

Yaroslav Riabinin

“Embodiment without Intelligence”

Introduction

Traditional accounts of the mind are generally rooted in the belief that some form of representation must be postulated in order to explain cognition and intelligence. Although this outlook has been widely embraced in cognitive science, successful research in situated cognition has challenged the classical conception of the mental. The rival view holds that the world is its own best model and that representational theories of the mind are inadequate and unnecessary. Instead, a different approach must be adopted, one where systems are designed to exhibit certain behaviour in response to their surroundings. In this essay, I will discuss the role of representation in cognition and consider the alternative of embodiment. I will then offer a critical evaluation of the behaviour-based model of intelligence, as exemplified by the work of Rodney Brooks.

Part I: Background on Representation and Embodiment

Mental Representation in Cognition

It is often the case that humans picture thought as resembling a computational process, which involves manipulating mental objects that have meaningful content. For example, after a period of careful introspection, most people would feel inclined to say that they have beliefs, ideas, and opinions about the world, and that they make decisions or reach conclusions by reasoning with such items. The classical model of the mind is based on these intuitions – that we form explicit mental representations and store them for later inference by cognitive mechanisms.

It is also evident that intelligent behaviour may be explained with reference to such mental representations. This stance is commonly known as “folk psychology” and it claims the

following: if we assume that people have mental states and propositional attitudes, then we can predict and rationalize their seemingly puzzling conduct. For instance, consider a situation where a student stays home to study for an upcoming test instead of going out to a bar. To explain this behaviour, one could imagine that the person has a general representation of the world, as well as various relevant mental states, such as the desire to get a good grade, the belief that studying harder will increase your chances of academic success, the belief that going out will distract you from studying, and so forth. Then, one could say that a series of inferences takes place in the person's mind and the outcome is the behaviour exhibited by the student.

However, the notion of mental representation has greater value than simply providing a means of accounting for a creature's actions. It allows us to reach a level of cognitive abstraction where we can deal with ideas and concepts that are not directly accessible to us in the physical world. Such deliberation – planning, philosophizing, speculating, etc. – is certainly characteristic of human intelligence and it would be unwise to treat it as a trivial concern. Consider our ability to discuss “fear” without actually experiencing it, or to contemplate “justice” while driving to work – would this be possible if we weren't able to represent these notions in the first place? It is difficult to imagine the process of thinking and reflection without entities to *think with* or *reflect on*.

Embodied and Embedded Cognition

It must be noted that mental representation is not always necessary to explain intelligent behaviour. In fact, it can complicate matters tremendously in some domains of activity. For example, consider instances of spatial navigation – walking, climbing, swimming, and so on. If we adopt a fully representational view of the mind, then we must think of action and movement as being guided by an internal model of the world, which represents a great deal of information and must be updated regularly. So it follows that any commonplace physical task must involve dynamic mental representation of the surroundings at frequent time intervals. The overall

complexity of this process is highly problematic, especially in light of the ease with which humans perform such tasks. Hence, it is no surprise that an alternate model of cognition was devised to deal with these issues.

In an article entitled “Intelligence without representation”, Rodney Brooks proposes that we eliminate the “intermediate step” between perception and action by abandoning centralized representation entirely. He advocates a belief that the world is its own best model, which means that we can make a creature exhibit complex and seemingly intelligent behaviour by having it rely on the environment itself, rather than on explicit mental states and computational processes. This is done by carefully engineering a system that senses the world and responds to it appropriately and in a timely fashion. The underlying architecture of such an embodied creature is composed of “layers” that are each responsible for a certain activity – like roaming around the room, or avoiding obstacles. Hence, intelligence is viewed not as a property of the agent, but rather as the interaction of the creature with its environment. Although this conviction greatly facilitates the design of intelligent beings, I will argue that such a behaviourist approach to cognition is fundamentally flawed in that it overlooks the importance of mental representation to human-level intelligence. But first I will briefly clarify what is meant by the term “representation”.

At first glance, the distinction between the classical view of the mind and the embodiment alternative is fairly clear – the latter completely erases the notion of representation from cognition. Upon further consideration, however, the matter becomes slightly more complicated. It could be claimed that meaningful content is *implicitly* represented by the electrical currents in Brooks’ robots – in other words, that the notion of representation can be extended to encompass plain sensory input. Moreover, it could be said that the electrical signal is “interpreted” by the receiving mechanism and that a decision is made based on that *meaningful* interpretation, which causes the creature to produce the observed behavioural response. The question, then, is whether it is appropriate to describe a merely causal process (e.g., electric

impulse causing robot to turn left) in the same manner as a computational process involving explicit representation of the world.

In order to gain more insight into the problem, let us consider a simpler example of one billiard ball (A) colliding with another (B) and propelling it forward. The obvious observation is that there is a cause and effect relationship between A and B. But if we were to be more generous than that, we could contend that B interprets the contact with A as an expression of hostility and immediately decides to take flight from the intruder, since B has a strong sense of self-preservation. Both explanations are plausible, but the burden of proof is on those who advocate the latter account – they must justify their liberal construal of “representation” and prove that the attribution of mental states to objects that are otherwise unequipped for such luxury is warranted. For our present purposes, we will require mental representations to be explicitly symbolic entities with semantic properties, such as truth value and reference. They must be stored in some sort of repository, so that they could be accessed and manipulated in a way that may be analogous to computational processes. I will now proceed to offer a criticism of the “embodied and embedded cognition” model in order to illustrate the downfalls of a behaviour-based approach to intelligence and emphasize the importance of representation.

Part II: The Trouble with Situated Cognition

Before painting a negative portrait of embodiment, however, we must first consider the significance of Brooks’ contribution to the study of cognition. It would be rather unfair to claim that his research did not yield valuable results. On the contrary, his success in robotics has drawn much needed attention to the problem of dynamic interaction with the environment. This issue has been largely ignored by classical approaches to artificial intelligence, but it is now apparent that such disregard for embodiment is unwarranted and potentially detrimental. The idea of relying on the physical world for guidance is extremely powerful because it allows us to account

for complex behaviour without alluding to elaborate mental processing. The obvious benefit is an enormous reduction in the computation that must be carried out by the system, especially for routine tasks, such as jogging or looking for your keys. Despite these considerable advantages, however, it would be wrong to conclude that embodiment alone is sufficient for human-level cognition. Far too much is omitted in this model, as will be shown next.

First, let us turn our attention to the “layered” architecture proposed by Brooks. Its strength lies in its simplicity – it consists of nothing more than pre-determined responses to sensory input obtained from the external world, which gives rise to behaviour that is often complex and seemingly *goal-oriented*. But the reality is that such goals are *implicit*, which means that they have no representation in the system. This position does not seem to be consistent with our intuitions about human cognition. For instance, how does one account for the fleeting and spontaneous nature of our goals? They can be formed at any instant and discarded with the same ease (e.g., I decide to go fishing and then change my mind five minutes later). Moreover, it raises the question of whether an agent could even be said to possess goals if it is not able to represent them explicitly. For example, imagine seeing a creature that attacks an approaching predator, and kills it. One person could claim that the creature’s goal is to become the dominant life form by eliminating all rivals. Another person could say that the creature’s goal is self-preservation (after all, the threat was removed). Both claims appear to be reasonable given the observed behaviour of the creature – so which is the correct one? It seems the choice is arbitrary if one was to believe that goals are implicit in the dynamic interaction between the agent and its environment, which is the case in Brooks’ model. Hence, it is not entirely clear that we can afford to eliminate mental representation from any adequate theory of cognition.

There is a similar concern with the claim that an agent’s behaviour is determined by its contact with the environment. Once again, this statement does not seem to do justice to human-level intelligence. Why is it that two people can behave differently under the same

circumstances? Moreover, how can an individual decide to perform one action over another? Without addressing the question of free will, one can still assert that humans make choices, or at least they are able to. For example, if I am at my computer, I can choose to check my email, or listen to music, or leave the room, or do a handstand on the carpet, or engage in an infinitely many activities, in any order and location. The environment does not dictate what I must do next, although it certainly imposes limitations – for instance, if it's nighttime then I won't be able to go outside and enjoy the warmth of the sun. But the surroundings clearly do not play an exclusive role in determining my next move. Is the same true of a creature with a layered, behaviour-based architecture? It seems not, since it was designed with the opposite view in mind – there are no internal variables or rules to follow. For example, recall Brooks' best known creation – *Herbert*. Its three-second memory does not allow it to do anything but *respond* to the environment. Hence, this model may not offer a satisfying account of cognition.

Although the general notion of letting the world be its own model can considerably simplify the design of autonomous agents, it also deeply restricts one's ability to make sense of human behaviour. The reason is that such an approach ignores mental deliberation as a cause of action and instead is only concerned with the immediate physical surroundings of the individual. Consider the following scenario: you are sitting in the kitchen, having breakfast. You remember visiting the mall yesterday and having a great time. Suddenly, you recall that it is your uncle's birthday tomorrow and you forgot to buy him a gift. You imagine yourself in your uncle's place and conclude that he would feel neglected. You project the future embarrassment that you would experience at your uncle's party and decidedly vow to avoid that fate. So you storm out of the house and head back to the mall. Such mental sequences are common in humans and surely classify as intelligent thought, but the resulting behaviour cannot be explained with reference to the environment alone. Hence, it is clear that some form of representation is necessary for inference to take place in the mind.

The final challenge to situated cognition concerns the view of intelligence under that model. Brooks and other proponents of embodiment might claim that intelligence is not actually a property of the individual, but rather it is the observer that ascribes intellect to an entity based on its behaviour, which emerges out of the local chaos of interactions between “layers” of its system. The fundamental question, then, is whether this conviction is correct, or at least justified. Is there any reason for us to attribute human properties to other living creatures – or even inanimate objects - solely on the basis of their conduct? For example, consider a robot that detects sunlight and moves towards it. If a person encounters this machine and concludes that it “likes the warmth” or “becomes happy when it feels light”, does this imply that the robot is genuinely intelligent? There seems to be no grounds for holding that belief. The creature possesses a “derivative intelligence” at most, which may be the result of a human tendency to personify things that are otherwise unfit to be endowed with such qualities. In any case, it is certainly troubling that non-representational theories of mind rest so heavily on a behaviourist approach to cognition.

Conclusion

Although I have argued in this essay that situated models of cognition do not by themselves offer a satisfying account of human-level intelligence, I agree with Brooks when he declares that they provide a necessary foundation for the cognitive scientific endeavour. In order for us to transcend existing boundaries and engineer genuinely intelligent artifacts, I believe that classical theories of representation and reasoning must be integrated with newer models, such as connectionist networks and embodiment in the physical world. It is only if we achieve unity and establish a division of labour between various subsystems that I think we can hope to account for the full scope of cognition. Otherwise, each camp will only get a glimpse at success, without ever seeing the full picture.

References

Brooks, R.A. (1991). *Intelligence without representation*. Artificial Intelligence 47, 139-159.

JUP302 Winter 2006. *Lecture Notes 10b*. Brian C. Smith & Jun Luo.