THE RELEVANCE REALIZATION FRAMEWORK OF INTELLIGENCE

Relevance realization: An emerging framework of intelligence in Garlick, Van der Maas, Mercado and Hawkins

Zachary C. Irving* and John Vervaeke*
University of Toronto

Abstract

Two separate conceptions of intelligence persist within the cognitive science community: The psychometric conception in terms of general intelligence (g), and the categorical conception in terms of the criteria that an entity must meet to be an intelligent cognitive agent. In this paper, we argue that a framework of intelligence in terms of relevance realization (RR) is emerging in the literature on psychometric and categorical intelligence. The RR framework involves three pervasive constraints on processing – cognitive scope, cognitive tempering, and cognitive prioritization – that emerge from interactions on the level of neurons and cognitive modules. Both Garlick and van der Maas presuppose and require RR in order to solve central problems in their accounts of psychometric intelligence. Moreover, the apparent clash between Mercado and Hawkins brain-based conceptions of categorical intelligence can be resolved through the RR framework. Taken together, Mercado and Hawkins give brain-based plausibility to Vervaeke, Lillicrap, & Richard's computational models and philosophical arguments for a RR account of categorical intelligence.

*The priority of authors is not meant to denote primacy or extent of contribution.

Relevance realization: An emerging framework of intelligence in Garlick, van der Maas, Mercado, and Hawkins

Introduction

There is confusion within the cognitive science community regarding intelligence. Broadly, "intelligence" has been conceived of in two ways throughout the literature – categorically and psychologically – and seldom has there been a theory that integrates these two perspectives. Yet integration is needed because of important possible connections between these two conceptions of intelligence. Our paper draws upon the empirical and theoretical findings of recent integrative work to posit a new framework of intelligence that integrates these conceptions in a more satisfactory manner. Specifically, we think that an explanation of both categorical and psychological intelligence in terms of *relevance realization* (RR) is both plausible and efficacious at solving parallel problems in both the literature on both conceptions of intelligence.

The categorical conception holds that intelligence is some foundational set of criterion that an entity must meet to be a cognitive agent. For example, a human is intelligent in this sense and an ant is not. The foundational insight for artificial intelligence is that the way to produce cognitive agents is to artificially create intelligent machines. In contrast, the psychological conception pertains to intelligence difference between individual humans, typically as measured by intelligence tests. Though Steven Hawkings is no more categorically intelligent that the authors of this paper (one would hope), he likely has an edge with respect to psychological intelligence.

The term "intelligent" has a similar ambiguity to the term "rational". Often, people can lack the distinction between the categorical sense of rationality that contrasts with arationality and the psychological sense that contrasts with irrationality. It makes sense to say that a chair is arational (that is, not subject to the norms of rationality) but it makes no sense to say that a chair is irrational. In order to evaluate an entity according to the norms of rationality in the psychological sense, that entity must meet the precondition of being rational in the categorical sense. Similarly, note the ambiguity in the claim that a chair is, "unintelligent". Do we mean that the chair is not subject to the categorical norms of intelligence, or do we mean that the chair has a psychological disability? Moreover, the categorical use of intelligence would allow us to classify a rat as a highly intelligent animal even though it is incapable of solving the most rudimentary psychometric test. In order to evaluate an agent according to the norms of psychological intelligence, that agent must be categorically intelligent, but the converse does not hold.

More specifically, the psychological conception of intelligence found in the psychometric literature holds that intelligence has something to do with an individual's performance on related tasks. The most pervasive result in the psychometric literature is the positive manifold (Jensen, 1998). This is a consistent and reliable statistical result that performance on one of the standard individual psychometric tests is highly predictive of performance on other such tests. This is generally regarded as a phenomenon that is in need of explanation. The traditional approach has been to posit a theoretical placeholder – named the general factor intelligence or g – that will be eventually filled in with the causal mechanism(s) responsible for the positive manifold. Our question is: can this explanatory place-holder be filled with the features that would determine an entity's categorical membership as a cognitive agent? Our question is not merely whether categorical intelligence is a precondition for psychological intelligence. Clearly, we can see that this is true. Instead, we are interested in whether variance in the set of features that make entities categorically intelligent can account for individual differences in psychological intelligence between human beings.

What makes an entity categorically intelligent, roughly, is its ability to solve problems and interact successfully with the world. Thus, if we could describe foundational processes that are central to these abilities, we would have a working definition of categorical intelligence. Throughout the literature on problem solving and intelligence, there are certain core theoretical difficulties that bedevil all purported explanations of problem solving and environmental interaction. One possibility is that these families of difficulties entail that all our current explanations of problem solving an environmental are irretrievability wrong. Another more appealing, and we believe more likely possibility, is that these explanations assume and require a set of until recently unexplained processes. Thus, if we can identify and explicate the set of processes presupposed by theoretical explanations that attempt to address these difficulties, then we would have found foundational processes for categorical intelligence.

Three of the aforementioned core theoretical problems in the cognitive science literature on intelligence are: combinatorial explosion (Holyoak, 1995; Newell and Simon, 1972), ill-defined problems (Haugland, 1989) and the frame problem (for an anthology, see Plyshyn, 1988). The problem of combinatorial explosion was first introduced by Newell and Simon's insight that, with an increase in the number of steps to solve a problem, the possible alternative pathways that might constitute a solution exponentially increase (1972). Thus, an algorithmic search of any realistic problem space is practically impossible for any finite problem solver. An intelligent problem solver with limited time and resources, thus, must constrain a problem space so that it can be effectively searched. One of the most important contributors to constraining the search space is how the problem is formulated. The importance of problem formulation is amplified by the fact that most real-world problems (i.e. following a conversation, taking good notes, drawing a good sketch or map) are ill-defined when compared to the well defined problems of Newell and Simon (such as a logical derivation or an algebra problem). Thus, the central task of the ill-defined problem solver is precisely to formulate the problem in an appropriate manner.

This issue of problem formulation is further exacerbated by a problem solver's causal interaction with the world, as demonstrated by a generalized version of the frame problem. In order to be capable of action as opposed to mere behavior, any purported cognitive agent must be able to track the intended effects of its behavior. The frame problem arises because, in addition to the intended effects of the behavior, there is always an indefinitely large set of unintended side effects. The entity cannot ignore all of these side effects, because some of them have the potential to interfere or aid it in its effort to produce the intended effect. Yet the entity cannot search all of the actual or potential side effects, if it is to avoid combinatorial explosion. The frame problem reveals that problem formulation requires the ability to intelligently zero in on relevant information and ignore irrelevant information in both effects and side effects. In other words, the entity must be able to frame its cognition. Note that a theory of intelligence must therefore explain the process(es) of framing cognition, and not presuppose it. Part of what is meant by intelligent framing is that the entity cannot absolutely ignore what is outside of its frame, because once-irrelevant information can often unexpectedly become relevant information and vice versa. This ability to create a flexible frame for information in order to zero in on relevant information in a way that facilitates effective problem formulation is the hallmark of intelligence. Of course an entities ability to frame the environment is constituted by its ability to realize relevant information. Without "relevance realization" an entity cannot solve problems and interact with the world; in other words it is not categorically intelligent.[1]

To return to the central question of this paper, we claim that relevance realization can both determine an entity's categorical membership as a cognitive agent (categorical intelligence) and fill the explanatory placeholder of g (psychological intelligence). Throughout this paper, we will review important efforts to come up with a comprehensive and integrated account of general and categorical intelligence. From the psychometric perspective, we focus on Garlick's Plasticity Thesis and van der Maas' Mutualism Model and show how both, in different ways, fit into and require a relevance realization framework of intelligence. From the categorical perspective, we demonstrate how the seemingly contrary work of Mercado and Hawkins (the Cognitive Complexity Thesis and Memory Prediction Framework, respectively) can be resolved under the relevance realization framework.

Single Latent Factor Explanations of g

In the first section of this paper, we will give a brief overview of theories that have posited a single, latent psychological factor that can explain the positive manifold (psychometric g). In addition to significant empirical and theoretical problems, few single latent factor theories posit an explanation of psychometric g that is related to categorical g.

Garlick's Plasticity Thesis is a step in the right direction in both of these respects. His theory explains many of the empirical results associated with a variety of other theories of general intelligence (the correlations between intelligence and brain size, neural speed, and cortical efficiency among other results) while evading some of the empirical problems of these theories. Moreover, Garlick explicitly sets out to create a theory that is consistent with the Connectionist Framework of the brain (p118), which shows that he is considering theoretical approaches as to what makes humans categorically intelligent.

Though we think that Garlick has the right approach when he attempts to give an account of psychometric intelligence in terms of a definition of categorical intelligence, we think that his theory must be refined. Specifically, Garlick's theory is unclear about whether *more plasticity* or *more appropriate plasticity* causes intelligence. We argue for interpreting Garlick's theory as a *more appropriate plasticity* account of intelligence and explain this account in terms of relevance realization.

Explanations from Single Task-Oriented Factors

In order for a single *task-oriented factor* to explain the positive manifold, it must be a basic process that is used when performing most, or all, cognitive tasks. This factor could be a single cognitive process, as is proposed by the working memory hypothesis. On the other hand, this factor could be one of the basic biological properties of the brain, such as speed of information processing (the rate at which neurons fire), neural efficiency (the amount of neural resources used to perform a task) or brain size (for reviews, see Bartholomew 2004; or Jensen 1998). Note that these biological explanations focus on processes that are directly involved in the way that tasks are carried out (hence, they are task oriented).

Single task-oriented factor theories – which suggest that variance in a psychological trait g causes variance in psychometric g – need to show parity or near parity between psychometric g and some process or feature of the brain. Unfortunately, no theory has been able to do so (Bartholemew, 2004; Jensen, 1998). In order to be an adequate single-factor explanation of g, the working memory (WM) hypothesis, which has received some support over the past decade, must assert parity or near parity between psychometric measurement of g and WM. However, a review has found the relationship between the constructs of g and WM to be "substantially less than unity ([approximately] .479)" (Ackerman, Beier, and Boyle, 2005). The brain size hypothesis argues that larger brain size, and with this size, more neurons causes greater g (Willerman et al., 1991) . This hypothesis suffers from the fact that children actually loose neurons with development after birth (Morgan and Gibson, 1991) while increasing in g. Furthermore, this hypothesis should predict that, because body-size and brain size are correlated, larger people should be more intelligence on balance, which is factually incorrect (Jensen, 1998). Speed of information processing theory asserts that g is due to faster connections between neurons. This theory is highly indebted to the empirical fact that more intelligent people tend to have faster reaction times. However, Luciano et al. suggest that there is no causal link between speed of information processing and intelligence and vice versa (2005).

No single-factor theory has been consistent with a categorical definition of intelligence. Putting aside the more complex questions of combinatorial explosion, ill-defined problems, and the frame problem, even the most basic definition of categorical intelligence must involve the ability to learn. Categorical intelligence must, on some basic level, be defined as the ability for an organism to respond appropriately to the dynamic environment in which it is embedded. This does not simply entail, however, that an intelligent organism must be able to learn simple, stimulus-response type patterns. Rather, an organism must be able to develop its capacity for learning or, in Harlow's terminology, it must "learn to learn" (1949). Harlow's theory was based on the observation that monkeys progressively improved their ability to solve *new* problems, instead of simply learning to improve upon old solutions (1949). This requires that the brain be able to re-design itself to more appropriately respond to – and learn from – its environment.

Both speed of information processing and brain size theory implicitly posit a distinction between how a brain designs itself to be intelligent and how it is psychologically intelligent, as measured by g. It must be admitted that the relationship between speed of information processing theory and psychometric g is, to a degree, "conceptually sound" (Blair, 2007). If one were to compare two networks with identically connected neurons, one would expect that if one network had faster neural connectivity it would also be more intelligent. However, this argument relies on the pre-existence of a network with intelligently designed[2] connections. Otherwise, in an un-differentiated (un-designed) neural network (in which connections between neurons are random) variations in the speed of information processing have no effect on the network's intelligence (Garlick, 2002). Therefore, speed of information processing theory falls prey to the homuncular fallacy, in that it implicitly assumes the existence of a mechanism through which the brain re-designs itself. The same holds true for brain size theory. Both theories imply a conceptual distinction between the mechanism(s) that enable the brain to be categorically intelligent by re-designing itself (and thus enable the possibility of psychological intelligence), and the mechanism that generates differences in psychological intelligence between individuals.

The neural efficiency hypothesis is interesting in that it relates to the way that a neural network is connected. This model takes the finding that people with higher g ratings show lower cortical activation when they perform intellectual tasks than less intelligent individuals, and proposes that intelligence is the result of efficient connections in the brain (Haier et al, 1992; Grabner, 2004; Grabner et al., 2006). An efficiently connected neural network *is* one that has been intelligently designed. Therefore, the model implies parity between psychological and categorical intelligence. However, this theory has an irredeemable lack of explanatory power. Certainly, we cannot appeal to a network's neural efficiency to explain the design processes that make a neural network efficient.

Explanation(s) from Single Learning-Oriented Factor

Single learning-oriented factor explanations focus on the way a neural network develops rather than the way it operates while performing a specific task. The synaptic plasticity thesis takes this approach (Garlick, 2002), and argues that variance in the ability for a neural network to adapt its connections, i.e. plasticity, could account for variance in g. Another theory that may be grouped under this type is the neural pruning hypothesis (for a review, see: Jensen, 1998).

Garlick's basic argument for plasticity as an explanation of g is as follows: all intelligent human behavior is the result of patterns encoded on a human neural network. Given the same environmental stimuli, a brain that displays more (and/ or more effective) plasticity will encode more (and/ or more appropriate) patterns in its neural network. On balance, environmental stimulation is the same for all people taking an IQ test[3]. Therefore, people with more (and/ or more appropriately) plastic brains form more (and/ or more relevant) environmentally relevant patterns across their neural network. Therefore, these people are more intelligent.

Part of the strength of Garlick's theory is that it works in consort with a categorical definition of intelligence: the ability to encode appropriate patterns on a neural network. In this way, it can be seen as an embodiment of the neural efficiency theory. We have seen earlier that neural efficiency theory posits parity between psychological intelligence (neural efficiency) and categorical intelligence (intelligent, i.e. appropriate design). Thus, the cause of neural efficiency is better[4] plasticity.

Garlick's theory also explains many other empirical results associated with g, and thus subsumes the theories that were made largely to explain those results. The brain size theory rests heavily on the empirical result that g correlates with brain mass and volume when body size is controlled for (as cited in Garlick: Jensen, 1998; Wicket, Vernon, and Lee, 1994; Willerman, Schultz, Rutledge, & Bigler, 1991). Brain mass and volume also increases with intelligence over childhood (as cited in Garlick: Morgan & Gibson, 1991). This increase is likely due to an increase in connection complexity (as cited in Garlick: Blinkov & Glezer, 1968) and the related support tissue (as cited in Garlick: Diamond, 1991; Konner, 1991; Sirevaag & Greenough, 1997). Garlick argues that more plastic brains would, with the same environmental stimuli, develop more connections, and have larger brains (2002).

Speed of information processing theory relies on the empirical result that more intelligent people tend to have faster and less variable reaction times on even simple tasks (Jensen, 1998). Computational simulations have shown that a network with more connections would show both of these reaction time effects. As discussed earlier, this is not the case for a network that simply has faster connections. Garlick argues, as with his explanation of brain size, that a more plastic brain would develop more connections. Thus, neural plasticity explains reaction time data better than speed of information processing (Garlick, 2002).

Lack of Theoretical Clarity: More or More Appropriate Plasticity

A major shortfall in Garlick's theory of intelligence is that he seems to be conflating two definitions of "effective plasticity": the frequency with which new connections are generated (*more* plasticity) and the appropriateness of those connections (*more appropriate* plasticity).

Garlick's explanation of brain size falls into the *more* plasticity camp. Garlick explains brain size in terms of variations in the number of connections in a network. Connections and the related support tissue add to the mass and volume of the brain (see above). To explain the number of connections in a network in terms of plasticity, Garlick's argument must be that a more effectively plastic network must simply produce more connections for the same environmental stimuli.

Garlick's explanation of neural efficiency appeals to a definition of effective plasticity as *more appropriate* plasticity. His argument is that "*better* connections… [in a neural network would result in] only the *appropriate relations* being activated" (Garlick, 2002 [my emphasis]). Therefore, a neural network that is connected "better" would use fewer cortical resources (be more efficient) when it processes information. But Garlick cannot mean that a network that has more neural connections is *prima facie* "connected better" and will exhibit "more appropriate relations". Therefore, it is not enough to say that a more effectively plastic network simply produces more connections. Instead, effective plasticity must involve the creation of more appropriate connections.

Garlick implicitly conflates *more* plasticity and *more appropriate* plasticity within his explanation of speed of information processing. Garlick explains reaction time data with an appeal to Anderson's simulations, which account for reaction time data by variance in the number of connections in a network (Anderson, 1994; Anderson & Donaldson, 1995). In this sense, effective plasticity only needs to be *more* plasticity. However, Garlick clarifies his argument for the explanatory power of plasticity over reaction time in the following terms: "a network with *stronger* and more *appropriate connections* would be able to process even relatively simple tasks faster and with less error" (Garlick, 2002 [my emphasis]). While one could equate "stronger" with "more" in this sentence, the same cannot be said for the term "more appropriate".

Regress in Garlick

This conflation between *more* and *more appropriate* plasticity may be the result of a homoncular fallacy in Garlick's model. The homuncular fallacy is to presuppose a process within cognition that is identical to the various features you are trying to explain (e.g. design, intelligence, vision etc.). For example, to explain vision by postulating a little man in the head who looks at a screen obviously leads to an infinite regress.

Garlick's model seems to work based on the presumption that a neural network that generates more connections will be more appropriately connected upon one condition: on balance, new connections in the network tend to be appropriate. Of course, new connections must tend to be appropriate in a human neural network. If not, humans would not, on balance, learn to respond appropriately to our environment, and would not be categorically intelligent. However, the assumption of connection appropriateness is equivalent to an assumption of intelligent design (indeed, this was our criticism of neural efficiency theory). Therefore, if Garlick's theory equates neural efficiency with more connections, it does not escape the regressive appeal to a network's appropriate design that is found in single factor task-oriented explanations of intelligence. In this case, Garlick would not, as I suggested, be any more consistent with a categorical definition of intelligence than previous theories.

More connections would actually generate a less efficient neural network without relevant constraints on encoding. For example, take the exploding variance problem. In a connectionist framework, information is encoded as a pattern across a neural network. As new information is encoded, new patterns would be overlaid onto the old patterns already encoded in the network. The more similar a new pattern is to an old pattern, the more the individual pieces of information in both patterns will be connected to each other. Therefore, as the number of similar patterns encoded in a network increases, so does the amount of neural resources required to activate any of these patterns (Metcalfe, 1993). Therefore, if intelligence was simply the result of more connections, then it should follow that intelligent people use more cortical resources, i.e. are less efficient, when they perform the same tasks.

This homoncular fallacy also seems to be present in Garlick's computational model on two counts: it relies on a generative learning algorithm and on a human who provides desired outputs. Garlick's computational model works as follows: the real output produced by each set of inputs is compared to the desired output. If the discrepancy between the two is beyond a certain threshold, an algorithm is used to modify the connection weights of the units in between the input and output, based on the magnitude of the error. Garlick uses the same connection weight modification algorithm for all trials, and only varies the "neural network learning rates" between models to simulate differences in plasticity (Garlick, 2002). What these learning rates represent is that, for a more "plastic" model, the weights will be altered more for each given error.

Higher neural network learning rates do increase performance (as predicted by Garlick's plasticity model) but only do so when filtered through two forms of intelligent design. A higher learning rate guarantees that the network will be modified more, which is only beneficial because of a generative learning algorithm that tends to lessen the difference between the network's actual output and desired output. Moreover, this algorithm is dependent on accurate measures of "desired outputs", provided by the programmer. Thus, the algorithm is not just given feedback on its behavior. It is supplied with a constrained version of the relevant behavioral features that it should emulate. Therefore, Garlick's model is parasitic upon a designer with (human) intelligence with respect to realizing the relevance of feedback on its behavior.

A charitable, yet plausible, reading of Garlick would say that he had appropriate plasticity (not just more plasticity) in mind when he laid out his thesis. After all, Garlick defines a more plastic brain as one that is "*more able to* adapt its connections to the environment" (2002 [my emphasis]). His focus on the "ability" rather than "rate" of plasticity seems to suggest a focus on appropriateness. Yet, while Garlick is relatively clear theoretically and computationally as to the nature of "more plasticity", he doesn't seem to specify the concept of appropriate plasticity. Without any such specification, his explanation of intelligence is viciously circular. Our belief is that "appropriate plasticity" can be explained in terms of relevance realization.

Appropriate Plasticity Through Relevance Realization

The relevance realization framework has the potential to provide non-recursive constraints to answer one of the most intractable and pervasive problems within current explanations of intelligent behavior and design in cognitive agents (Vervaeke, Lillicrap & Richards, forthcoming). We saw earlier that the ability of a cognitive agent to seamlessly realize relevant environmental information is critical if it is to solve problems and interact with the world. However, given that the cognitive agent is embedded in a dynamic environment, this agent's action is inseparable from its development. Thus, if a cognitive agent is not only to act, but act in a *relevant* way, then it must not only develop, but develop in *relevant* ways. Within a neural network, this is equivalent to "appropriate plasticity".

The relevance realization framework provides three basic constraints on processing in a neural network: cognitive scope, cognitive tempering and cognitive prioritization (Vervaeke et al. forthcoming). Although only the first two constraints deal explicitly with the development of a neural network, all three constraints are consistent with the highly interwoven relationship between the action of a cognitive agent and the development of its a neural network. Thus, the constraints of relevance realization provide an account, in general terms, of Garlick's "appropriate plasticity".

Cognitive scope involves the tendency for a neural network to develop general-purpose versus special-purpose machinery in order to interact with the world (Vervaeke et al., forthcoming). On the level of a cognitive agent, for example, a hunter tracking a young, male deer could use the experience to develop a generalized ability to track mammals, or a specialized ability to track this particular buck. Obviously, neither extreme is ideal. If one generalizes from a case, she risks ignoring relevant details that may be applicable to other, similar cases (i.e. tracking a deer as opposed to a rabbit). If one specializes from a case, she engages in an untenable lack of compression of her information and risks ignoring the analogies between this case and other similar cases. Vervaeke et al. suggest that opponent processes must be at play in the basic processing of a neural network, some that push new machinery towards specialization, and others that push it towards generalization (forthcoming). For cognitive scope, Vervaeke et al. give the engineering analogy of a strategy used to train neural networks. Networks are given performance metrics, which may be increased to induce specialization from learned information, and weight decay terms, which may be increased to induce generalization of information (forthcoming).

Cognitive tempering involves a balance between a neural network's tendencies to develop exploitative versus exploratory machinery. In other words, does a network stick to actions that it knows will have a high payoff now, or does it explore new actions that may result in higher payoff later (Vervaeke et al., forthcoming)? To look at the equivalent case in a cognitive agent, a foraging animal may stick to the environment and actions it is conditioned to know, or may explore new environments that may include more plentiful and easily accessible food. As with cognitive scope, neither extreme is satisfactory, and a moderate position is reached through opponent processes. For cognitive tempering, Vervaeke et al. give the engineering example of neural networks that couple temporal difference (TD) learning with an inhibition on return (IOR) trace. TD learning involves strengthening or weakening recent responses based on the presence of reinforcement or punishment. This tends to create exploitative-type machinery in an artificial neural network by reinforcing returning to known action patterns (Vervaeke et al. forthcoming; Sutton & Barto, 1998). IOR traces inhibit responses to stimuli that the system has recently interacted with, which moves the system toward exploratory behavior (Vervaeke, et al., forthcoming; IOR?). Thus, a system that couples TD learning with an IOR trace implements opponent processes to balance between exploitative and exploratory machinery and meet the constraint of cognitive tempering.

Zac/Irving Zac 6/3/09 4:56 PM
Comment: Cite?

Interaction between the opponent processes involved in cognitive scope and tempering should cause the neural network to encode new information that is relevant with respect to these constraints. In other words, both of these constraints are determinants of "appropriate plasticity". Because action and learning are interwoven for the relevance realization framework and all learning oriented theories of intelligence, these constraints on relevant encoding are central to categorical intelligence.

Cognitive prioritization pertains to whether a system attempts to avoid misses or mistakes in the face of information in the environment that is ambiguous due to factors such as signal noise or multiple potential causes for the same stimuli. Thus, cognitive prioritization is more directly related to action than either cognitive scope or tempering. At any given time, a system may either diversify or focus the range of representations and actions it employs in its interactions with the environment. Diversification will tend to avoid missing relevant information in the environment, whereas focus will tend to avoid mistaking irrelevant information as relevant information. Thus, a system must be able to flexibly gamble between diversification and focusing, depending on which strategy is best to solve the problem(s) at hand. For example, if an animal is very hungry, it will tend to focus on food acquisition at the cost of many other potentially relevant actions (e.g. predator-avoidance, mate-winning, shelter-building). All other factors being equal, if an animal is sated, it is free to pursue a wider range of potentially relevant actions (Vervaeke et al.). Although cognitive prioritization directly involves action, since the signals that a cognitive agent attends to or ignores will shape the information that it encodes, then cognitive prioritization is also critical to learning.

Thus, taken together, the three constraints of relevance realization provide a strong account of "appropriate plasticity".

First Order Relevance Realization and the Modular Approach

An important question is whether relevance realization emerges at the level of neuronal interaction or at the level of modular interaction. An implicit strategy in the psychological literature has been to suggest that relevance realization emerges solely from the interaction between higher order modules. Later in this article, we will show how van der Maas (2007) implicitly takes this approach when he claims that "intelligence is based on underlying cognitive processes, such as perceptual, memory, decision and reasoning processes". While we believe that interactions between modules may be *part* of the picture, we will provide a plausibility argument for why First Order RR (RR1) – in other words, RR emergent from the basic properties of neurons – is necessary for a brain to be self-organizing.

A particularly pervasive example of the modularity approach within the literature is memory compartmentalization. This particular approach suggests that you can deal with issues of relevance realization by organizing memory into compartments so as to vastly constrain the search space (for a review, see Cherniak, 1986). This is a radically insufficient solution to the relevance realization problem. The first problem with this strategy is that relevance realization is pre-supposed in the construction of said compartments. Typically, memory compartments are thought to hold sets of information that are relevant to each other or, in other words, belong together in some relevant fashion (Chiappe and Vervaeke, 1997). Secondly, relevance realization is presupposed in the size and scale of the compartments. I can have fewer compartments that are larger, but that tends to make the search through any compartment difficult. I can have more compartments that are smaller, but then I will need to use information from multiple compartments, and will face combinatorial explosion when I search the many possible combinations of information from multiple compartments. Furthermore, there is a problem of scale, because most memory is organized into multiple levels of abstraction. Thus, relevance realization is presupposed in the selection of the appropriate level of abstraction when the information is used. For example, when I access the "car compartment" do I need to access "automobile, car, Ford, Steve McQueen's 1969 Ford Mustang"?

Relevance realization is not only presupposed in the search of memory compartments, it is presupposed in the use of these compartments in non-stereotypical situations (Chiappe and Vervaeke, 1997). For example, if you enter a restaurant and you are served rubber chicken or your waiter pulls out a gun, you can deal with those situations even if they're not part of a pre-fabricated script. Haughland gives the example of a person who comes into your house and puts her raincoat in your bathtub so as not to get your floors wet. This information is not contained in the concepts of raincoat or bathtub; nor can it be derived from a simple combination of these concepts. Hence, any theory of intelligence based on compartments presupposes intelligence.

We believe that similar problems exist in all formulations of the module approach. One may claim, as van der Maas does, that intelligence emerges because of the mutually beneficial interaction between different cognitive modules (for instance, short term memory helps to develop better cognitive strategies, which in turn increase short term memory (Siegler & Alibali, 2005)). However, a great deal of relevance realization, and thus intelligence, was necessary in the construction, access and application of each of those modules, not to mention the interaction between them. To explain any of these features of cognitive modules, we need an account of what we will call First Order Relevance Realization (RR1), in other words, relevance realization that is immanent to processing at the neuronal level (Vervaeke et al., forthcoming). We also think the fact that all three constraints of relevance realization can be implemented in simple algorithms in a computational neural network strengthens our evidence for the plausibility of RR1 (Vervaeke et. al, forthcoming). This is especially case considering some aspects of the constraints, such as TD learning and IOR trace have known correlates in human and animal neuronal systems. (For an example of TD learning, see Ainslie's work on hyperbolic discounting of more temporally distant rewards (2005); for a review of IOR in the human visual system, see Klein, 2000 and for an example of evidence for IOR enhancing exploratory behavior see Klein & MacInnes, 1999).

However, this is not to claim that the presence of RR1 is sufficient to label a cognitive agent categorically intelligent, as we believe intelligent agents apply relevance realization to higher order modules, not simply lower order processing. Many categorically stupid animals possess some degree of RR1. For example, lizards possess the capacity for classical and operant conditioning, but do not display cognitive flexibility, or in Harlow's terminology "learning to learn" (1949). We think that Mercado provides compelling evidence that development in intelligence across species is correlated with an increasing number of cortical modules (2008). This is consistent with a basic feature of the RR framework: the constraints of RR are applied recursively to many levels of processing (note that this is not simple recursion, but instead recursion applied on higher and lower levels). In other words, the same constraints apply to the basic interaction between neurons (RR1) as apply to the interaction between cortical modules (RR2). Later in this paper, we will suggest that Mercado (2008) and Hawkins (2001), when considered together, provide a strong argument for the existence of RR2.

Our overarching proposal is that g can be filled with recursive RR. We have demonstrated the efficacy of RR1 as an explanation of Garlick's "appropriate plasticity", and believe that Garlick provides a strong argument for why "appropriate plasticity" can fill the explanatory placeholder g. Next, we will look at theories of categorical intelligence that involve more than one factor and how a lack of theoretical clarity in van der Maas' Mutualism Model (2007) can be clarified by appeal to RR2.

A Multiple Latent Factor Theory and Mutualism

Sampling: An Early Theory

Sampling theory claims that the positive manifold does not arise because of any unitary psychological g factor, but instead because of a measurement problem: intelligence-related tasks draw upon many uncorrelated lower order processes, and these processes are involved in many different tasks. Therefore, performance on any intelligence-related task arises from a set of underlying processes that subserve many other intelligence-related tasks. Thus, performance on one intellectual task should predict performance on other intelligence-related tasks.

While sampling theory represents a coherent escape from the necessity of a single factor explanation of intelligence, it runs into three strong empirical critiques from Jensen (1998) and Eysenck (1987). First, some very narrowly defined intelligence tasks (which should draw upon only a few processes) heavily load on the g factor (in other words, highly correlate with other processes) whereas sampling predicts that this should only be true of complex intelligence tasks. Second, seemingly unrelated intellectual tasks – such as visual and memory span tasks – that should not draw upon the same cognitive processes, are highly correlated whereas some seemingly related intellectual tasks – like backward and forward digit span – are uncorrelated. Third, brain damage should, under sampling theory, lead to general impairments in cognitive processing, when in some cases, damage leads instead to very specific impairments. We will analyze these critiques, and the mutualism model's ability to escape them, later in our paper.

Another criticism of sampling theory, which should be familiar from above, is that the processes it claims are responsible for psychological intelligence have very little to do with the learning-oriented nature of intelligence as defined philosophically. In sampling theory, lower order processes subserve a person's ability to *perform* cognitive tasks. Like single factor theories, psychological intelligence in sampling theory is something that varies within, rather than helps to design an intelligent neural network.

Mutualism

Van der Maas' mutualism argues that multiple underlying psychological factors of intelligence could, through mutual developmental relationships, generate a positive manifold. The first premise of the argument for the mutualism model is that "intelligence is based on underlying cognitive processes, such as perceptual, memory, decision and reasoning processes" (van der Maas et al, 2006). The second, implicit premise of the mutualism model is that these underlying cognitive processes develop over the lifespan. This second premise shows that mutualism, unlike most of its predecessors including sampling theory, engages with the necessarily dynamic nature of intelligence from a philosophical perspective.

The third premise of the mutualism model – its *mutualism* premise – is the most important and controversial aspects of van der Maas' theory: the cognitive processes that underlie intelligence tend to have beneficial mutual relationships toward each other's development. In other words, as each of these processes develops, it supports or facilitates the beneficial development of other processes, whether directly or through intervening variables (van der Maas et al, 2006). Later in this paper, we will question the assumption of "beneficial" as opposed to simply causal interactions between processes. Regardless, the Mutualism model is in line with the notion, espoused earlier, that a dynamical neural network does not just learn, but learns how to learn. On one hand, intellectual development facilitates performance on a variety of tasks. On the other hand, intellectual development also facilitates the development of a variety of processes, each of which in turn facilitate further performance and development. Thus, a network of intellectual processes is, according to the mutualism model, extremely self-organizing and bootstraps its own development.

Van der Maas argues that the positive manifold can be explained, through mutualism, not as the result of (a) latent psychological factor(s) involved in intelligence-related tasks, but as a statistical correlation between cognitive processes that arises through development. If one has above average ability at any given intelligence-related process, this process should tend to have a greater than average beneficial effect on the development of all other processes. Therefore, after any period in which intellectual processes develop (e.g. development before IQ testing), one's ability at each intelligence-related process should tend to correlate with one's ability at most other such processes. Thus, performance on any intelligence-related task should tend to predict one's performance on other intelligence-related tasks, which involve different, but developmentally related processes. Therefore, the positive manifold arises without having to suggest that the same processes underlie many different intelligence-related tasks.

Van der Maas supports his argument by using a mathematical analogy to show that processes that mutually benefit each other's development will generate the positive manifold. Van der Maas builds his mathematical "model" (see the following section for a discussion of whether van der Maas' mutualism should be considered a model) like a predator-prey population matrix used to map ecosystems (as cited in van der Maas et al: May, 1973; Murray, 2002). In this matrix, processes develop in a way that is analogous to the growth of populations of organisms. These processes are initially uncorrelated. They also have uncorrelated growth parameters, which represents that there is no single factor behind intelligence in the mutualism model. However, these processes tend to mutually benefit each other's development (this amounts to taking predation out of the population matrix). Given these conditions, which are analogous to the assumptions of mutualism, a positive manifold arises. In fact, the "mutualism" matrix has similar results (i.e. a similar positive manifold) to a model of the "single factor" matrix (van der Maas et al, 2006).

Mutualism: Model, Argument or Framework?

To his credit, van der Maas seems to realize the limitation of the mutualism model, since he explicitly refuses to clarify what he means by the "underlying cognitive processes" of intelligence, citing a lack of consensus as to what these processes are (as cited: Deary, 2002, p.153). Even the term "cognitive process" is vague, for almost everything that humans do is based in cognition in some sense, develops over the lifespan, and even potentially has mutually facilitating relationships with other processes. Thus, mutualism is not a "model" in the sense that it specifies the processes, or even types of processes that generate intelligence. Compare this to speed of information processing or the *more* plasticity thesis. Even though neither of these models is entirely effective, both of them attempt to provide clarification as to the mechanisms that create intelligent behavior. This is important, because any model of intelligence must attempt to answer a question that is inherently grounded in the processing of the brain: what *is* intelligence?

A useful heuristic to analyze whether a theory of intelligence is a model, then, is to look at whether this theory can be implemented in a computational neural network, our best approximation of the brain. Speed of information processing, brain size, and extent of plasticity theory are conceptually clear enough that each of them have been mapped onto computational analogies (speed of information processing: Anderson, 1994;

Anderson & Donaldson, 1995; brain size: Sejnowski & Rosenberg, 1986; extent of plasticity: Garlick, 2002). Van der Maas' mathematical model of mutualism is, tellingly, based on an ecosystems metaphor, a very general level of ecology. One cannot take a mathematical predator-prey population model and glean anything about how the body of a wolf works. Rather, a population model simply supplies statistics on the growth of and relationships between populations of animals. The one restriction on cognitive processes in mutualism – that these processes must have mutually beneficial developmental relations – can be represented perfectly in a modified population model (van der Maas et al, 2006). However, in the same way as an ecosystems population model does not tell you details about the animals within it, a mutualism model cannot tell you details about cognitive processes. Therefore, even if the mutualism model were provided with statistics of the development of and developmental relationships between cognitive processes (which it is not) it would tell you almost nothing about the processes themselves. While the RR framework is also a higher-level theory than, say, speed of information processing, it can nonetheless be directly specified within a computational neural network (Vervaeke et. al, forthcoming), as we have seen above.

This is not to say that one could not create a model of intelligence – computational or otherwise – that incorporates mutualism; rather, simply argue that van der Maas' mutualism model does little theoretical work to clarify what process(es) would account for the mutualism in such a model. A useful analogy can be drawn with Garlick's *more appropriate* plasticity thesis. It is of course the case that new connections must tend to be appropriate in an intelligent brain. To avoid circularity, however, it is necessary to specify the design processes that generate appropriate connections. Similarly, it is true that cognitive processes must tend to benefit each other's development in a self-organizing brain. However, simply evoking the fact that some degree of mutual bootstrapping is critical to the dynamical systems approach does no work to develop a model of how mutualism is implemented in the brain.

One reasonable interpretation would argue that van der Maas intends for mutualism to be taken as an argument to be used by future models rather than a model in and of itself. Mutualism shows effectively that if unrelated processes tend to have mutually facilitating developmental relationships, then a positive manifold will arise. Therefore, any multiple component model of intelligence that shows developmental mutualism between the processes underlying intelligence can, using the mutualism argument, explain the positive manifold. Considering that the positive manifold is arguably the single most important and established effect relating to intelligence, this is a significant development. Yet mutualism remains, in this view, only an argument that can be used within a functional model of intelligence.

Another interpretation, not mutually exclusive with the last, is that mutualism is a theoretical framework, a conceptual space in which future models can be built. In this view, mutualism is a meta-model that argues about the kinds of models that should be built to explain intelligence. This view is supported by the fact that mutualism gives between two and three constraints on the processes involved in future models of intelligence. First, the cognitive processes that any theory of intelligence refers to must be directly involved in the performance of IQ tests. This constraint exists because, without it, the term “cognitive process” is so vague that literally every aspect of brain functioning is involved in intelligence. Note that this constraint is dangerously circular without clarification. We appeal to RR, applied on the level of neural and cortical module connections, as a process that should be related to intelligence in principle (and not just as a matter of fact). Second, the cognitive processes behind intelligence must have mutually beneficial effects on each other’s development.

A third constraint of the mutualism framework seems to be as follows: cognitive processes must develop over time. For most cognitive processes, this should be the case. However, we have argues in our section on RR1 that there must be some pervasive constraints on the brain's processing that tend to realize relevance and subserve plasticity on the level of neurons and their connections and on which other processes are asymmetrically dependent. Otherwise, there is no way that any part of the brain could solve the frame problem. These first order constraints would certainly *affect* plasticity, but would not necessarily be *subject to* plasticity themselves. Nevertheless, given the global importance of RR1 for plasticity, well-functioning RR1 constraints should correlate strongly with most other processes that are subject to plasticity.

We believe that mutualism should be viewed as a theoretical framework that puts various constraints on future models of intelligence and provides a valuable argument to explain the positive manifold, and many other empirical results, for any model that can meet these constraints[5]. It is important to understand that the mutualism framework is not a freestanding model. The mutualism framework is reliant on some model of intelligence that can meet its constraints.

Clarified Mutualism through Relevance Realization

While van der Maas can neglect to specify the processes involved in his framework without undermining the framework itself, one premise upon which he is reliant is that mutualism between intelligence-related processes is plausible. Furthermore, this is the most important constraint of the mutualism framework. Thus, a clarification of the mutualism premise helps to examine the plausibility of and to develop the mutualism framework. These are both important steps in order to guide future work within the mutualism framework.

In fact, without clarification, the mutualism framework simply shifts, and does not avoid, two important problems: Thorndike (1927) and Thompson's (1951) critiques of sampling theory (see above) and an assumption of intelligent design. It should be noted that our work here is theoretical, meant to develop the mutualism framework rather than a model within this framework. Thus, we do not provide definitive solutions to these problems. Rather, we provide clarified constraints that mutualism models must meet in order to hope to avoid these problems.

The constraints on mutualism models that we provide are as follows: 1) mutualistic relationships between a mutualism model's processes need to be (relatively) contained within a causal nexus; 2) they need to demonstrate beneficial mutualism, not just causal mutualism.

The Pervasive Mutualism Problem

One interpretation of van der Maas' argument for mutualism is that all cognitive processes tend to have mutually beneficial relationships to each other, so it is plausible that the cognitive processes behind intelligence display mutualism as well. Indeed, one of his arguments for the mutualism premise is to give a veritable laundry list of "reciprocal causal relationships [that] are well known in the psychological literature" (van der Maas et al, 2006). These "reciprocal causal relationships" are between pairs of cognitive processes that span from language and cognition, to short term memory (STM) and cognitive strategies, to intellectual abilities and emotional abilities. Van der Maas cites six such pairs in all (2006). All of these relationships could be framed as mutually beneficial developmental relationships. Therefore, the argument seems to be that mutualism is a pervasive property of human cognition and likely exists within the realm of intellectual cognition as well.

The more general form of this argument is that mutualism is consistent with "views of dynamical systems" (van der Maas et al, 2006). One of the most crucial properties of dynamical systems is that they must be self organizing. One of the ways that self-organization could be implemented in a cognitive system is through mutualism. If all cognitive processes, on balance, were to benefit or even facilitate each other's development, then perhaps there is no need for a unitary intelligent designer within a cognitive system. Instead, the development of each of the processes contributes to the design of the system through its interaction with other processes. Thus, armed with only individual processes as opposed to an intelligent designer, the cognitive system is organized.

The greatest problem with this line of argumentation – that mutualism is pervasive throughout the cognitive system – is that it does not allow for a point at which mutualism stops. If all cognitive processes tend to display mutualism with each other, then why don't we see strong correlations between all tasks that are supported by cognitive processes? Implicitly, the positive manifold itself is only interesting because intelligence-related tasks tend to be highly correlated *in comparison* with other tasks. For example, motor skills, communication skills and emotional control are just some of the many cognitive processes that are obviously involved in sports. If all cognitive processes tended to be correlated, then general athletic ability should predict intellectual ability (and all the other phenomena predicted by g) to the same extent as fluid cognitive skill. If one questions the measurability But g is important precisely because one cannot just look at athletic performance (or just any given task) to predict performance on most other cognitive tasks.

It is critical, therefore, that any mutualism model should delineate the nexus of mutualistic (and even inhibitory) relationships between processes in the cognitive architecture. Indeed, the ability of the mutualism framework to avoid the second critique of sampling theory is reliant on specifications of and limitations on the mutualistic relationships between processes. One half of this critique is that seemingly unrelated intellectual tasks – such as visual and memory span tasks – that should not draw upon the same cognitive processes are highly correlated (Thorndike, 1927; Thompson, 1951). Especially since very different processes are likely behind these tasks, the nexus of mutualistic relationships between these processes needs to be clarified. However, one must be careful not to extend this nexus without bounds in order to avoid pervasive mutualism. Another reason why the nexus must be limited is that success on seemingly related intelligence tasks – like backward and forward digit span – is uncorrelated (Thorndike, 1927; Thompson, 1951). Thus, one must not just limit the broad types of processes that are correlated (i.e. perceptual, memory) but clarify why there is no correlation between processes that would seem, in the very least, to have mutually beneficial relationships. One possibility, which should generate this result, is that such processes directly inhibit each other's development. If this were the case, these inhibitory relationships would directly negate the correlation between two processes while maintaining mutualism between these and other processes. This forces any "mutualism model" to include inhibitory, as opposed to simply mutualistic, relationships, and also to consider what could give rise to such relationships.

Extended Mind Analogy

Through the analogy of extend mind theory, it can be seen that drawing the boundaries to this nexus of mutualistic relationships will be difficult. In extended mind theory, the concept of mind leaks out of the brain, into the body, and even into the tools that being used by the cognitive agent. For example, Andy Clarke suggests that a paper and pencil can be a part of the mind, because these tools can represent an essential part of a cognitive agent's ability to perform certain tasks (2007). The paper and pencil are not only being used by various aspects of the brain and body, but constrain the mind's processing in turn. For example, notes can focus the mind on relevant information and force the mind to ignore, to a certain extent, irrelevant information. In this sense, a paper and pencil can constrain thought in a similar way to working memory.

Extended mind theory, like pervasive mutualism, threatens to arise because no cognitive process, part of the body or even tool is independent from causal influences and effects. Extended mind theory is essentially a radical extension result of computational functionalism, where the brain changes depending on its function. Extended mind theory says that the mind can leak even into the non-human objects involved in the performance of a function. However, unless one constrains the mind to the physical body of a person, then we will encounter the frame problem when we try to find the boundaries of the mind. Should we consider light to be part of the mind when we see, for example, that more light expands our processing capacity and feeds back into the brain? Even if we do not believe in determinism (which would mean that the entire universe is generating our processing) there are an almost limitless number of inputs that can directly feed back upon our experiences and actions at any given time; thus, the extended mind should leak almost without bound. Similarly, with the mutualism model, every cognitive process will likely influence, directly or indirectly, the development of every other process. Therefore, the boundaries of the nexus of mutualistic relationships threaten, like an extended mind, to leak outward unbounded.

Constraints Through Causal Nexuses

We can constrain the extended mind by appealing to a nexus of causal relevance. We can think of the dynamic, functionally defined mind as consisting of a nexus of parts that influence and are influenced by each other. From this metaphor, whenever the brain is engaged in a function, there will be some parts that are closer to the center of the nexus, in other words, influencing and being influenced by the environment and other active parts of the cognitive agent. The various cognitive processes of the brain are extraordinarily interwoven and are therefore more likely going to be toward the center of the nexus than the inanimate parts of "the mind" at any given time. While there is no hard line, as something moves farther out from the center of the nexus, we should be less willing to call it a part of the mind.

Similarly, we can constrain our nexus of cognitive processes by functional and developmental causal relevance. The functional nexus would be defined by the extent to which, during the performance of an intelligence-related task, a cognitive process is feeding into and receiving feedback from the environment and other active processes. The developmental nexus would be defined by the extent to which a cognitive process has mutually beneficial developmental relationships (direct or indirect) with other cognitive processes that are relatively close to the center of the functional nexus. If a cognitive process was peripheral on both an aggregate of the functional nexuses of IQ tasks and on the developmental nexus, it should not be considered to underlie intelligence. One potential setback of the mutualism model is that centrality on the developmental and aggregate functional nexuses should have an indistinguishable marker from a psychometric perspective: high g.

One benefit of this last point is that it provides a potential explanation of the first critique of sampling theory. Even if a narrowly defined cognitive process was peripheral on the aggregate functional nexus, if it was central on the developmental nexus, that process should have a high g loading. However, a careful examination of these narrowly defined, high g processes is in order to see if it is plausible that they show strong mutual effects on the development of other processes.

Causal as Distinct from Beneficial Mutualism

The mutualism assumption is not simply that cognitive processes will have mutually *causal* developmental relations, but instead that they will have "mutual *beneficial or facilitating* relations" (van der Maas et al, 2006) in development. Causal mutualism is an obvious, entirely uncontroversial premise. It is clear that cognitive processes must interact to perform any cognitive task. From Hebb's basic insight, "neuron's that fire together wire together" (1949), it is also clear that each cognitive process, implemented through patterns of neural connections, must develop through use. Therefore, cognitive processes interact to change the way that other processes are used, and subsequently develop. However, causal mutualism shows only that cognitive processes interact through development, not that this interaction is beneficial. This is not enough for the mutualism framework, as causal interaction could just as easily be inhibitory as it could be beneficial in a neural network.

An argument that seems to be implicit in van der Maas' appeal to the consistency between mutualism and dynamical systems is that beneficial mutualism between cognitive processes is necessary in order for a cognitive system to be self-organizing. If new cognitive development tended to be neutral or inhibitory toward further development, then humans would never develop intelligence in the first place.

One can draw an analogy between this argument and the central assumption in Garlick's *more plasticity* thesis, explained above, which explains g in terms of variations in a networks capacity to make new connections (2002). More connections should indeed generate higher g, in so far as one assumes that new connections in the network tend to be appropriate (the network should, on balance, be capable of learning if the new connections are not appropriate). Of course, if a network is self-organizing in the sense that it makes appropriate connections, or that its processes tend to appropriately bootstrap each other's development, then that network will be intelligent. This last statement amounts to the following: if a network has the capacity to *intelligently design* itself, then it will be intelligent. Thus, neither the more appropriate plasticity nor the beneficial plasticity theses escape the assumption of intelligent design that the dynamical systems approach seeks to avoid. Both Garlick and van der Maas' theses simply shift the assumption of intelligent design onto an "appropriate" or "beneficially" developing cognitive architecture. Clarification of the way in which plasticity is "appropriate" and what Mutualism is "beneficial" is critical if these theses are to carry explanatory weight

Better Mutualism through Relevance Realization

We argue that relevance realization is at the center of the functional and developmental causal nexuses of mutualism. In the functional nexus, we have argued that almost all tasks related to intelligence have a problem of relevance realization at their core. Therefore, the constraints of relevance realization will be causally related to performance on these tasks. Moreover, since relevance realization is recursive in the cognitive architecture, many different opponent processes will subserve the same constraints, and will even influence the extent to which other opponent processes are implemented (Vervaeke et al., forthcoming). For example, a rational strategy might help implement an opponent process on the level of cortical modules and vice versa. Furthermore, any relevance realization process is likely in opponent process type relationships with many other processes. Therefore, any relevance realization process will be pushed toward many different equilibriums and will have a distributed effect through the functional nexus. In the developmental nexus, the strong causal effect that relevance realization processes have on each other functionally will extend to development. Remember that, in the dynamic system of relevance realization, functionality (behavior) strongly causes and is caused by development. Therefore, the immediate effects that each processes has on the functioning of other processes will be both reflective of prior development that related these processes, and will result in future development that will strengthen the causal relationships of these processes.

Therefore, the various interrelating opponent processes of relevance realization will have strong causal relevance for the functional implementation and the behavioral development of intelligence. Based on our argument in III.1, these processes should then be near the center of the functional and developmental nexuses of intelligence, and should have large explanatory power over g.

One downside of a relevance realization explanation of beneficial mutualism is that success on relevance realization is defined by the maintenance of equilibriums in support of broad constraints, and therefore it becomes difficult to define good performance on an individual process scale. However, measurement problems will arise in any model of intelligence in terms of dynamic functionally and developmentally related processes (this includes van der Maas' mutualism). As we pointed out in III.1, the mutualism model cannot separate the psychometric effects of centrality on the functional and developmental nexuses. Moreover, the inability to measure individual processes may not be as large of a problem as it seems, because most intelligence tasks likely measure opponent processes. Indeed, one critique of his mutualism model that van der Maas points out is that few tasks measure individual processes. Importantly, a relevance realization account of mutualism does not, like sampling theory, require that only broadly defined tasks display high g loadings, because the function of one opponent process (plausibly contained, even within a specific task) might have broad effects.

Higher Order Relevance Realization from Mercado and Hawkins

A Critique of the "Complexity Thesis"

The beneficial mutualism premise shares an implicit central assumption with recent work by Mercado: beneficial development amounts to an increase in the complexity of cognitive processes and representations. In Mercado's expansive, cross species analysis of cognitive plasticity,

> the brain's capacity to differentiate representations is specified as the primary determinant of which cognitive skills an organism can learn as well as how long it will take for an organism to learn a particular cognitive skill (2008).

This capacity for complexity generates differences in cognitive plasticity between species, across the lifespan and between individuals (Mercado, 2008). In other words, if a cognitive agent has a greater capacity to sustain a complex network of differentiated representations of the environment, that organism will be more intelligent.

One possible explanation of beneficial mutualism, then, is to propose that, as cognitive processes develop, they tend to make a cognitive agent's representations more complex. A very simple example is short term memory (STM): as STM capacity increases over childhood, a cognitive agent can hold a greater number of simple representations of the world at any given time. Also, the development of language allows a cognitive agent to increase the conceptual complexity of the stimulus representations that they currently hold. Development in the vast majority of cognitive processes can be interpreted, in some way, as an increase in the complexity of stimulus representations. If we believe Mercado's argument, this development in complexity will benefit the development of new and existing cognitive processes and, thus, explain beneficial mutualism.

Unfortunately, Mercado's definition of cognitive plasticity and our subsequent explanation of beneficial mutualism both expose a developing cognitive agent to a strong version of the frame problem. Vervaeke et al. argue that a fundamental question behind most theories of intelligence is how the brain constrains potentially boundless problems into a representation that contains only a (relatively) small number of relevant features (forthcoming). According to this argument, intelligence must, contrary to Mercado, involve the ability to reduce the representation complexity of any problem into its relevant features.

The need for intelligence to be involved in the constraint of representation complexity exists because the vast majority of problems that a human solves – for example, walking, talking, writing an essay or making lunch – are ill-defined. In other words, there are a vast number of potential ways to represent any such problem (i.e. whether one should eat a sandwich, whether food is poisoned, undercooked, or whether using the stove will blow up the house), most of which must be dismissed most of the time. Essentially, a cognitive agent must be able to focus on the central features of the functional cognitive nexus for any given problem without even considering any of the peripheral features. The brain simply does not have the processing capacity to consider all the potential representations for any given problem (there are about as many ways to play a chess game as there are electrons in the known universe). Clearly, an account of beneficial mutualism, and thus intelligence, in terms of increases in as opposed to constraints on representation complexity is incomplete.

The Memory-Prediction Framework: A Contrast to Mercado (with problems)

The *memory-prediction framework* of intelligence given by Hawkins et al (2004) is the stark opposite of Mercado's account for intelligence. The central thesis of this framework is that intelligence is based on a cognitive agent's capacity to make predictions about the environment based on invariant representations stored in memory. This theory, from a number of perspectives, credits intelligence to a reduction, not an increase, in representation complexity. In Mercado's model, a more intelligent brain was able to represent the same object, situation or problem in the environment from various different perspectives (2008). In Hawkins' model, a more intelligent brain notices the invariant features of any stimulus that it encounters in order to compress its memory of that stimulus into one invariant, generalized representation (Hawkins)[6]. In short, Hawkins' intelligence involves collapsing the distinctions between perspectives to form unitary, invariant representations. Another example is that Mercado stresses that evolution has provided more modalities through which to represent the same stimulus through the increase in cortical modules, which perform different functions in the brain. Hawkins, to the contrary, stresses that evolution has created a more developed neocortex, the greatest impact of which is to coordinate many disparate modalities (i.e. vision, hearing, motor etc.) into unitary representations (2004).

Hawkins' attempts handle the frame problem through his appeal to prediction. The brain, according to Hawkins, uses the invariant representations it stores in memory to make predictions about current stimuli based on what similar stimuli have done in the past. Therefore, the brain needs only to check if current stimuli fit its predictions in a general way, not analyze each individual feature of the environment. If something in the environment is different than predicted, then the brain attends to that anomalous feature (Hawkins, 2004).

Though Hawkins recognizes and attempts to circumvent the frame problem, he does not avoid it nor recognize the centrality of the brain's capacity to realize relevance to his account. Hawkins' approach is not new. In 1990, Kaplan and Simon appealed to the *notice invariants heuristic* as a way to form accurate stimulus representations and apply these representations to current problems. Moreover, many have suggested that the ability to make analogies between past and present stimuli and problems – which is implicitly at the center of Hawkins' memory-prediction – is critical to intelligent behavior (Polya, 1973; Getner, 1983; Holyoak, 1982).

However, the problem of each of these theories (including Hawkins') is that making an analogy is itself an ill-defined problem. As Goodman famously noted, there are an indefinitely large number of similar and different features between any two stimuli (1972). For example, think of all the similarities between a wooden chair and a buffalo (both have four legs, both are made of organic material, both are brown, you can sit on both, neither is a particularly good weapon). This list is indefinitely long. All of these comparisons are true, yet you will naturally find some more relevant than others. This is our point. The similarity is based on the relevant comparisons and is therefore completely dependent on the capacity to realize relevance. To make an analogy between any invariant representation and current stimuli one must first recall the relevant representation out of a vast bank of potential representations (i.e. why don't you apply the representation of a buffalo, or even a bookshelf, to a chair you are currently viewing). Then, one must pick out the relevant similarities between said representation and the current stimuli. Both of these processes require that the brain is able to realize relevance if we are to avoid combinatorial explosion.

Hawkins' response would likely be to say that the "relevant similarities" are what is invariant through different representations of the same thing. But each of these representations is a different stimuli, and thus, similar to the problem of analogy, there are an almost infinite number of similarities (invariants) and differences (variants) between different representations of the same thing. For example, there is much more invariance between two different people's faces from the front than from the same person's face from the front and side. Invariance detection, like analogy, requires the ability to pick out relevant similarities between two stimuli without considering all other possible features of comparison.

There is a deeper problem of trying to equate relevance realization with invariance detection, which has to do with the formation of categories. As Wittgenstein famously noted, for very many categories there is no invariant set of features shared by all the members of the category. Learning how to interact with games (to play them, identify them, talk about them etc.) does not involve learning the invariant features possessed by all games. This derives from the fact that games are related, as Wittgenstein said, by "family resemblance", and not by sharing a Platonic essence (1958). Similarly, Rosh was directly influenced by Wittgenstein to argue that concepts are not mental definitions (1978). Note that, in order to structure concepts by family resemblance, one must be able to integrate the invariance detection that explains overlaps between subgroups (e.g. board games or ball games as a subset of games) with the differentiation that actually separates class into subgroups that do not possess any shared essence. We believe that relevance realization can account for the integration of invariance detection and differentiation.

Mercado and Hawkins as Relevance Realization 2

Perhaps the most interesting feature of Mercado and Hawkins is not that either represents a stand-alone theory, but that they both present plausible, yet opposing functions of the brain. Simply, they talk about the importance to intelligence of, respectively, complexity and simplicity of cognitive representations. Thus, we must either reject both theories, or argue for a framework that sustains their seeming contradictions. We believe that the RR framework does just that.

Before we discuss how Mercado and Hawkins fit into the RR framework, it is important that we clarify RR2. In this respect, it is helpful to draw analogies between RR and the role of fitness in evolution and market forces (Vervaeke et al., forthcoming). There is no particular characteristic (i.e. running speed, size, intelligence etc.) that defines fitness from an evolutionary perspective. Rather, fitness is defined as the ability to interact with the environment in such a way as to survive (forthcoming). Another analogy would be market forces in an economy, none of which benefit the economy independently. For example, an increase in demand only results in economic growth if met with an appropriate increase in supply. If demand (for goods in general) were to continue to stay high, while supply remained constant, the economy would be faced with a period of little growth and inflation. Economies function, at least in part, because the opponent processes of supply and demand tend to move each other to a state of relative equilibrium. Ideally, whenever either force becomes too high or low, the other will tend to move it toward a middle point.

To draw out the analogy, relevance realization is not defined by the functioning of any single process such as novelty monitoring or invariance detection (forthcoming). Rather, the benefit of cognitive processes for relevance realization is defined by how they interact with other opponent processes to generate equilibrium with respect to cognitive scope, cognitive tempering and cognitive prioritization. Mercado and Hawkins, with their opposing theories of intelligence, provide brain-based plausibility to the presence of the implementation of relevance realization on the level of cognitive processes (RR2).

Cognitive scope represents the extent to which the cognitive architecture is designed toward special purpose or general-purpose machinery (Vervaeke et al., forthcoming). Mercado argues that intelligence involves the differentiation of cortical modules into special purpose machines. Hawkins argues that intelligence involves the integration, through the neocortex, or many cortical modules into one general-purpose machine.

Cognitive tempering represents the extent to which the cognitive architecture is exploratory – tending to look for new, better ways of solving problems – or exploitative – tending to apply what is currently known to new problems (Vervaeke et al., forthcoming). Mercado's theory implies that intelligence involves the development of new ways to represent stimuli, in order for cognition to simultaneously explore various, differentiated representations of the environment. Hawkins theory implies that intelligence, for the most part, involves exploitation of invariant representations stored in memory.

Cognitive prioritization pertains to whether a cognitive agent diversifies or focuses its representations of the world in order to, respectively, avoid misses or mistakes. This allows it to select context-appropriate strategies to deal with ambiguous information. Mercado and Hawkins seem to suggest that intelligence involves, respectively, diversification and focusing of representations. Mercado's intelligence involves diversification of a cognitive agent's representations in order to miss less potentially relevant perspectives. Hawkins intelligence involves, primarily, dismissal of as much irrelevant environmental information as possible by viewing the environment as a series of compressed (i.e. focused), generalized and invariant representations. Cognitive prioritization suggests that a cognitive agent must be able to flexibly move between such diverse and focused representational stances, depending on environmental constraints such as satiety. Hawkins and Mercado, taken together, provide brain-based evidence that categorical intelligence involves both opponent processes of cognitive prioritization. We believe that this makes it plausible that a categorically intelligent brain does flexibly move between these opponent processes.

Through Mercado and Hawkins, it seems as if the evolution of cognitive agents into species with greater categorical intelligence involves a simultaneous development of both opponent processes in each constraint on RR. Thus, they provide brain-based evidence not only for the existence of RR2 (relevance realization as implemented on the level of cognitive processes), but also that RR2 is deeply involved with categorical intelligence.

Conclusion

We believe that the central problem to prior explanations of intelligence is the presupposition of a set of intelligent processes within a system that designs the system in such a way as to make it intelligent. This leads to a viciously circular explanation, which presupposes the very phenomenon to be explained. The strength of Dennis Garlick's plasticity account is as a first step beyond task-oriented accounts of intelligence, such as speed of information processing and brain size, which account for psychometric measures of intelligence in terms of variations within an intelligently designed network. Though Garlick even moves beyond the neural efficiency account by suggesting that the *process* of unsupervised intelligent design (plasticity) is central to intelligence in a human brain, he does not describe the design process itself. The only element of plasticity that Garlick elucidates with computational and theoretical clarity is *more plasticity*, that is, the variations in the rate at which connections are made. But in order to escape the trappings of circularity, one must specify the constraints by which a network enacts *more appropriate plasticity*, which cannot be provided outside a theory of a self-organizing, relevance-realizing brain.

Indeed, much contemporary literature on intelligence from the philosophical, artificial intelligence and psychometric communities seems to point towards the importance of dynamical, self-organizing systems. The computer theory for the mind, the dominant metaphor in psychology until the 1980s (Haugeland, 1989), is gradually being replaced by connectionism and dynamical self-organization. Part of Garlick and Van der Maas' insight was to realize that the emergence of new frameworks for the mind entails the necessity of new frameworks for intelligence. Yet in order for their theories not to viciously invoke an unspecified mechanism of relevance realization, we must specify not only that intelligence arises *because of* appropriate plasticity or mutualism, but also the mechanisms *from which* appropriate plasticity and mutualism arise.

We believe that the opponent process constraints of RR can lend this specificity to dynamical systems theories. We also argue that these RR constraints are a plausible link between the powerful, brain-based, yet seemingly contradictory theories of categorical intelligence provided by Mercado and Hawkins. Much work is needed to specify how the opponent processes of RR are implemented in the brain, and to elucidate the specific empirical connections between RR and psychometric g. Yet a new philosophical and psychological framework must arise to reflect the emerging knowledge that our world is peopled by dynamic, self-organizing and relevance realizing creatures like human beings.

Acknowledgements

References

Ackermanm, P. L., Beier, M. E., & Boyle, M. O. (2005). "Working memory and intelligence: the same or different constructs?" *Psychological bulletin*, *Vol. 131*. 30-60.

Ainsley, G. (2005). "Précis of breakdown of will" *Behavioral and brain sciences*, *Vol. 28*. 635-673.

Anderson, B. (1994). "Speed of neuron conduction is not the basis of the IQ-RT correlation: Results from a simple model" *Intelligence, Vol. 19.* 317–323.

Anderson, B., & Donaldson, S. (1995). "The backpropagation algorithm: Implications for the biological bases of individual differences in intelligence" *Intelligence, Vol. 21.* 327–345.

Bartholomew, D. J. (2004). *Measuring intelligence: facts and fallacies.* Cambridge, England: Cambridge University Press.

Blair, C. (2006). "How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability" *The behavioral and brain sciences, Vol. 29, No.* 2. 109-25; discussion 125-60.

Cherniak, C. (1986). *Minimal Rationality*. Cambridge, Mass: MIT press.

Zac/Irving Zac 9/3/09 10:30 AM
Comment: Fix

Clarke, A. Book

Clarke, A. (2007). "Messy Minds: Embodiment, Action and Explanation in 21st Century Cognitive Science. Keynote Speaker" *University of Toronto Interdisciplinary Symposium on the Mind.*

Eyesneck, H. (1987). "Thomson's "bonds" or Spearman's "energy": Sixty years on" *Mankind quarterly, Vol. 27*. 259-274.

Garlick, D. (2002). "Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism" *Psychological review, Vol. 109*, *No. 1*. 116-136.

Gentner, D. (1983). "Structure Mapping: A Theoretical Framework for Analogy" *Cognitive Science: A Multidisciplinary Journal, Vol. 7, No. 1*. 155-170.

Goodman, N. (1972). "Seven strictures on similarity" in N. Goodman (Ed.) *Problems and projects*. New York: Bobbs-Merril.

Grabner, R. H., Fink, A., Stipacek, A., Neuper, C., & Neubauer, A. C. (2004). "Intelligence and working memory systems: Evidence of neural efficiency in alpha band ERD" *Cognitive brain research, Vol. 20*, *Vol. 2*. 212-225.

Grabner, R. H., Neubauer, A. C., & Stern, E. (2006). "Superior performance and neural efficiency: The impact of intelligence and expertise" *Brain research bulletin, Vol. 69, No.* 4. 422-439.

Haier, R. J., Siegel, B. V., Tang, C., & Abel, L. (1992). "Intelligence and changes in regional cerebral glucose metabolic rate following learning" *Intelligence special issue: Biology and intelligence, Vol. 16 No. 3-4*. 415-426.

Happé, F., Frith, U., (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal or autism and developmental disorders*, *Vol. 36, No. 1*. 5-25.

Harlow, H. F. (1949). The formation of learning sets *Psychological Review*. *Vol. 56*. 51-65

Haugeland, J. (1989). *Artificial intelligence: the very idea.* Cambridge: MIT press.

Hawkins, J., Blakeslee, S. (2004). *On Intelligence*. New York: Owl Books.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Holyoak, K. J. (1982). Analogical Thinking and Human Intelligence, in R. J. Sternberg (Eds.) *Advances in the Psychology of Human Intelligence*. *Vol. 2*. New Jersey: Lawrence Erlbaum and Associates.

Holyoak, K. J. (1995). Problem Solving in D. N. Osherson & E. E. Smith (Eds.) *An invitation to cognitive science: Thinking*. 2^{nd} Ed., *Vol 3*. Cambridge: MIT Press. Pp. 267-296.

Kaplan, C. A. & Simon, H. A., In search of insight *Cognitive psychology*, *Vol. 22*. 274-419.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.

Klein, R. (2000). Inhibition of Return. *Trends in Cognitive Science, Vol 4, No 4*. 138-147.

Klein, R., MacInnes, W. J. (1999). Inhibition of Return is a foraging facilitator in visual search. Psychological Science. Vol 10, No 4, 346-352.

Luciano, M., Wright, M. J., Geffen, G. M., Geffen, L. B., Smith, G. A., & Martin, N. G. (2004). A genetic investigation of the covariation among inspection time, choice reaction time, and IQ subtest scores. *Behavior genetics, Vol. 34, No. 1*, 41-50.

May, R. M. (1973). *Stability and complexity in model ecosystems.* Princeton, NJ: Princeton University Press.

Mercado, E. III. (2008). Neural and Cognitive Plasticity: From Maps to Minds. *Psychological Bulletin*, *Vol. 134, No 1*. 109-137.

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for korsakoff amnesia. *Psychological review, Vol. 100, No. 1*, 3-22.

Morgan, B., & Gibson, K. R. (1991). Nutritional and environmental interactions in brain development, In K. R. Gibson & A. C. Petersen (Eds.), *Brain maturation and cognitive development.* New York: Aldine de Gruyter. 91-106.

Murray, J. D. (2002). *Mathematical biology: I. An Introduction*, 3rd ed. Betlin: Springer Verlag.

Newell, A., Simon, H.. (1972). *Human problem solving.* Inglewood Cliffs, NJ: Prentice-Hall.

Polya, G. (1973). *Mathematics and plausible reasoning, Vol. 1.* Princeton, NJ: Princeton University Press.

Pylyshyn, Z. W. (1988). *The robot's dilemma: The frame problem in artificial intelligence*. NJ: Ablex Publishing Corporation.

Rosch, E. (1978). Principles of Categorization, in R. Rosch and B. Lloyd, (Eds.), *Cognition and categorization.* New Jersey: Lawrence Erlbaum Inc.

Sejnowski, T. J., & Rosenberg, C. R. (1986). NETtalk: a parallel network that learns to read aloud In J. A. Anderson & E. Rosenfeld, *Neurocomputing: Foundations of research* Cambridge, MA: MIT Press. 663-672.

Siegler, R. S., & Alibali, M. W. (2005). *Children's thinking* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Spearman, C. (1904). *The abilities of man: Their nature and measurement.* New York: Macmillian.

Richard S. Sutton and Andrew G. Barto (1998). *Reinforcement Learning*. Cambrige, Mass: MIT Press.

Thorndike, R. L. (1927). *The measurement of intelligence*. New York: Teachers College.

Thorndike, R. L. (1994). g. *Intelligence*, *Vol. 19*. 145-155.

Thomson, G. (1951). *The factoral analysis of human ability*, 5th Ed. London England: University of London Press

van der Maas, H. L. J., Dolon, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., Raijmakers, M. E. J. (2006). A Dynamical Model of General Intelligence: The Positive Manifold of Intelligence by Mutualism *Psychological Review*, *113*, 842-861.

Vervaeke, J., Lillicrap, T., Richards, B. (forthcoming). Relevance realization and the emerging framework in cognitive science. *Journal of logic and computation: special issue on relevance.*

Willerman, L., Schultz, R., Rutledge, J. N., & Bigler, E. D. (1991). In vivo brain size and Intelligence *Intelligence, 15* 223–228.

Wittgenstein, L. (1958). *Philosophical Investigations.* Worchester, England: Billing & Sons Ltd.

Author Note

Zachary C. Irving, Department of Philosophy, University of Toronto; John Vervaeke, University of Toronto, Cognitive Science Program & Department of Psychology, University of Toronto.

Correspondence concerning the article should be addressed to Zachary C. Irving, at 274 Rhodes Avenue, Toronto, Ontario, Canada, M4L 3A3. Correspondence by email can be sent to zac.irving@utoronto.ca.

The authors would like to acknowledge the contributions of Dominic Alford-Dugid, Leo Ferraro, Mitchell Gerskup, and Adriana Robertson toward the preparation of this document.

Footnotes

1. For a more detailed introduction to the central importance of relevance eealization to Cognitive Science, see Vervaeke, Lillicrap and Richards (forthcoming). This includes in depth and computational models of the relevance realization constraints.

2. Our invocation of "intelligent design" is mean to point to the circularity of presupposing a well designed network in any explanation of intelligence. Our point is that design must be immanent to the network and not have an external or unexplained source. We are not in any way engaging with the debate about the existence of God.

3. Of course, environmental stimulation varies between individuals. As Garlick's is a theory that focuses on learning from the environment, it concedes that this variance should have an effect on IQ. However, it seems unlikely that environment can account for g, although this discussion is outside the bounds of this paper (for a review, see Jensen, 1998).

4. Note how this word "better" is vague. This plays into our critique of both Garlick and van der Maas.

5. Van der Maas et al also argue that mutualism has explanatory value over: developmental effects; integration/ differentiation effects; the increase in the heritability of intelligence; the Jensen effect; the Flynn Effect; and even given high heritability of g. Therefore, any model that is created within the mutualism framework also explains these effects.

6. As an example of invariant representations, Hawkins gives the example of one's ability to notice a friends face as the same regardless of the varying conditions – different visual perspectives (e.g. profile, front etc.), lighting etc. – in which one sees her face. Instead, one notices, and builds a representation in memory out of her invariant features (e.g., the proportions between his facial features, textures etc.)