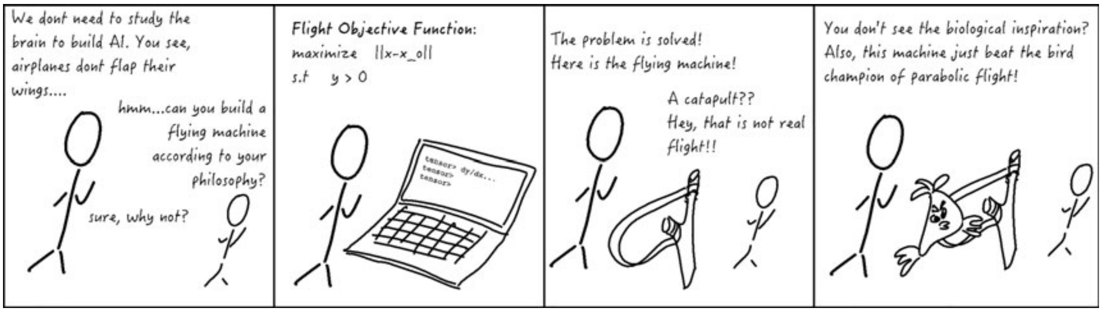

Figure 1 (George). A humorous take on the current debate in artificial intelligence.

clues about what kind of generative model the brain implements and how this model differs from models being developed in the AI community.

For instance, spatial lateral connections between oriented features are a predominant feature of the visual cortex and are known to play a role in enforcing contour continuity. However, lateral connections are largely ignored in current generative models (Lee 2015). Another example is the factorization of contours and surfaces. Evidence indicates that contours and surfaces are represented in a factored manner in the visual cortex (Zhou et al. 2000), potentially giving rise to the ability of humans to imagine and recognize objects with surface appearances that are not prototypical – like a blanket made of bananas or a banana made of blankets. Similarly, studies on top-down attention demonstrate the ability of the visual cortex to separate out objects even when they are highly overlapping and transparent (Cohen & Tong 2015). These are just a handful of examples from the vast repository of information on cortical representations and inference dynamics, all of which could be used to build AGI.

The conundrum of "human-level performance": Benchmarks for AGI. We emphasize the meaninglessness of "human-level performance," as reported in mainstream AI publications, and then use as a yardstick to measure our progress toward AGI. Take the case of the DeepQ network playing "breakout" at a "human level" (Mnih et al. 2015). We found that even simple changes to the visual environment (as insignificant as changing the brightness) dramatically and adversely affect the performance of the algorithm, whereas humans are not affected by such perturbations at all. At this point, it should be well accepted that almost any narrowly defined task can be "solved" with brute force data and computation and that any use of "human-level" as a comparison should be reserved for benchmarks that adhere to the following principles: (1) learning from few examples, (2) generalizing to distributions that are different from the training set, and (3) generalizing to new queries (for generative models) and new tasks (in the case of agents interacting with an environment).

Message passing-based algorithms for probabilistic models. Although the article makes good arguments in favor of structured probabilistic models, it is surprising that the authors mentioned only Markov chain Monte Carlo (MCMC) as the primary tool for inference. Although MCMC has asymptotic guarantees, the speed of inference in many cortical areas is more consistent with message passing (MP)-like algorithms, which arrive at maximum *a posteriori* solutions using only local computations. Despite lacking theoretical guarantees, MP has been known to work well in many practical cases, and recently we showed that it can be used for learning of compositional features (Lázaro-Gredilla et al. 2016). There is growing evidence for the use of MP-like inference in cortical areas (Bastos et al. 2012; George & Hawkins 2009), and MP could offer a happy medium where inference is fast, as in neural networks, while retaining MCMC's capability for answering arbitrary queries on the model.

Building brains that communicate like machines

doi:10.1017/S0140525X17000152, e266

Daniel Graham
Department of Psychology, Hobart & William Smith Colleges, Geneva, NY 14456.
graham@hws.edu **http://people.hws.edu/graham**

Abstract: Reverse engineering human cognitive processes may improve artificial intelligence, but this approach implies we have little to learn regarding brains from human-engineered systems. On the contrary, engineered technologies of dynamic network communication have many features that highlight analogous, poorly understood, or ignored aspects of brain and cognitive function, and mechanisms fundamental to these technologies can be usefully investigated in brains.

Lake et al. cogently argue that artificial intelligence (AI) machines would benefit from more "reverse engineering" of the human brain and its cognitive systems. However, it may be useful to invert this logic and, in particular, to use basic principles of machine communication to provide a menu of analogies and, perhaps, mechanisms that could be investigated in human brains and cognition.

We should consider that one of the missing components in deep learning models of cognition – and of most large-scale models of brain and cognitive function – is an understanding of how signals are selectively routed to different destinations in brains (Graham 2014; Graham and Rockmore 2011).

Given that brain cells themselves are not motile enough to selectively deliver messages to their destination (unlike cells in the immune system, for example), there must be a routing protocol of some kind in neural systems to accomplish this. This protocol should be relatively fixed in a given species and lineage, and have the ability to be scaled up over development and evolution.

Turning to machine communication as a model, each general technological strategy has its advantages and ideal operating conditions (grossly summarized here for brevity):

Circuit switched (traditional landline telephony): high throughput of dense real-time signals

Message switched (postal mail): multiplexed, verifiable, compact addresses

Packet switched (Internet): dynamic routing, sparse connectivity, fault tolerance, scalability

We should expect that brains adopt analogous – if not homologous – solutions when conditions require. For example, we would expect something like circuit switching in somatosensory and motor output systems, which tend to require dense, real-time communication. However, we would expect a dynamic, possibly packet-switched system in the visual system, given limited windows of attention and acuity and the need for spatial remapping, selectivity, and invariance (Olshausen et al. 1993; Poggio 1984; Wiskott 2006; Wiskott and von der Malsburg 1996).

There could be hybrid routing architectures at work in brains and several that act concurrently (consider by way of analogy that it was possible until recently for a single human communicator to use the three switching protocols described above simultaneously). Individual components of a given routing system could also be selectively employed in brains. For example, Fornito et al. (2016) proposed a mechanism of deflection routing (which is used to reroute signals around damaged or congested nodes), to explain changes in functional connectivity following focal lesions.

Nevertheless, functional demands in human cognitive systems appear to require a dynamic mechanism that could resemble a packet-switched system (Schlegel et al. 2015). As Lake et al. note, the abilities of brains to (1) grow and develop over time and (2) flexibly, creatively, and quickly adapt to new events are essential to their function. Packet switching as a general strategy may be more compatible with these requirements than alternative architectures.

In terms of growth, the number of Internet hosts – each of which can potentially communicate with any other within milliseconds – has increased without major disruption over a few decades, to surpass the number of neurons in the cortex of many primates including the macaque (Fasolo 2011). This growth has also been much faster than the growth of the message-switched U.S. Postal Service (Giambene 2005; U.S. Postal Service 2016). Cortical neurons, like Internet hosts, are separated by relatively short network distances, and have the potential for communication along many possible routes within milliseconds. Communication principles that allowed for the rapid rise and sustained development of the packet-switched Internet may provide insights relevant to understanding how evolution and development conspire to generate intelligent brains.

In terms of adapting quickly to new situations, Lake et al. point out that a fully trained artificial neural network generally cannot take on new or different tasks without substantial retraining and reconfiguration. Perhaps this is not so much a problem of computation, but rather one of routing: in neural networks, one commonly employs a fixed routing system, all-to-all connectivity between layers, and feedback only between adjacent layers. These features may make such systems well suited to learning a particular input space, but ill suited to flexible processing and efficient handling of new circumstances. Although a packet-switched routing protocol would not necessarily improve current deep learning systems, it may be better suited to modeling approaches that more closely approximate cortical networks' structure and function. Unlike most deep learning networks, the brain appears to largely show dynamic routing, sparse connectivity, and feedback among many hierarchical levels. Including such features in computational models may better approximate and explain biological function, which could in turn spawn better AI.

Progress in understanding routing in the brain is already being made through simulations of dynamic signal flow on brain-like networks and in studies of brains themselves. Mišić et al. (2014) have investigated how Markovian queuing networks (a form of message-switched architecture) with primate brain-like connectivity could take advantage of small-world and rich-club topologies. Complementing this work, Sizemore et al. (2016) have shown that the abundance of weakly interconnected brain regions suggests a prominent role for parallel processing, which would be well suited to dynamic routing. Using algebraic topology, Sizemore et al. (2016) provide evidence that human brains show loops of converging or diverging signal flow (see also Granger 2006). In terms of neurophysiology, Briggs and Usrey (2007) have shown that corticothalamic networks can pass signals in a loop in just 37 milliseconds. Such rapid feedback is consistent with the notion that corticothalamic signals could function like the "ack" (acknowledgment) system used on the Internet to ensure packet delivery (Graham 2014; Graham and Rockmore 2011).

In conclusion, it is suggested that an additional "core ingredient of human intelligence" is dynamic information routing of a kind that may mirror the packet-switched Internet, and cognitive scientists and computer engineers alike should be encouraged to investigate this possibility.

The importance of motivation and emotion for explaining human cognition

doi:10.1017/S0140525X17000164, e267

C. Dominik Güss[a] and Dietrich Dörner[b]

[a]*Department of Psychology, University of North Florida, Jacksonville, FL 32224;* [b]*Trimberg Research Academy TRAc, Otto-Friedrich Universität Bamberg, 96047 Bamberg, Germany.*

dguess@unf.edu **dietrich.doerner@uni-bamberg.de**
https://www.unf.edu/bio/N00174812
https://www.uni-bamberg.de/trac/senior-researchers/doerner

Abstract: Lake et al. discuss building blocks of human intelligence that are quite different from those of artificial intelligence. We argue that a theory of human intelligence has to incorporate human motivations and emotions. The interaction of motivation, emotion, and cognition is the real strength of human intelligence and distinguishes it from artificial intelligence.

Lake et al. applaud the advances made in artificial intelligence (AI), but argue that future research should focus on the most impressive form of intelligence, namely, natural/human intelligence. In brief, the authors argue that AI does not resemble human intelligence. The authors then discuss the building blocks of human intelligence, for example, developmental start-up software including intuitive physics and intuitive psychology, and learning as a process of model building based on compositionality and causality, and they stress that "people never start completely from scratch" (sect. 3.2, last para.)

We argue that a view of human intelligence that focuses solely on cognitive factors misses crucial aspects of human intelligence. In addition to cognition, a more complete view of human intelligence must incorporate motivation and emotion, a viewpoint already stated by Simon: "Since in actual human behavior motive and emotion are major influences on the course of cognitive behavior, a general theory of thinking and problem solving must incorporate such influences" (Simon 1967, p. 29; see also Dörner & Güss 2013).

Incorporating motivation (e.g., Maslow 1954; Sun 2016) in computational models of human intelligence can explain where goals come from. Namely, goals come from specific needs, for example, from existential needs such as hunger or pain avoidance; sexual needs; the social need for affiliation, to be together with other people; the need for certainty related to unpredictability of the environment; and the need for competence related to ineffective coping with problems (Dörner 2001; Dörner & Güss 2013). Motivation can explain why a certain plan has priority and why it is executed, or why a certain action is stopped. Lake et al. acknowledge the role of motivation in one short paragraph when they state: "There may also be an intrinsic drive to reduce uncertainty and construct models of the environment" (sect. 4.3.2, para. 4). This is right. However, what is almost more important is the need for competence, which drives people to explore new environments. This is also called diversive exploration (e.g., Berlyne 1966). Without diversive exploration, mental models could not grow, because people would not seek new experiences (i.e., seek uncertainty to reduce uncertainty afterward).

Human emotion is probably the biggest difference between people and AI machines. Incorporating emotion into computational models of human intelligence can explain some aspects that the authors discuss as "deep learning" and "intuitive psychology." Emotions are shortcuts. Emotions are the framework in which cognition happens (e.g., Bach 2009; Dörner 2001). For example, not reaching an important goal can make a person angry. Anger then characterizes a specific form of perception, planning, decision making, and behavior. Anger means high activation, quick and rough perception, little planning and deliberation, and making a quick choice. Emotions modulate human behavior; the *how* of the behavior is determined by the emotions.