specialized circuitry that enables vocal imitation, this same system requires specific input (e.g., species-specific song), is not engaged for other vocalizations (e.g., alarm calls), and in some species, shows plasticity throughout life as individuals create new songs each season. In addition, many researchers have recognized and detailed other DG processes that go beyond what Burkart et al. discuss. For example, there is considerable comparative work exploring the concept of "sameness," analogical reasoning, and algebraic computations (Martinho & Kacelnik 2016; Smirnova et al. 2015; ten Cate 2016). These are not part of the executive system, have not typically been linked to general intelligence, and yet they cut across domains and appear evolutionarily ancient.

Putting these strands together suggests that any approach to exploring the evolution of intelligence must consider the interaction between DSM and DG, understand the specificity of the content of DSM, examine a diversity of DG systems (i.e., beyond executive functions), and document how maturational changes in DG can impact the ontogeny of DSM. The content of a domain is particularly relevant as tasks within the general intelligence battery are often assumed to be part of a given domain without rigorous testing. Take, for example, work on tool use. Many researchers have considered tool technology a domain, one based in part on the functional design features of its objects. Thus, when animals such as chimpanzees and New Caledonia crows – natural tool users – show sensitivity to an object's design features, using those objects that are most likely to lead to successful outcomes, we consider this to be evidence of domain-specificity. And yet, cotton-top tamarins – a species that never uses tools in the wild and shows virtually no interest in object manipulation in captivity – show the same kind of sensitivity to an object's design features as chimpanzees and crows; furthermore, this sensitivity appears early in ontogeny in the absence of experience (Hauser et al. 2002a). This suggests that we should be more cautious with our claims of DSM capacities, and thus, how we classify the tasks within a general intelligence battery.

In conclusion, although Burkart et al. introduce a tension between DSM and DG that doesn't exist, incorrectly consider DSM perspectives as innate and inflexible, and falsely accuse other scholars of classifying nonhuman animals as rigidly DSM, they are correct in emphasizing the importance of looking more deeply at general intelligence in animals. Progress will depend on a clear articulation of the different skills tapped in the general intelligence battery, and standard methods that can be implemented across a diversity of species.

The evolution of general intelligence in *all* animals and machines

Kay E. Holekamp[a,b,d] and Risto Miikkulainen[c,d]

[a]*Department of Integrative Biology, Michigan State University, East Lansing, MI 48824-1115;* [b]*Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing, MI 48824;* [c]*Departments of Computer Science and Neuroscience, University of Texas, Austin, TX 78712;* [d]*BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824*

holekamp@msu.edu   http://www.holekamplab.org/
risto@cs.utexas.edu   https://www.cs.utexas.edu/~risto/

Abstract: We strongly agree that general intelligence occurs in many animals but find the cultural intelligence hypothesis of limited usefulness. Any viable hypothesis explaining the evolution of general intelligence should be able to account for it in all species where it is known to occur, and should also predict the conditions under which we can develop machines with general intelligence as well.

In their rich and thought-provoking review, Burkart et al. use impeccable scholarship to produce a heroic synthesis of multiple complex literatures. Their two main goals are to critically evaluate the question of whether general intelligence exists in nonhuman animals, and to evaluate the implications of general intelligence for current theories about the evolution of cognition. In our view, they accomplish the first goal extremely effectively, making a compelling argument that general intelligence is indeed widespread among animals. Regarding their second goal, they argue that existing data from vertebrates support the cultural intelligence hypothesis, which stresses the critical importance of social inputs during the ontogenetic construction of survival-relevant skills. However, the general intelligence explained by the cultural intelligence hypothesis is actually quite limited, so we must seek a more robust explanation for its evolution.

We believe that the cognitive buffer hypothesis (Allman et al. 1993; Deaner et al. 2003; Sol 2009a; 2009b; Lefebvre et al. 2013) offers a better alternative because it can account for phenomena the cultural intelligence hypothesis leaves unexplained. The cognitive buffer hypothesis posits that general intelligence is favored directly by natural selection to help animals cope with novel or unpredictable environments, where it enables individuals to exhibit flexible behavior, and thus find innovative solutions to problems threatening their survival and reproduction. In our view, Burkart et al. dismiss the cognitive buffer hypothesis prematurely. They argue that fundamental preconditions for the evolution of large brains include a slow life history and high survivorship, possible only in species not subject to unavoidable extrinsic mortality such as high predation pressure (van Schaik et al. 2012). However, much can be learned by considering apparent exceptions to "rules" like these, so we offer the octopus as one such exception.

Most octopuses are strictly solitary except when copulating, have very short lives, have countless predators, and produce thousands of offspring, most of which die. Nevertheless, they have some of the largest brains known among invertebrates (Hochner et al. 2006; Zullo & Hochner 2011); they exhibit a great deal of curiosity about their environments (Montgomery 2015); they recognize individual humans (Anderson et al. 2010); they exhibit pronounced individual differences (Sinn et al. 2001; Mather et al. 2012); they use tools; and they play (Mather 1994; Mather & Anderson 1999). Octopuses thus appear to exhibit a considerable amount of general intelligence without any opportunity whatsoever for social learning. Clearly, the cultural intelligence hypothesis cannot account for the general intelligence apparent in creatures like these.

Similarly, the cultural intelligence hypothesis offers little promise with respect to evolving general intelligence in machines. Computer scientists and robotic engineers have understood for decades that the embodiment of intelligent machines affects their ability to adapt and learn via feedback obtained during their interactions with the environment, mediated by sensors and activators (Brooks 1990; 1991; Sharkey & Ziemke 1998; Goldman & de Vignemont 2009). Most hypotheses forwarded to explain the evolution of intelligence in animals, including the cultural intelligence hypothesis, fail to address the question of how morphological traits outside of the nervous system might have shaped intelligence. In creatures such as octopuses and primates, mutations affecting nervous system structure or function, which might generate less-stereotyped and more-flexible behavior, are visible to selective forces in the environment because they can be embodied in the limbs. Thus, greater intelligence is likely to evolve in these animals than in those whose interactions with their environments are more highly constrained.

Roboticists have also realized that logic alone cannot generate much intelligent behavior in their machines, and that to achieve better performance, their robots must also want things. The skills discovered by evolutionary algorithms are diverse, and many such skills may occur within a single population of digital organisms, but individual agents are rarely motivated to acquire a large array of skills. As a result, most current evolutionary algorithms produce domain-specific intelligence in machines that

rarely possess more than a small set of skills, and they are thus suited to performing only tasks that demand that particular skill set. Although an intrinsic motivation to explore the environment has been imitated in artificial agents via machine learning (Schmidhuber 1991; Oudeyer et al. 2007), the production of generalist learners within an evolutionary context remains highly problematic (Stanton & Clune 2016).

Any selection pressure that promotes behavioral diversity or flexibility within the organism's lifetime, including the ability to learn from experience, should theoretically result in enhanced general intelligence. Novel or changing environments should select for individuals who can learn as much as possible in their lifetimes, as suggested by the cognitive buffer hypothesis. Indeed, Stanton and Clune (2016) recently developed an evolutionary algorithm that produces agents who explore their environments and acquire as many skills as possible within their lifetimes while also retaining their existing skills. This algorithm encourages evolution to select for curious agents motivated to interact with things in the environment that they do not yet understand, and engage in behaviors they have not yet mastered. This algorithm has two main components: a fitness function that rewards individuals for expressing as many unique behaviors as possible, and an intra-life novelty score that quantifies the types of behaviors rewarded by the algorithm. Agents are also provided with an intra-life novelty compass that indicates which behaviors are considered novel within the environment. The intra-life novelty compass may simply identify and direct agents toward areas of high expected learning because new knowledge often promotes the ability to perform new skills. Aligned with these results, we suggest that the primary value of the cultural intelligence hypothesis is to offer social learning as an intra-life novelty compass, but that this hypothesis provides neither the requisite fitness function nor anything analogous to an intra-life novelty score.

A viable hypothesis explaining the evolution of large brains and general intelligence should be able to account for general intelligence in any species where it is known to occur, and it should be able to predict the conditions under which we can develop machines with general intelligence as well. The cultural intelligence hypothesis simply cannot do these things.

Where is the evidence for general intelligence in nonhuman animals?

Ludwig Huber

Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna, University of Vienna, 1210 Vienna, Austria.
ludwig.huber@vetmeduni.ac.at   www.vetmeduni.ac.at/messerli

Abstract: This commentary contrasts evolutionary plausibility with empirical evidence and cognitive continuity with radiation and convergent evolution. So far, neither within-species nor between-species comparisons on the basis of rigorous experimental and species-appropriate tests substantiate the claims made in the target article. Caution is advisable on meta-analytical comparisons that primarily rely on publication frequencies and overgeneralizations (from murids and primates to other nonhuman animals).

In this thought-provoking, highly inspiring article, Burkart et al. explore the possibility of the existence of general intelligence in nonhuman animals. Given the evidence for *g* in humans, it is a reasonable and worthwhile endeavor to look for its existence in other taxa. However, to pursue a psychometric approach to nonhuman intelligence, it is necessary to obtain relevant and reliable data. As the authors themselves admit, evolutionary plausibility does not amount to empirical evidence.

Within-species comparisons. For more than a century, psychometricians have devised IQ tests to measure human intelligence. However, the breadth of test items is quite narrow. The tasks are, for the most part, administered in the same manner, with no or only modest variation of test-taking situation, motivation, or sensory domain (Locurto et al. 2006). For instance, the WAIS-IV (Wechsler et al. 2008) comprises four index scores, focusing on verbal comprehension, perceptual reasoning, working memory, and processing speed. This paper-and-pencil task may be enough to represent major components of human intelligence, but it does not tap the most interesting cognitive abilities in nonhuman animals, especially in the technical and social domains.

A crucial question in the search for the influence of an underlying general mental ability is the rationale behind which tests are included in the test batteries and the reliability of those tests for uncovering cognitive abilities. Tests measure performance, not cognitive abilities per se. A huge number of possible noncognitive factors may influence performance, from anatomical to perceptual and motivational. Therefore, it is important to know which cognitive tasks and which controls are included in the test battery. Human IQ tests are often constructed in the manner of a best-case scenario, in that tasks are included in the final battery only if they correlate positively with other tasks and loaded positively on the first component. That is, the presence of *g* is assumed and tasks chosen that verify its presence (Locurto et al. 2006). Furthermore, human IQ tests are standardized with several hundreds to thousands of people of all age classes. This is not feasible with (most) nonhuman animals.

Between-species comparisons. Large data sets for valid comparisons are only possible if we collect data from different labs. But can we rely on data sampled in different labs, using (slightly) different methods (different stimuli, apparatuses, procedures, etc.) and groups of subjects differing in important features like housing and rearing conditions, individual experiences, age, and sex composition? This is both a practical and a theoretical problem. It would demand an enormous amount of labor, money, space, and other resources to test a large sample of species in one lab. Even if one has access to a zoo or game park, testing the abilities that tap reasoning in nonhuman cognition is a difficult and time-consuming business. Furthermore, if the tasks were designed to tap different response systems, sensory modalities, and motivations, it would be a huge undertaking.

Therefore, the evidence for general intelligence on the interspecific level so far rests on meta-analyses. This strategy is based on the assumption that the frequency of reported observations of complex traits associated with behavioral flexibility is a reflection of that species' intellectual capability. For instance, Reader and Laland (2002) used indices of innovation, tool use, and social learning for their correlations. But is innovation really a direct outcome of a cognitive trait of a species? The relation is vague and the behavioral definitions are rather slippery. Furthermore, most of these meta-analyses rely on observation frequency, which may deviate widely from the experimentally proven existence of a cognitive trait in a species. For instance, reports of true imitation in callithrichids are very rare, but rigorous laboratory tests have proven its existence (Voelkl & Huber 2000; Voelkl & Huber 2007). The same is true with invisible displacement in *Callithrix jacchus* (Mendes & Huber 2004). Tool use may be the best example of the problem with drawing conclusions about species differences in general intelligence based on publication counting. It is an important ability in chimpanzees, New Caledonian crows, and Galápagos woodpecker finches. However, these species have no clear, experimentally proven cognitive superiority over their non-tool-using relatives, bonobos, carrion crows, or tree finches, respectively (Gruber et al. 2010; Herrmann et al. 2010a; Teschke et al. 2011; 2013). This led to the conclusion that habitual tool use is not a clear predictor of general intelligence, not even physical intelligence (Emery & Clayton 2009). Although