



[Home](#)

COGNITIVE GADGETS

Cecelia Heyes

Reviewed by
Philip Gerrans

Cognitive Gadgets: The Cultural Evolution of Thinking[□]

Cecelia Heyes

Cambridge, MA: Harvard University Press, 2018, £40.95

ISBN 9780674980150

Cite as:

Gerrans, P. [2023]: 'Cecelia Heyes's *Cognitive Gadgets*', *BJPS Review of Books*, 2023

How does the infant mewling and puking in its mother's arms develop into an adult equipped with the knowledge and skill to navigate and transform her physical and social world? The human ability to learn is the

topic of this fascinating book. It looks past the products of human cognition (artefacts, practices, natural languages, theories, and stories, the warp and woof of 'culture') to the cognitive processes on which they depend.

Heyes returns us to arguments about the nature and extent of heritable specialization (henceforth, domain specificity) in the cognitive phenotype synthesized in the 1990s by Evolutionary Psychology (pp. 9–16). Evolutionary Psychology integrates evidence about developmental canalization,¹ selective deficits and dissociations,² neural localization, inheritance, encapsulation of so-called system 1 from system 2 processes,³ learning theory, and evolutionary modelling,⁴ and argues that domain specificity is the result of genetic evolution. Of course, domain specificity associated with any or all of these 'markers' does not entail genetic evolution. However, Evolutionary Psychology links these considerations to the poverty of stimulus argument for innateness to provide an inference to the best explanation that genetic evolution explains domain specificity.

The most general form of the poverty of stimulus argument is that evidence available in the learning environment is insufficient to explain the knowledge acquired by the learner. Specifically, a domain-general learning device—one applying general principles of knowledge acquisition such as inductive reasoning, logic, reinforcement learning (of associations among perceived objects, properties, and relations) or statistical analysis—could not possibly acquire and deploy relevant domain-specific knowledge in time and on time (Cowie [1999]). The classic modern version is Chomsky's argument that the grammatical knowledge rapidly and automatically acquired by children could not be acquired from information in their linguistic environment. Some core generative linguistic knowledge must be innate.

Heyes's book is so important because she is engaging on this terrain. She agrees with Evolutionary Psychologists that we have some evolved cognitive structures that are responsible for our cognitive phenotype. However, she argues that these templates are constructed by domain-general processes of associative learning and reliably transmitted from generation to generation by cultural learning. They are a product of cultural not genetic evolution. She calls these structures 'apps'. One way of thinking of the difference between Heyes and Evolutionary Psychologists is that Heyes argues that these apps are installed through cultural learning. Evolutionary Psychologists argue that these apps are genetically specified and the role of cultural learning is to initialize them.

It is agreed on all sides that fluent cognition requires—and humans develop, more or less quickly—cognitive specialization. It is also agreed on all sides that humans have a range of domain-general abilities, in particular, associative and reinforcement learning, statistical inference and logical reasoning, executive function, explicit memory, and meta-representation (the ability to represent lower-order representational relationships). Debate continues over the form, extent, and neural implementation of these domain-general abilities (Arciuli [2017]).

We also have some low-level adaptations for learning from our peers. Human infants are predisposed to attend selectively to, and process, sources of information such as faces, emotional expressions, and intentional movements from which they can rapidly acquire necessary input for other forms of learning. Humans are also unique among primates in their tolerance for, and affectively scaffolded engagement with, juveniles. The protracted human childhood that renders children and mothers in ancestral environments uniquely vulnerable is an adaptation for social learning (Konner [2010]).

As Heyes reminds us, the degree of innate bias necessary to create this learning niche can be very slight: a strongly rewarded bias to attend to the most minimal representation of facial geometry and towards intentional movement can tune a child into its social world at the outset (pp. 60–66). Not only that, but this tuning need not depend on intrinsically social mechanisms. Emotional expressions, affective touch, and vocal tone, for example, produce social rewards in the form of positive affect; but the mechanisms of association between reward and stimulus are not intrinsically social or specialized. In Harlow's famous experiment, monkey infants preferred soft, nurturing female bodies to wire dummies that provided food. He concluded that the preference for affective maternal contact was innate. We are now conducting a massive experiment by babysitting infants with smartphones that provide cunningly engineered rewards from non-social stimuli. Will the human infant of 2027 prefer the phone or its parent?

In any case, someone impressed by the poverty of stimulus argument for innatism will still want to know how an infant endowed with initial biases to social interaction and maturing higher-order, domain-general abilities develops the suite of abilities characteristic of our species absent a pre-existing set of templates to filter and structure the information to which she is exposed. After all, it is a long way from being immersed in a socio-linguistic niche to inferring rules regulating anaphoric dependence and the scope of negative disjunctions from ambiguous sentences such as 'he said he did not order either pasta or sushi' (Crain et al. [2017]). Such sentences, processed in the same way by Mandarin and English speaking children, are taken by Chomskian linguists as evidence for presence in the cognitive phenotype of a shared, innately specified, language-processing app. Heyes needs to provide an alternative account of this ability.

This is why the example of reading turns out to be important to Heyes's overall argument. The way to dispute the Evolutionary Psychology inference to genetic specification is to show that the learning environment provides sufficient structure to allow a domain-general learning system to construct (install) a specialized app (pp. 19–20, 148–51). Reading depends on a cognitive system constructed by non-specialized processes (grapheme–phoneme mapping, a species of associative learning). The result is a specialized ability with a discrete neural substrate that has all the 'markers' mentioned above without being genetically specified. There are genes associated with reading, but these genes turn out to be relevant to other capacities as well. Reading requires expensive and extensive installation via a structured learning environment.

This idea applies and develops a familiar one exemplified by the co-dependent relationship between cultures of dairying and the LCT gene for processing lactose. The gene creates and reproduces the adaptive environment and vice versa. In the case of cognitive evolution, the mechanism of inheritance is not a gene but what Heyes calls an app: a neural circuit whose cognitive architecture is necessary for construction of the cognitive niche. The niche is necessary for the inheritance of the app in a feedback loop.

Consider reading once again. In environments where reading is adaptive (that is, literate societies), the development of reading circuitry is reliably canalized and reproduces both the niche and itself. Stable changes in the adaptive environment will produce changes in the mechanism. As societies become illiterate, specialized circuitry for grapheme–phoneme mapping will disappear from the phenotype. The app will be deleted.

Can this idea be extended? In particular, Heyes wants to extend it to explain the 'big three' apps that, together or in combination, are often invoked to explain the human cognitive phenotype: imitation, language, and mind-reading. The arguments and the evidence here are complex and span disciplines.

The idea that human imitation (as opposed to mimicry, contagion, or emulation) is a consequence of genetic evolution is intuitive. It is also supported by two arguments that help to create an 'urban myth' (Gopnik [2007]) or 'revolutionary grand theory' (Ramachandran [unpublished]) to explain cognitive modernity in humans. 'Cognitive modernity' is the name given to the period marked by the appearance in the archaeological record of evidence of artefacts and practices requiring sophisticated cognition.

The first comes from studies of neonatal imitation (of tongue protrusion). The idea that infants imitate tongue protrusion from birth was readily taken as evidence of an innate capacity for imitation (Meltzoff and Moore [1983]). Yet, as Heyes notes, the most recent comprehensive controlled study found no evidence for the phenomenon, or of any other form of imitation in infants. Infants stick their tongues out when they are aroused, and faces in close proximity sticking their tongues out arouse them (Oostenbroek et al. [2016]).

She then turns to the other main argument in support of an innate capacity for imitation. So-called mirror neurons, initially found in the parietal cortex of macaque monkeys, respond both to observation and performance of actions such as grasping or reaching. If this circuitry constitutes a mechanism for imitation, then perhaps it is leveraged in humans in combination with our other capacities to provide the basis for sophisticated imitative and even mind-reading abilities. If such a mechanism were innate, one of the most difficult cognitive problems in skill acquisition would be solved, namely, the correspondence problem of mapping a perceptual representation of movement to the representation of a motor plan for the movement, without visual feedback from one's own body. As Heyes points out, this mapping is actually not automatic and effortless but depends on associative learning, which is why the same neurons in the macaque monkeys respond not just to observations of reaching for a peanut but the sound of the food container being opened.

Some of Heyes's best-known experimental work (Brass and Heyes [2005]; Catmur et al. [2009]) addresses the way the correspondence problem is solved. If imitation is innate (based on a pre-existing template), it should be harder to learn and easier to unlearn non-imitative associations between self- and other-produced movements. If there is no significant difference between the two cases, the hypothesis is that imitation is a sub-class of associative learning, not a distinct process. This is the conclusion Heyes reaches. She argues that development of imitation beyond motor resonance into a complex skill in humans depends on a range of scaffolds including mirrors, self-referential language, and instruction. She also suggests that in cases like rhythmic dance or chant, where people synchronize movements, the goal is not imitation of object-directed action but communicative or gestural performance.

This latter point turns out to be important because some champions of imitation overestimate its scope as a way to learn complex object-directed skills. The reason is that imitation is the reproduction of a precise movement sequence as a means to realize a goal. I can imitate you closing the door with your hand by using the same movement sequence, or I can emulate you by closing the door with my foot, a different means to the same end. Much human learning of complex skills is a mixture of imitation and emulation, with the former a special case of the latter.

What is loosely described as learning by imitation is a complex suite of higher-order conceptual and lower-order sensorimotor skills. In particular, it requires three related skills: the ability to evaluate and make adjustments in light of goals, responsiveness to instruction in natural language, and the capacity to understand a teacher's explicit or implicit intentions. If mirror neurons were all we needed, we could all learn to dance by watching videos of Black Pink.

The need to share goals and intentions leads to the second of the big three: mind reading. Mind reading refers to the ability to infer or predict how others (mis)represent their world and to behave accordingly. Explicit mind reading is manifest around primary school age in explanations like, 'she thinks he thinks the teddy bear is in the toybox'. An influential Evolutionary Psychology explanation of mind reading is that the ability to predict the behaviour of others in 'representationally opaque' contexts, where others are acting on false information (perhaps someone has moved the teddy bear!), depends on an innate mind-reading module.

For Heyes, in contrast, 'language comes first' and, in concert with the maturation of other high-level capacities, explains the development of mind reading. An immersive linguistic environment in which psychological states are constantly referred to provides both a syntactic framework (recursive embeddings of mental state terms in structures like, 'She thinks that >') and semantic knowledge about mental states as unobservable causes of behaviour. The trajectory of mind reading is very sensitive to these scaffolds. Similarly, close attention to the role of executive functioning and general cognition seems to show interaction with, rather than independence from, (explicit) mind reading (Qureshi et al. [2010]). Which is, as Heyes says, what we might expect if mind reading is more like print reading than face recognition. The range of evidence and depth of discussion exceed the scope of a review, but Heyes mounts a persuasive case that a genetically specified module for mind reading is unlikely. On her view, children are already tuned into an immersive environment, which, as they mature, provides a wealth of stimulus for the hypothesis (implicit or explicit) that others have inner mental states (pp. 148–55). Variations in acquisition of mind reading should thus depend on cascading consequences of differences in very early tuning mechanisms (autism), contributions of other components of the system such as executive systems and languages, or on environmental scaffolds such as social interaction (Everett [2012]).

A difficulty she acknowledges for her view is the presence of early implicit mind reading or 'anticipatory looking' in very young infants and chimpanzees. In anticipatory looking, an observer watches while a salient object previously seen by a subject in location *A* is moved to location *B* while the subject is not present. The subject then returns to the scene of the crime. Eye-tracking equipment shows that infant and chimpanzee observers (but not autistic children) rapidly saccade to the location *A*. This is taken to indicate that they 'know' the subject will look to *A*. A richer interpretation is that they have an implicit representation that the subject believes the object is at *A*. In other words, they implicitly meta-represent the subject's mental state. Deceptive behaviour of wrasse and scrub jays (who cache food and only retrieve it when conspecifics are not looking) invite similar mind-reading interpretations.

Anticipatory looking might be an early stage of mind reading (in which case, why the dissociation with explicit mind reading?) or a separate mind-reading system (in which case, why evolve two mind-reading systems?). Or it may be another precursor mechanism of social cognition that does not depend on intrinsically social processes. This latter interpretation is naturally the one favoured by Heyes. Heyes explains anticipatory looking as 'submentalizing'. In submentalizing, the subject associates a salient feature of the environment with the standard behaviour of the agent (gaze) and retrieves that association when cued (by return of the agent in typical scenarios). The advantage of Heyes's view is that it reduces anticipatory looking to 'behaviour reading' or prediction rather than mentalizing. The disadvantage is that there is no neat unifying explanation here. Each case will have to be examined and tested to see which elements of the stimulus array are predictive, available for processing, associative learning, and cued for retrieval. This seems messy and *ad hoc*, even behaviouristic, to those who prefer the elegance and parsimony of the poverty of stimulus argument

and Evolutionary Psychology. But Heyes might well reply that if cued associative learning can do the job, then postulating extra meta-representational cognitive machinery is otiose. It is an empirical question.

Which takes us to the last of the big three, language. The poverty of stimulus argument is *a priori* (which is why Plato used it), but whether and how its premises are satisfied is an empirical matter. So anti-Chomskians argue that children's primary linguistic data does contain enough information in terms of both positive and negative examples to enable them to build a model of the correct grammar ('Bianca was too angry to eat'. To have lunch? Or to be cannibalized?) using domain-general cognition. A complicating factor is an inability to agree on what counts as decisive evidence. The reason these matters are so difficult and recondite is a feature not a bug of Chomskian theory. Namely, Chomsky, although a linguist, is not essentially interested in natural language but in the recursive combination of representational atoms into molecular structures and compounds according to rules. On his view, this is what thought is. Thought is language, but not natural language. His view of language is like Wittgenstein's rather perverse view of architecture: 'I'm not interested in buildings but what it is to be a building'.

A recent version of Chomsky's ideas is that the architecture of linguistic processing (now named MERGE) is hierarchical, recursive, and binary—that is, two linguistic atoms can be combined and nested inside a higher-level atom to create molecules, and so on. This is a simple and powerful computational procedure and its simplicity is part of its appeal: 'the operation Merge, linked to atomic conceptual resources to create a "language of thought", perhaps near optimally' (Chomsky [2007]). Note that, as described, MERGE is domain general. The atoms could be any form of representation. Its application to natural language to create trees that represent phrase structure is of course domain specific. So much of linguistics is about the way this 'core' of linguistic cognition is applied to natural language (the interface) to allow the receiver to decode it back into the language of thought. It is quite consistent with Chomsky's view that this application might recruit or reuse circuitry that evolved for other purposes and depend on domain-general processes. Indeed, MERGE is conceived of as a cognitive tweak within domain-general processing, a small cognitive mutation with massive cascading consequences for the mind.

These debates have been renewed by the advances of artificial intelligence in the form of large language models (LLMs) like GPT-4. LLMs are trained on huge datasets of internet-based text (often at a sub-word but supra-character encoding) to predict upcoming linguistic material. The training is by gradient descent learning in which the parameters (up to a trillion in GPT-4) of the system are optimized to minimize errors in input-output mappings. LLMs are quintessentially domain general. Their input is text strings from sentences about all topics on the internet. And their parameters are set using algorithms that implement domain-general statistical learning (Buckner [2023]; Piantadosi [unpublished]).

Earlier versions of LLMs struggled with theory of mind tasks and with syntactically complex sentences (such as centre embedding). But GPT-4 passes classic mind-reading tests at the same rate as seven-year-old children (Marchetti et al. [2023]). Similarly, anyone who has interacted with GPT-4 can attest to its ability to parse and produce syntactically complex sentences.

As Heyes says, the present state of cognitive neuroscience offers no decisive evidence for MERGE. Equally, given the complexity and opacity of LLMs, verifying a theory of their cognitive architecture is difficult. However, all is not dark inside. It does seem clear the activities LLMs are not structureless, producing 'order out of chaos', as Wittgenstein said. Rather, they are constructing hierarchical feature maps in activation and weight (parameter) space. How closely these spaces approximate cognitive architectures proposed by

innatists of various domains is now terrain on which these debates will be conducted (Buckner [2023]). For example, Manning et. al. ([2020], p. 30053) recently argued that a 'representation learner trained via self-supervision on word prediction tasks [...] implicitly learns to recover the rich latent structure of human language'. What they meant was that patterns of activation in hidden layers of their LLM corresponded to syntactic categories hypothesized by innatists to be unlearnable by domain-general systems. This approach to interpreting LLMs is part of a wider one in which 'LLMs can be studied using methods designed to investigate another capable and opaque structure, namely the human mind' (Hagendorff et al. [2023], p. 837).

Heyes has produced an outstanding piece of synthetic philosophy of science, cognitive science, and theoretical psychology. It makes one look again at the evidence and the adequacy of theoretical explanations advanced in almost every domain of human, and now artificial, thought.

Philip Gerrans

University of Adelaide

philip.gerrans@adelaide.edu.au

References

- Arciuli, J. [2017]: 'The Multi-component Nature of Statistical Learning', *Philosophical Transactions of the Royal Society B*, 372.
- Brass, M. and Heyes, C. [2005]: 'Imitation: Is Cognitive Neuroscience Solving the Correspondence Problem?', *Trends in Cognitive Sciences*, 9, pp. 489–95.
- Buckner, C. [2023]: *From Deep Learning to Rational Machines: What the History of Philosophy Can Teach Us about the Future of Artificial Intelligence*, Oxford: Oxford University Press.
- Catmur, C., Walsh, V. and Heyes, C. [2009]: 'Associative Sequence Learning: The Role of Experience in the Development of Imitation and the Mirror System', *Philosophical Transactions of the Royal Society B*, 364, pp. 2369–80.
- Chomsky, N. [2007]: 'Biolinguistic Explorations: Design, Development, Evolution', *International Journal of Philosophical Studies*, 15, pp. 1–21.
- Cowie, F. [1999]: *What's Within? Nativism Reconsidered*, Oxford: Oxford University Press.
- Crain, S., Koring, L. and Thornton, R. [2017]: 'Language Acquisition from a Bilingual Perspective', *Neuroscience and Biobehavioral Reviews*, 81, pp. 120–49.
- Duchaine, B., Cosmides, L. and Tooby, J. [2001]: 'Evolutionary Psychology and the Brain', *Current Opinion in Neurobiology*, 11, pp. 225–30.
-

- Everett, D. L. [2012]: 'What Does Pirahã Grammar Have to Teach Us about Human Language and the Mind?', *Wiley Interdisciplinary Reviews: Cognitive Science*, **3**, pp. 555–63.
- Frankish, K. [2010]: 'Dual-Process and Dual-System Theories of Reasoning', *Philosophy Compass*, **5**, pp. 914–26.
- Gopnik, A. [2007]: 'Cells That Read Minds? What the Myth of Mirror Neurons Gets Wrong about the Human Mind', *Slate*.
- Hagendorff, T., Fabi, S. and Kosinski, M. [2023]: 'Human-Like Intuitive Behavior and Reasoning Biases Emerged in Large Language Models but Disappeared in ChatGPT', *Nature Computational Science*, **3**, pp. 833–38.
- Konner, M. [2010]: *The Evolution of Childhood: Relationships, Emotion, Mind*, Cambridge, MA: Harvard University Press.
- Manning, C. D., Clark, K., Khandelwal, U. and Levy, O. [2020]: 'Emergent Linguistic Structure in Artificial Neural Networks Trained by Self-Supervision', *Proceedings of the National Academy of Sciences USA*, **117**, pp. 30046–54.
- Marchetti, A., Di Dio, C., Cangelosi, A., Manzi, F. and Massaro, D. [2023]: 'Developing ChatGPT's Theory of Mind', *Frontiers in Robotics and AI*, **10**.
- Meltzoff, A. N. and Moore, M. K. [1983]: 'Newborn Infants Imitate Adult Facial Gestures', *Child Development*, **54**, pp. 702–9.
- Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S. and Slaughter, V. [2016]: 'Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans', *Current Biology*, **26**, pp. 1334–38.
- Piantadosi, S. [unpublished]: 'Modern Language Models Refute Chomsky's Approach to Language'.
- Qureshi, A. W., Apperly, I. A. and Samson, D. [2010]: 'Executive Function Is Necessary for Perspective Selection, Not Level-1 Visual Perspective Calculation: Evidence from a Dual-Task Study of Adults', *Cognition*, **117**, pp. 230–36.
- Ramachandran, V. S. [unpublished]: 'Mirror Neurons and Imitation Learning as the Driving Force behind "the Great Leap Forward" in Human Evolution'.
- Shallice, T. [1988]: *From Neuropsychology to Mental Structure*, Cambridge: Cambridge University Press.
- Waddington, C. H. [1942]: 'Canalization of Development and the Inheritance of Acquired Characters', *Nature*, **150**, pp. 563–65.
-
-