater, and G. D. A. Brown. 2006. "Decision by Sampling." *Cognitive Psychology 53*: 1–26.

Sutton, R., and A. Barto. 1998. *Reinforcement Learning*. Cambridge, MA: MIT Press.

Theiner, G., C. Allen, and R. L. Goldstone. 2010. Recognizing Group Cognition. *Cognitive Systems Research 11*: 378–395.