The missing link: Dynamic, modifiable representations in working memory

doi: 10.1017/S0140525X08003634

Graeme S. Halford,^a Steven Phillips,^b and William H. Wilson^c

^a*School of Psychology, Griffith University, Nathan, Queensland, 4111, Australia;* ^b*Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, 305-8568, Japan;* ^c*School of Computer Science and Engineering, University of New South Wales, Sydney NSW 2052, Australia.*

g.halford@griffith.edu.au

[http://www.griffith.edu.au/school/psy/Professor Graeme Halford](http://www.griffith.edu.au/school/psy/Professor%20Graeme%20Halford)

steve@ni.aist.go.jp

<http://staff.aist.go.jp/steven.phillips>

billw@cse.unsw.edu.au

<http://www.cse.unsw.edu.au/~billw>

Abstract: We propose that the missing link from nonhuman to human cognition lies with our ability to form, modify, and re-form dynamic bindings between internal representations of world-states. This capacity goes beyond dynamic feature binding in perception and involves a new conception of working memory. We propose two tests for structured knowledge that might alleviate the impasse in empirical research in nonhuman animal cognition.

We agree with Penn et al. that the ability to recognise structural correspondences between relational representations accounts for many distinctive properties of higher cognition. We propose to take this argument further by defining both a conceptual and a methodological link between animal and human cognition. The conceptual link is to treat relational processing (Halford et al. 1998a) as dynamic bindings of chunks to a coordinate system in working memory (Oberauer et al. 2007). Such a coordinate system consists of slots and relations between them, and includes relational schemas (Halford & Busby 2007). Dynamic bindings are defined structurally, the governing factor being structural correspondence, which gives the flexibility that characterises higher cognition. It enables bindings to be modified, and it permits representations to be combined, giving the property of compositionality that is essential to higher cognition. It also permits premise integration, the core process of reasoning. Dynamic bindings involve the prefrontal cortex as well, which is late evolving and late developing (Wood & Grafman 2003), and is characterised by the sort of sustained activations needed to maintain a representation of task structure across different task instances. Working memory is at the core of higher cognitive processes, being the best single predictor of reasoning performance, accounting for more than 60% of the variance (Kane et al. 2004). We propose that dynamic binding to a coordinate system in working memory is a prerequisite for relational representations and therefore well worth studying in humans and nonhuman animals.

Humans' dynamic binding ability can be tested by briefly presenting words in separate slots, such as frames on a computer screen, then testing for recognition of the frame to which a word belonged (Oberauer 2005). This ability underlies the capacity for relational processing because the explicit representation of relational information requires binding to slots (the relation "larger than" comprises sets of ordered pairs in which the larger and smaller elements are bound to specific slots). We need a test for mapping to coordinate schemas that can be used with inarticulate participants. The delayed response task could be adapted for this purpose. For example, animals could see food hidden in one of two boxes, placed one above the other; then the boxes would be moved to a different location to remove environmental cues, and, after a delay, the animals could attempt to retrieve the food from one box. This requires dynamic binding of the food to a box, where the correct box is defined by its relation, above or below, to the other box. Thus, the spatial relationships within the set of boxes provide a coordinate system. There are potentially many variations on this paradigm, once the significance of dynamic binding to a coordinate system is recognised.

Another paradigm is the generativity test. A relational schema is induced by training on sets of isomorphic problems. Then elements of a new problem can be predicted by mapping into the schema. This is a form of analogical inference, and provides a good test for relational knowledge in humans (Halford & Busby 2007). The test can be applied to nonhuman animals using the learning set paradigm (Harlow 1949) comprising series of two-object discrimination tasks, in which the choice of one object is rewarded and the other is not. At the asymptote of training, typically after hundreds of isomorphic problems, discrimination between a new pair of objects is very rapid. In some higher primates it is close to perfect after one information trial (Hayes et al. 1953).

To illustrate, consider a new pair of objects. If A is chosen on the first trial and the response is rewarded (A+), A will continue to be chosen on a very high proportion of subsequent trials. If,

however, B is chosen on the first trial, resulting in no reward (B-), there will be a reliable shift to A on subsequent trials (win-stay, lose-shift). This paradigm has not been widely interpreted as inducing relational knowledge, but it does have potential for that purpose (Halford 1993). At the asymptote of inter-problem learning, participants could acquire a representation of a relation between slots, one rewarded and the other not. When a new pair is encountered, following an information trial when one object is found not to be rewarded (B-) it will be mapped to the non-rewarded slot, and the other (A) will be mapped to the rewarded slot of the relation (by structural correspondence rules which provide, *inter alia*, that each object will be mapped to one and only one slot). This inference can be made before the participant has any experience with the second object (A) and is a form of analogical inference. This interpretation of learning set acquisition is supported by findings that participants learn less about specific objects near the asymptote of learning set acquisition than early in acquisition (Bessemer & Stollnitz 1971). This suggests a switch to a different mode of learning late in acquisition, consistent with our proposal that the ability to process relational schemas is acquired near the asymptote of learning set acquisition. This paradigm can be used with inarticulate species, because the types of stimuli presented and responses required remain the same as in simple discrimination learning. We propose that this paradigm has been under-utilised as a measure of relational knowledge in inarticulate species. It can also be applied to more complex concepts such as oddity and conditional discrimination (Halford 1993), as well as to structures based on mathematical groups (Halford & Busby 2007).

The difficulty in resolving controversies in animal cognition is partly attributable to limitations in the power of empirical methods, as Penn et al. note. The two paradigms that we propose might break this impasse. The generativity test is adaptable for inarticulate subjects and can be used to assess induction of relational schemas. Dynamic binding in the context of a coordinate system (relational schema) can be assessed with nonhuman animals, and it affords the missing conceptual link between externally driven, perceptually grounded representations and internally driven, structurally reinterpreted representations.