Autism and Bounded Rationality

Frederik Eaton
Department of Engineering
University of Cambridge

February 13, 2008

Abstract

In this essay we explore connections between bounded rationality and autism. In particular, we consider the possibility that autism results from an impairment in the brain's ability to behave in a boundedly rational way.

Contents

1 Overview of Bounded Rationality

Boundedly rational (BR) reasoning can be defined as "reasoning about the value of reasoning". BR is a concept which arose in contrast to a common simplifying approximation to rational behaviour, called **perfect rationality**. A perfectly rational agent will always select from a set of available actions the one that maximises his expected utility. In reality, the utility maximisation problem may be difficult or impossible to solve exactly. Perfect rationality ignores the cost of deliberation. For a more accurate model of rational behaviour, we should take

this cost (and related costs) into account, at which point we are said to be modelling BR. We can hold on to the goal of maximising expected utility, but now our utility function should account for computational costs. These may differ depending on how we arrive at a decision. In effect, with BR we want to choose a utility-maximising *procedure*, subject to architectural and computational constraints, for making a series of resource-constrained decisions (this is sometimes referred to as **procedural rationality**) - rather than, as in the case of perfect rationality, making separate utility-maximising decisions directly.[1]

BR is a form of meta-reasoning. Humans seem to be quite good at it. It seems especially important when we are solving difficult problems, where exact solutions are intractable. Such problems require approximation, and choosing an appropriate level of approximation means finding the best trade-off between accuracy and computation, which requires us to reason about the costs and values of the computations we perform. Without BR, we would not be able to give up on solving a problem, or to forget a fact which is not useful; games like chess or go, which require us to limit ourselves to exploring a tiny subspace of possible plays before selecting one of them, might be difficult. On the other hand, lack of awareness of the value of reasoning would not preclude our following a deterministic, predefined set of rules to reach a conclusion, so problem domains where such a process is effective would still be accessible to us. This is akin to what computers do now - executing programs, but not thinking about them in the process. We view modern computer systems as being somewhat poor at BR. They don't have much awareness of the value of their computations, as evidenced by the fact that they generally don't get better at them with practice.[2]

2 Our Cognitive Model

2.1 Transaction Costs

It is natural to think of various costs associated with transactions in an economy as analogous to costs of computation. The concept of BR was initially studied in this form, such study being motivated by a desire to understand the economics of organisational structure and administrative behaviour.[Sim47] The existence of firms is not well-explained by neoclassical economics and general equilibrium theory, which ignores transaction costs. In an idealised economy without transaction costs, employees would always get the best value for their labour by continuously selling it on the open market. As a result, there would be no need for long-term employment contracts or any of the hierarchy that we see in real-world firms. However, when we take into account the fact that transaction costs are in reality non-zero, we see that the existence of firms can be explained as resulting from the desire to reduce these transaction costs. The principal way in which a firm reduces transaction costs is by simplifying the decision problems of employees, which it accomplishes by restricting the possible actions they must choose between in the usual course of work. By working for

[1] In the BR regime, what we can continue to ignore is the cost of *designing* a decision-making procedure, which may be very high - such as, for instance, billions of years of biological evolution in the case of humans.

[2] We see no reason why modern computers couldn't be given a BR-mechanism, aside from engineering difficulties.

a new employer each day, a labourer might be able to take advantage of the efficiency of the free market to get an "optimally" competitive wage, but when we account for the costs of advertising his services, evaluating potential offers, meeting a new boss, learning a new trade, relocating to a new city, and so forth, we see that the transaction costs associated with changing employers every day would, in most cases, erase the possible gains from getting a higher salary in the free market. From experience, we know that such transaction costs can be reduced to a more reasonable level if the employee engages in longer-term employment contracts, specialises in a certain area, and works for the same boss every day. So, transaction costs give a natural explanation for the existence of firms.

This analogy to economics is suggestive. Most modern theories of cognition suppose that the brain, like an economy, operates in a highly parallel manner. This is sensible because most cognitions are likely to serve independent purposes. For instance, whether I conclude A first and then B, or B first and then A, I will end up having concluded both A and B - the actions are said to be **commutative**. Similarly, I can buy two commodities in any order, and in either case I will own both of them. Although there are ways in which the order of thinking or trading is important (often involving BR or transaction costs), to first approximation the value of making a set of trades or thinking about a set of ideas is independent of the order in which those actions are carried out. When one learns a new language, it may be easier to learn the more common words first, but generally the order in which new words are learned can be made quite flexible. Note that, compared to computation and economics, most real-world problem domains are not very commutative. For instance, getting dressed requires a somewhat inflexible ordering of actions - socks before shoes, and so on. Commutativity is thus somewhat special. Noticing that commutativity is shared by both (cognitive) beliefs and (economic) allocations, it might seem more plausible that methods from economics would be useful for reasoning about reasoning.

2.2 Decentralisation

To build on this analogy to economics, we note that in an economy, a decision whether or not to make a particular trade is primarily made by the two parties to that trade. A government may restrict the kinds of trades which are allowed, or impose taxes, but it is generally found that the most efficient way to run an economy is to leave trading decisions to concerned parties. They have the best information about their own constraints and preferences, and also the best motivation to make a good decision. In other words, transaction costs must be taken into account independently by each agent. To refine an economic model by adding a notion of transaction costs mostly involves adjusting the behaviour of each agent individually, rather than, for instance, creating a new branch of government to centralise such decisions.

Similarly, we expect that when the brain takes into account the cost and value of reasoning, it does so in a distributed fashion. We would be surprised, for instance, to find a particular region of the brain which is solely responsible for BR. More likely, since every aspect of cognition has its own costs and values, we expect decisions concerning the value of a particular region's functioning to be mostly local to that region. The more decentralised these decisions are,

the simpler they become, so it is plausible that much (but probably not all) of the mechanism for boundedly rational reasoning in the brain is implemented at the level of individual neurons. We will use this assumption in the next section when we try to imagine the behavioural implications of a hypothetical BR-impairment.

3 Autism

3.1 Overview

Autism is a condition which is characterised by a wide range of behavioural symptoms, which can each be placed broadly into one of two somewhat dissimilar groups. We shall call these groups "social" and "general". The social group is the most well-known of the two, and consists of behaviours such as

(Social group)

- lack of empathy
- wanting to leave when engaging with others
- attachments to objects rather than people
- inability to reason about others' beliefs
- one-sided verbosity
- failure to interpret body language
- failure to interpret signs of boredom
- non-verbal learning disabilities
- alexithmia (lack of language for feeling)
- theoretical understanding of other peoples' emotions, but difficulty acting on this in real life situations
- lack of understanding of own beliefs and emotions

Because we live in a social world, impairments in social functioning are often the most visible symptoms of a cognitive disorder, and are the first to be observed and researched. However, social symptoms are often one part of a larger pattern. In autism, what we are calling the "general" group of symptoms takes a somewhat different shape from the above:

(General group)

- interest in unusual topics
- passionate pursuit of narrow fields of interest, possibly without understanding of broader phenomena
- preoccupations with parts of objects
- good memory for trivial facts
- inability to approximate
- inflexible adherence to routines or rituals
- excessive rule-making
- need for "sameness"
- repetitive patterns of behaviour
- stereotyped mannerisms
- compulsive movements, hand flapping, tics and stims
- sensory hypersensitivity [3]

[3] to sound, touch, taste, sight, smell, pain, temperature, texture of foods, etc.

- synaesthesia [4]
- poor motor planning
- toe walking
- impaired tactile perception (in children but not adults)
- poor muscle tone
- sleep problems
- atypical or selective eating behaviour
- gastrointestinal symptoms (anecdotal)
- poor immune response (anecdotal)

We have tried to order the symptoms on the two lists from "most social" to "least social", so that it is possible to see that the dividing line between them is somewhat arbitrary. For example, many people become interested in things that those around them are interested in, so we could arguably reclassify the first symptom on the "general" list ("interest in unusual topics") into the "social" category; likewise, the last symptom on the "social" list ("showing a theoretical understanding of other peoples' emotions, but difficulty acting on this in real life situations") seems to also suggest more of a *general* cognitive or perceptual impairment.

According to Wikipedia, existing theories of autism may explain one or the other group of symptoms well, but not both. Thus, current theories of autistic behaviour can be put into two categories according to the group of symptoms which they best explain. **Social cognition theories**

> [focus] on deficits in social cognition. **Hyper systemizing** hypothesizes that autistic individuals can systematize–that is, they can develop internal rules of operation to handle internal events–but are less effective at empathizing by handling events generated by other agents. It extends the **extreme male brain theory**, which hypothesizes that autism is an extreme case of the male brain, defined psychometrically as individuals in whom systemizing is better than empathizing. This in turn is related to the earlier **theory of mind**, which hypothesizes that autistic behavior arises from an inability to ascribe mental states to oneself and others. ...

while **nonsocial theories**

> [focus] on nonsocial or general processing. **Executive dysfunction theory** hypothesizes that autistic behavior results in part from deficits in flexibility, planning, and other forms of executive function. A strength of the theory is predicting stereotyped behavior and narrow interests; a weakness is that executive function deficits are not found in young autistic children. **Weak central coherence theory** hypothesizes that a limited ability to see the big picture underlies the central disturbance in autism. One strength of this theory is predicting special talents and peaks in performance in autistic people. A related theory–**enhanced perceptual functioning**–focuses more on the superiority of locally oriented and perceptual operations in autistic individuals. ...

[4] in which stimulation of one sensory or cognitive pathway leads to involuntary stimulation of another

We can see it as a goal for autism research to find a theory which explains all of the above symptoms of autism well. Ideally, such a theory would eventually be able to describe the causes of autism at a very low level, for instance at the neural or genetic level, and to provide a coherent description of higher-level behavioural symptoms in terms of these low-level differences. In what follows, we will not be able to provide such a detailed picture of the autistic condition, but we do hope to persuade the reader that the right foundation for a unified cognitive-neurological theory of autism is the concept of bounded rationality.

3.2 Connections to Bounded Rationality

We will start by hypothesising an individual whose ability to carry out boundedly rational metareasoning is somehow impaired - we will say that he is BR-impaired. Let us think about how we might expect his behaviour and cognition to differ from that of a normal individual.

3.2.1 General Symptoms

In section 2.2, using an analogy to economics, we tried to argue that reasoning about the cost of reasoning is most efficiently carried out in a distributed fashion. We think that value judgements are "woven in" at the lowest levels of computation, rather than being performed separately by a special module.

Thus, we would expect a BR-impaired person to have, just like an autistic person, structural abnormalities all over the brain, rather than localised to a particular functional unit.

We would expect a BR-impaired person to suffer from general over-stimulation, for instance sensory hypersensitivity (which may cause **atypical eating**, **toe-walking**, and **sleep problems**). That is, he would have difficulty assessing the value of information coming from the senses, which might then be "overprocessed" by his brain.[5] If neurons controlling muscles are also prone to excessive activity, then it would be natural to expect **tics and stims** and **compulsive movements**, and perhaps **gastrointestinal dysfunction**. Under-stimulation of neurons connected to the immune system would lead to a **poor immune response** and thence **poor muscle tone**. **Synaesthesia** may also be the result of nervous over-stimulation.

At a higher level, decision problems where approximation is necessary are generally difficult or impossible to solve exactly, so they also tend to be the kind of problems which benefit most from BR. Therefore, we would expect a BR-impaired person to suffer from an **inability to approximate**.

These are all symptoms from the "general" group, thus we have a good match so far between autism and BR-impairment.

3.2.2 Coping Mechanisms

To explain symptoms which are more social in nature - starting with those at the top of the "general" group, such as interest in unusual topics, narrow fields of interest, or need for sameness - it becomes necessary to make a minor extension to the theory.

[5]It might also be "underprocessed" - but given limited resources, if some input is overprocessed then it seems unavoidable that other input would be underprocessed

Although on the surface some of these behaviours can be seen as stemming from an impaired ability to prioritise cognitions, or in other words somewhat directly from a BR-impairment, it makes more sense to interpret them instead as *coping mechanisms* for such an impairment. For BR-impairment isn't the same as irrationality - just as an autistic person might be unable to implement certain social rules in a realtime exchange, but yet at the same time might be able to demonstrate a sound theoretical understanding of those same rules; in a similar way, we would expect a BR-impaired person to be able to make rational decisions about important long-term issues - such as choosing a field of interest - even if it takes longer than average for him to do so (or even if he has to go to others for advice). Once he has in hand the justification behind such a decision, checking its validity is a straightforward process which he can perform without a need for BR-metareasoning. So we don't expect these long-term behavioural symptoms to be dysfunctional in themselves.

Instead, we propose that the autistic person is displays **interest in unusual topics** not because he is "unaware that they are useless", but because he finds that they provide him with a way to engage with the world that takes advantage of his strengths, while avoiding his weaknesses. By becoming an expert on a **narrow field of interest**, someone who is BR-impaired can avoid having to make high-level decisions, for instance about usefulness of a fact, or proper amount of detail, which he is unskilled at making. By carefully controlling his environment (**need for sameness**), he can avoid having to make constant relevance judgements about new sources of sensory input. By **inflexibly adhering to routines and rituals**, and through **excessive rule-making**, he can minimise the amount of time which is spent on minor day-to-day decisions, or on reacting spontaneously to external demands - tasks which he finds difficult because they make heavy use of BR.

If one accepts this line of reasoning, we again find that the symptoms predicted by hypothetical BR-impairment closely match those of autism.

3.2.3 Social Symptoms

At this point, we will attempt to explain autistic symptoms in the social group. To do this, we propose a further extension to our theory, which will be the last but also the most substantial.

It is clear that BR, i.e. reasoning about the value of reasoning, is a kind of *introspection*. Noticing that autistic people often have trouble introspecting about their own beliefs or emotions, we are motivated to ask: what other basic kinds of introspection exist, distinct from those needed for BR? We conjecture that the answer to this question is: none.

For upon some reflection, we find that it is possible to express almost any form of introspection as a question about the cost of a computation.

If I want to ask myself whether I know the US pledge of allegiance, I can query "Is the cost of computing the words to the pledge of allegiance small?" - if the 'computation' is simply a matter of recalling the answer from memory, then the cost will indeed be small.

Now, if I want to ask myself if I am depressed, I can say, "Does my brain seem to be paying a greater price for sad thoughts than happy thoughts?".

If I want to know whether I believe a politician, I will ask "Do I put a high value on the information that the politician is telling me?".

In each case, the introspective query can be cast in terms of values of cognitions.

Since BR is more generally useful than other forms of introspection, we can suppose that it evolved first. Yet if that is the case, and if other forms of introspection can be implemented in terms of BR, it seems reasonable to assume that evolution would take the simplest approach, and select BR as a basis for implementing higher forms of introspection and self-consciousness. [6]

If BR underlies all introspection, then we would indeed expect a BR-impaired person to have trouble with introspective tasks such as reasoning about his own beliefs and emotions, just as autistic people are observed to do.

As for the rest of the social symptoms, although they deserve additional attention, it is hopefully not too difficult to imagine that impaired reasoning about one's own mental state would extend to impaired reasoning about the mental states of other people. It is possible that, for instance, empathy corresponds to introspection of the mirror neuron system. Thus, we predict that the more social symptoms of autism, such as **inability to reason about others' beliefs**, **lack of empathy**, and so forth, derive more or less directly from an impairment in introspection.

4 Conclusion

In this essay, we have tried to explain all of the symptoms of autism as arising from an impaired ability to carry out boundedly rational reasoning. We arrived at three hypotheses: (a) in section 2.2 and section 3.2.1, that the BR mechanism must be distributed, and that this explains the lowest-level symptoms of autism; (b) in section 3.2.2, that BR-impairment gives rise to certain general symptoms of autism as coping mechanisms; (c) in 3.2.3, that bounded rationality underlies all introspection, and that this explains the social symptoms of autism.

Our chain of reasoning has been quite high-level and philosophical, and when it comes to cognition there are often many different yet equally plausible ways of explaining the same observations. One direction for future investigation may be to explore the effects of changing the price of computation on the behaviour of computer algorithms that have been designed to operate in various boundedly-rational ways. For instance, it might be fruitful to explore how varying the price of computation in an inference algorithm (adjusted relative to the price of accuracy, for instance) changes the way certain information is encoded. More specifically, if we are able to find an algorithm for sensibly adjusting the shape of a network by adding and deleting nodes, then it would be a simple matter to measure, for instance, how the optimal network connectivity varies with price of computation - and to check whether the result is consistent with observed differences in neural structure in the brains of autistic people (e.g., having more neurons with fewer long-range connections).

However, we believe that the most effective way to verify the components of our theory may be, in the end, through psychological experiments involving autistic people. Since we are not psychologists, we refrain from guessing what form such experiments might take.

[6] The fact that we rarely if ever observe symptoms from the social group of symptoms, without symptoms from the general group as well, or vice-versa, can be seen as additional evidence that BR and introspection are based on the same underlying mechanism.

5 Acknowledgements

The author would like to thank Edgar Moose for a number of useful discussions.

References

[Sim47] H. Simon. *Administrative Behaviour*. PhD thesis, University of Chicago, 1947.