9

REPRESENTATIONAL REDESCRIPTION: AN APPRECIATION OF ONE OF ANNETTE KARMILOFF-SMITH'S KEY CONTRIBUTIONS TO DEVELOPMENTAL SCIENCE

James L. McClelland

Since I first heard of the idea of representational redescription, I have been intrigued by it. Annette presented the idea in a talk she gave at UCSD when I was still on the faculty there, sometime around 1980. Annette's ideas on the subject are described in several places, perhaps most thoroughly in Karmiloff-Smith (1986), hereafter KS86. There, Annette laid out her view that our abilities progressed from an early stage in which the representations of knowledge are implicit through later stages where they are re-described, becoming accessible to explicit cognition and therefore to reflection and extension. Annette called this idea 'Representation Redescription (RR)'. It has been tantalising to me since I first heard of it. As we shall see, recent work in my laboratory is beginning to address some of the issues Annette hoped to address with the idea of representational redescription.

In her inimitably frank and direct manner, Annette made no bones about her reactions to my own work on cognitive development when she first heard about it, shortly after the publication of the PDP volumes (Rumelhart, McClelland, & the PDP research group, 1986) and her own 1986 paper. She was in the audience at Oxford in 1987 at a conference focused on the wave of new neural network modelling research that emerged in the 1980s, where I presented my connectionist simulations of developmental transitions in children's judgements of the roles of weight and distance in balance (McClelland, 1989). There I showed how a simple neural network, trained with examples of balance problems, could capture the developmental pattern that had been extensively documented by Siegler (1976). When confronted with balance problems where weight and distance vary on both sides of a balance scale, children first respond based only on the number of weights (choosing the side with more weights as the one that will 'go down'), before later reaching a stage where they consider distance from the balance point as well as weight in their judgments. She came up to me afterwards and let me know that I had addressed what she thought might be a part of the problem – the implicit knowledge aspect – but had failed to address what for her was the more important part, the redescription. Indeed, in her paper not long thereafter with Andy Clark (Clark & Karmiloff-Smith, 1993), hereafter CKS93, she and Andy explicitly discussed the role of connectionist models (including my balance scale model) in capturing the representational changes underlying the development of cognitive abilities. Discussing the representations learned in connectionist models, they wrote: 'Such representations remain implicit in the network's functioning. While this is the ***endpoint*** of learning in a connectionist network, in the human case it is the ***starting* point** for generating redescriptions of implicitly defined representations. In other words, current connectionist models account rather well for children's ***initial*** learning in a domain, but they do not yet adequately model the subsequent representational change posited by the RR model" (p. 488, the emphasis is in the original).

Why is representational redescription important and what is involved in it? This is discussed in KS86 pp 100–116. Four levels of representation are distinguished, although for simplicity, we can consider three, corresponding to the three phases of development Annette describes. In a first phase, implicit representations are formed, allowing, for example, a language learner to produce the correct article, marking it properly for gender, number and definiteness, when talking about an object or set of objects. During the second phase, implicit comparisons among procedures occur, and the child may go beyond what would be necessary, perhaps making errors or repairs thought to reflect the consequences of the comparison process. The child might say, to use Annette's example (pp. 113–114) 'un de mouchoir' to mean 'one handkerchief' to contrast the numerical function with non-specific reference ('un mouchoir', meaning 'a handkerchief'). During the third phase the additional marking would be dropped, so that the behaviour would look the same as the behaviour in phase one, but based on '*qualitatively different representations'* (p. 114, emphasis Annette's). The phase three, representations are thought to be potentially accessible to consciousness, and to be used in a planful way to allow the child to coordinate referential structure across a discourse, an ability lacking in children still in the earlier developmental phases. Two central assumptions are stressed in KS86: First, that this process is not at all age-related but recurs for each aspect of language or cognition in an experience-dependent fashion; and second, that it is not failure that drives the redescription process but success. Only after mastery at phase one can the child then reflect on the learned representations and progress to phase two. They do so not to correct mistakes but to build a new level of understanding.

While the details of the RR process as described in KS86 may be open to debate, it is clear that Annette was deeply insightful in calling for something beyond implicit knowledge to characterise children's (and adults') representations and cognitive operations. We do come to have a degree of strategic control over our thought processes, and exactly how we do so needs to be explained. The ideas have also been influential among neural network modellers who seek to understand the relationships between neural networks and conscious experience. The idea expressed in

CKS93 – that standard neural networks do not explicitly represent or manipulate their own knowledge – seems correct. The question, what might explicit knowledge be and how might it arise from an implicit foundation in a standard neural network has been explored extensively by Axel Cleeremans and his colleagues, building from his dissertation work with me on implicit learning (Cleeremans & McClelland, 1991).

Now, 30 years later, during the current resurgence of interest in neural networks, Andrew Lampinen, another colleague who completed his PhD in my laboratory, has been exploring how the knowledge in a neural network can be made accessible to manipulation (Lampinen & McClelland, 2020). A starting place for this work is the observation that a pattern of activation across a set of neuron-like processing units can be used to specify the strengths of the connections between neurons in a target network, relying on an intermediary network to do the translation. The pattern that specifies the connection strengths in the target network can be viewed as a more explicit representation of the implicit knowledge that might be in the connection weights of the target network, thus corresponding to a description of the knowledge in the connections. The next step is the observation that once we have a pattern of activation representing the knowledge in the connection weights, we can transform it using a neural network – the bread and butter of neural networks is transforming one pattern into another. One example Andrew has explored is learning to transform a strategy that wins at a game into a losing strategy (something we can do if we chose or if we are asked to). The neural network learns to transform the pattern that directs the target network to win a given game into a pattern that directs the network to lose it. After learning to do this for a subset of games, it can then transform the strategy for other games. Andrew has also explored using language to directly specify the pattern of activation that specifies the connection weights in the target network. A long-term possibility is that Andrew's architecture will allow us to build networks that can perform a wide range of tasks that could be specified through a linguistic representation.

Something still separates Andrew Lampinen's work from Annette's ideas about the representational redescription process. Annette saw this process as arising from within, as an active and constructive process engaged in by the learner. So far in Andrew's work, this active and constructive process is being carried out by Andrew; it does not arisefrom an active constructive process generated by the neural network itself. I believe human learners do have this ability, but we have yet to reach the point where it has been captured in our existing network architectures. We still have a long way to go, therefore, before fully realising Annette's vision. She has certainly given us a worthwhile target to aim for.

Bibliography

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind & Language 8*(4), 487–519.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235–253.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, *23*(2), 95–147.

Lampinen, A. K., & McClelland, J. L. (2020). Transforming task representations to perform novel tasks. *Proceedings of the National Academy of Sciences*. DOI:10.1073/pnas.2008852117

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In Morris, R. (Ed.)., *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 8–45). New York: Oxford University Press.

Rumelhart, D. E., McClelland, J. L., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volumes I & II*. Cambridge, MA: MIT Press.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*(4), 481–520.