

teorema

Vol. XXXVII/3, 2018, pp. 145-150

ISSN 0210-1602

[BIBLID 0210-1602 (2018) 37:3; pp. 145-150]

COMMENT TO CLATTERBUCK

Daniel Dennett

Hayley Clatterbuck has given us a useful and constructive account of the problems facing those of us who want to explain the minds of human beings by thinking about how they got that way instead of trying to analyze the “finished product” independently of its history. As she notes, human minds are uncontroversially unique in some ways, and this very uniqueness is one of the features in need of explanation. It is worth highlighting her quotation from Frans de Waal, who opines otherwise:

By far the simplest assumption regarding the social behavior of the chimpanzee, for example, is that if this species' behavior resembles that of ourselves then the underlying psychological processes *must* be similar too [de Waal (1991), p. 298].

This is indeed the simplest assumption and also a deeply attractive assumption, but it is wise to set it aside and adopt, with strenuous discipline of the imagination, the tentative assumption that animal minds, including chimpanzee minds, are profoundly different from ours. Why? Because in the survey of life from bacteria to Bach we find ubiquitous cases of *competence without comprehension*, and theoretical reasons for concluding that comprehension is almost always optional, not necessary, and when it does arrive on the scene, it is both partial and variable, even in our own species. There remain pivotal unanswered empirical questions about what is cause and what is effect, the “chicken egg problem” and the “why only us” problem, and these can best be approached by treating *lack* of human-level comprehension as the null hypothesis, as advocated by Conwy Lloyd Morgan and others.

My canon of parsimony, illustrated by the Tower of Generate and Test, is an adaptation of earlier proposals, marking in idealized, oversimplified terms, a few of what I argue to be landmark differences. As Clatterbuck notes, Darwin himself was resolutely gradualist, and imagined that

something like Humean associationism, variously expanded and intensified, could take us all the way from “the lowest fishes” through the other apes to us. This was also Skinner’s view: operant conditioning all the way up. I aim to preserve gradualism, which abhors giant leaps and near-miraculous mutations *à la* Chomsky, while still marking plateaus that signify major transitions *à la* Szathmari and Maynard Smith. Clatterbuck rightly observes that Darwin’s naïve vision was both ways distorted: he was over-optimistic about the power of associationism and over-anthropomorphic in his assessment of the mental powers of animals. And, while she appreciates the rationale of the Darwinian/Skinnerian/ Popperian/Gregorian ladder, she sees, correctly, that “it is not obvious how turning the knobs on a set of existing traits could yield human behaviors that seem to be genuine discontinuities with what we find in even closely related species.” Indeed, it is not obvious, and moreover, she is right that I skimmed on any discussion of just how the transitions from Skinnerian to Popperian to Gregorian could be accomplished evolutionarily. I am grateful for her observation that it is the Skinner/Popper transition, or something very much in the same territory, that Lloyd Morgan identifies as requiring a new and different kind of mind. It is time to repair that omission, not with a worked out and tested theory, but with a sketch that could be filled out.

All four of my kinds of creatures are deliberately oversimplified, just cartoons or caricatures of actual kinds of minds, but I do think they illustrate the innovations that have to have evolved one way or another. Darwin was right, I maintain, to think that something like the process of natural selection must be reiterated in any non-question-begging account of learning or intelligence. My characterizations are really more like the “specs” of such minds, only hinting at how implementation might be accomplished. When we reverse-engineer these minds, we tend to look at them from the perspective of an intelligent designer who knows what task has to be accomplished, whereas evolution simultaneously discovers both means and ends in a hill-climbing process that opportunistically saves whatever appears to be local progress. From the intelligent designer’s perspective, a Skinnerian creature is like a lazy person who knows better but riskily tries out “all possible” avenues instead of looking before leaping, while a Popperian creature is like a person who wisely, on purpose, builds and tests off-line models before acting. In the real world neither Skinnerian nor Popperian creatures approach their predicaments with the understanding tacitly assumed in the caricature.

Neuroscience has made progress identifying Skinner’s postulated reinforcement mechanisms, and they prove to have features that can

serve as intermediaries on the gradual path to “wild type” Popperian creatures. There are creatures in between those who “blindly” submit Skinnerian operants to the environment’s verdicts and those who are Popperian offline hypothesis-*formulators*. These are creatures who benefit from hints embedded in their learning mechanisms. I suggest that in *rudimentary* Popperian creatures the available options are ranked by immediately aroused affective reactions: some Skinnerian *anticipations* have a premonitory “smell” that is either aversive or positive. From this perspective, the Garcia effect — one-shot learning of a particularly ominous regularity that is literally yoked to olfaction — can be seen as an extreme case on something like a continuum, with hard-to-learn and easy-to-learn affordances in between. In the caricature, the creature endowed with a Garcia effect mechanism says to itself “Ugh! Well I won’t do *that* again!” while a more Skinnerian creature stupidly takes more hard knocks before it eventually learns the lesson. Does a rat with a Garcia effect qualify as a Popperian creature because it has an internal model of the world that can be quickly updated? It has a design improvement with a Popperian rationale, but it can be readily tricked because it relies too mindlessly on the olfactory association; it has a cheap substitute for a more comprehending model.

We fancy human thinkers frequently exhibit these intermediate phenomena as well, succumbing to premature evaluations in our attempts at Popperian problem-solving. Many thinkers “instinctively” shy away from any hypothesis that “smacks of teleology,” for instance. Their resistance to even *considering* some hypotheses is not a reasoned criticism, an evaluation based on analysis, but a quick-and-dirty “I don’t like the smell of it!” reaction. Marcel Kinsbourne has suggested (personal communication) that what makes *any* problem difficult for us is that something attractive, and false, stands in the way. Perhaps the chief ingredient in becoming a truly creative and open-minded thinker is learning how to tolerate/withstand some of these negative affective reactions long enough to actually *play out* the testing of the option. Notice that not only can our thinking be shaped by such “subliminal” forces (we are typically not *aware* of being influenced by them), but also our ability to overcome such subliminal forces can grow, with practice, without our becoming aware that we are engaged in such mental exercising. Competence precedes comprehension.

Clatterbuck has the congenial habit of pushing my metaphors to the limit, usually to welcome effect, but occasionally I want to backpedal:

The DNA metaphor suggests, perhaps, that human culture already comprised a type of superorganism, where words serve as its method of reproduction. Or words may serve as a type of central nervous system, coordinating the far-flung activities of the organism of a human society.

These are both avenues to explore, but with caution; a core difference between human cultural organizations and organisms is that the myriad cells that compose organisms are incredibly myopic and incurious compared to the human beings who play the implementing roles in cultural institutions. We human operatives may often be dupes or dunces, deeply deluded about the larger scheme of things in which we are participating, but even so we can wonder what it's all about, a reflective talent far beyond any termite or neuron. Our kinds of signaling are profoundly unlike the signaling between cells and cell assemblies [Cao (2012)], and what the frog's eye tells the frog's brain is not in the same league with what the neuroscientist tells the cognitive ethologist.

Words are memes that can be pronounced, and memes are like viruses, I claim. Clatterbuck expands on this:

Often these viruses are harmful or neutral to their hosts, but occasionally they are beneficial and can indeed be integrated into the genetic code of the host and used to produce valuable new proteins.

I didn't say or mean to imply the latter point, and I don't know if there is any evidence for this. Maybe there is. I think it is possible. Whether viral nucleic acid is incorporated (the way the genes of early endosymbiont prokaryotes eventually got integrated into nuclear DNA in eukaryotes) is an open question so far as I know. Clatterbuck goes on to say, correctly, that the metaphors I have used

make different claims about what must have already been present in order for the evolution and subsequent de-Darwinization of words to take off, and by mixing models, we risk taking for granted that the initial conditions of the models obtained. This question regarding the preconditions of meme evolution becomes important when we consider the desiderata of an evolutionary account of human uniqueness that we started with.

Her "best attempt at briefly summarizing" my account is first-rate, and includes an emphasis I wish I had thought of expressing:

A word will earn its keep if it can yield beneficial effects for its user, regardless of whether the user's own actions or someone else's bring about

those effects and perhaps regardless of whether the user ever recognizes that the word reaps such benefits.

Once again, competence precedes comprehension, and this is as true of adult speech acts as infant babblings. Ruth Garrett Millikan's new book, *Beyond Concepts: Unicepts, Language, and Natural Information* (2017), expands on these issues and helps us break the bad habit of assuming, with tradition, that (as one might put it) "you can't really *use* words until you understand what words *are*, and what they *mean*." This intellectualist presupposition is embodied in all the GOFAI models of linguistic recognition, interaction, production (and hence translation), and is apparently an axiom of Chomskian linguistics in at least most of its variants, but from an evolutionary perspective one can see that this dogma characterizes an idealized end state ("the specs" for a speaker/*understander* of, say, English) that is only ever approximated by "adult native speakers" who gradually learned to use words, to notice that they were using words, to adjust their idiosyncratic meanings (mainly *without* noticing) to the meanings of their interlocutors, and so forth. The slow dawning and subsequent enhancement of (shared) comprehension is an *effect* of linguistic practice, not a prerequisite, both ontogenetically and phylogenetically.

But this practice does require a prior bias in favor of cooperation and joint attention, does it not? I don't think so; I think these can co-evolve together, with primordial varieties of primitive imitation (or imitation-like behavior-shaping) earning their keep much as they have done in all the species exhibiting "animal traditions" in the sense of Avital and Jablonka (2000) while providing the raw material to be further shaped by evolutionary pressures into "proper" cooperation and joint attention. This brings us back to the "why not everyone?" issue. Why haven't these same foundations in those species led to the development of language and the Gregorian explosion? Maybe because something about the social structure of our ancestral lineages favored the establishment of not just "pushmipullyu" signs [Millikan (1984)] such as alarm calls, but also *labels* for affordances offering multiple kinds of opportunities [Clark and Karmiloff-Smith (1993), Dennett (1993)]. Clatterbuck has raised the right questions, and set the table with much food for thought.

REFERENCES

AVITAL, E. and E. JABLONKA (2000), *Animal Traditions: Behavioural Inheritance in Evolution*, Cambridge, Cambridge University Press.

- CAO, R. (2012), "A Teleosemantic Approach to Information in the Brain," *Biology and Philosophy*, 27, pp. 49-71.
- CLARK, A. and A. KARMILOFF-SMITH (1994), "The Cognizer's Innards: A Psychological and Philosophical Perspective on the Development of Thought," *Mind and Language* 8 (4), pp. 487-519
- DENNETT, D. (1993), "Learning and Labeling," *Mind and Language* 8, (4), pp. 540-547.
- DE WAAL, F. B. M. (1991), "Complementary Methods and Convergent Evidence in the Study of Primate Social Cognition." *Behaviour* 118, pp. 297-320.
- MILLIKAN, R. (1984), *Language, Thought and Other Biological Concepts*, Cambridge, MA: MIT Press.
- (2017), *Beyond Concepts: Unicepts, Language and Natural Information*, Oxford, Oxford University Press.
- SZATHMÁRY, E. and J. MAYNARD SMITH (1995), *The Major Transitions in Evolution*, Oxford, Oxford University Press.