Cognitive Externalism meets Bounded Rationality

Eric Arnau (eric.arnau@gmail.com)
Department of Philosophy, Universitat Autònoma de Barcelona
Edifici B, Campus UAB, 08193 Bellaterra, Spain

Saray Ayala (sarayayala@gmail.com)
Department of Humanities: Philosophy, Language and Literature
University Carlos III, Madrid 126, 28903 Getafe, Spain

Thomas Sturm (tsturm@mpiw-berlin.mpg.de)
Department of Philosophy, Universitat Autònoma de Barcelona
Edifici B, Campus UAB, 08193 Bellaterra, Spain

Abstract
When proponents of cognitive externalism (CE) have turned to empirical studies in cognitive science to put the framework to use, they have typically referred to perception, memory or motor coordination. Not much has been said about reasoning. One promising avenue to explore here is the theory of bounded rationality (BR). In this paper, we try to clarify the potential relationship between these two programs. We start by discussing Andy Clark’s interpretation of BR, which we find unconvincing in several respects. Next, we take a closer look at CE in order defend a version of it that stands against mainstream internalism without committing itself to constitutional claims about the mind. We then turn to analyze BR from the CE perspective. Finally, we argue that internalism about cognition cannot explain important aspects of the BR program.

Keywords: extended cognition; bounded rationality; heuristics.

Introduction

By Cognitive externalism (CE) we refer to the framework that accommodates the initially differentiated challenges to the internalist picture of cognition, developed under the flag of extended cognition and distributed cognition. CE departs from the original proposal introduced by Clark and Chalmers (1998) in which what determines the cognitive status of an extended process is its functional parity with an intracranial cognitive process. The core of this shift is the complementarity of internal and external elements (Menary 2007, Sutton 2010, Wilson and Clark 2009). The argument for extended cognition turns on the way different inner and outer components co-operate so as to yield an integrated system capable of supporting intelligent behavior.

One consequence of this revised CE is a broadening of the span of the putative cases considered, as it points our attention to the many ways in which “the computational power and expertise is spread across a heterogeneous assembly of brains, bodies, artifacts, and other external structures” (Clark 1998, p. 77). Once parity is not required, we can tackle any manipulation of structures or elements that is integrated –e.g. measurement instruments, information storage devices, representational systems, etc.

Typically, proponents of externalist accounts of cognition have sought for empirical support from research in lower cognition, mostly memory (Clark and Chalmers 1989, Sutton et. al. 2010), and perception (Wilson 2010). But seldom the literature has addresses the realm of higher cognition. Here we want to explore the benefits that CE could offer to empirical research on human reasoning, and whether such research can vindicate CE.

Bounded Rationality

A promising avenue to explore is the interplay between CE and the program of Bounded Rationality (BR) as it is developed by Gerd Gigerenzer and colleagues into the Fast and Frugal Heuristics theory. The core research question of BR is: “How do people make judgments and decisions in everyday life, when time and information is limited and the future uncertain?” (Gigerenzer 2008, p. 5).

From here, the program unfolds in three dimensions: (1) The core descriptive tenet is that people typically rely on fast and frugal heuristics. There are strategies, conscious or unconscious, that search for minimal information and exploit evolved capacities and environmental structures (Gigerenzer and Brighton, 2009). (2) The core normative tenet is that such reliance on heuristics is at least oftentimes desirable. Heuristics work remarkably well; as good as –or even better than– optimizing strategies. They are not merely second-best options for “wanna-be optimizers”. (3) The core prescriptive tenet is that the BR picture of reasoning affords prescriptive guidance, opening a space in which to enhance reasoning by intervening on the environment rather than in the inner processing.

There is a strong prima facie affinity between BR and CE. In the BR literature one can find many claims that echo the externalist tenets. We find claims such as that “the heuristic lets the environment do much of the work” (Gigerenzer and Hutchenson 2005, p. 101), or that “in the ecological view, thinking does not happen simply in the mind, but in interaction between the mind and its environment” (Gigerenzer 2008, p. 17, emphasis added). Somewhat surprisingly, BR has attracted only one discussion within
the externalist camp, by Andy Clark. In the next section we turn to his discussion of BR.

**Clark on Bounded Rationality**

Clark (2001, 2003) addresses the relation between CE and BR guided by his own defense of CE, but also by his favoring of a “naturalistic” or “mechanical” account of rationality, i.e. one that is merely descriptive. Clark claims that there are two ways to develop a “naturalistic” or “mechanical” account of rationality based on a non-internalist approach in cognitive science. First, a “biological cognitive incrementalism” (BCI), according to which “full-scale human rationality is reached, rather directly, by some series of tweaks to basic biological modes of adaptive response” (2001, p. 121). Defenders of BCI view reasoning or rationality as continuous to more “basic” processes, such as perceptual responses to one’s environment, that do not require the manipulation of symbols. A second option assumes that rationality is at best indirectly based upon processes of BCI and, instead, strongly based upon “symbiotic relationships with knowledge-rich artifacts and technologies” (2001, p. 122). Not surprisingly, Clark favors the latter view, since it obviously comes down to his version of CE. His argument accordingly should have two sides, one being directed against BCI, the other in favor of CE.

In Clark’s view (2001, p. 126), BCI can be characterized by three core assumptions: (1) the thesis of organism-environment interaction, or anti-representationalism; (2) the modularity thesis (heuristics, short-cuts which are locally rather than globally active); (3) the thesis of distributed cognition. He illustrates these assumptions with research on the wing-flapping of houseflies and phontaxis in robot crickets (ibid., pp. 126-129). All behave successfully in their environments using extremely simple means. This “breeds skepticism” that “symbols, internal representations and the like play little role even in advanced human problem-solving” (ibid., p. 129). However, as Clark rightly points out, the anti-representationalism entailed by this view must not be taken too far, since typical instances of reasoning, such as drawing an inference or making a choice, are all “representation-hungry” activities (ibid.). Clark does not claim that BR requires that assumption, but he does think that BR is committed to assumptions 2 and 3. Simple heuristics, after all, are supposed to make us smart because of their joint exploitation of evolved capacities and the structure of (the information in) the environment. To that extent, BR seems to be a naturalistic account of rationality that also subscribes to BCI.

Let us assume for the moment that the program of BR is indeed to be understood as endorsing (2) and (3). What, then, is Clark’s objection to these contentions? His claim is that BR, thus understood, still does not grasp what we understand as full-scale human rationality (Clark 2001, p. 131). As he also says, “In much of this recent work, traditional conceptions of thought, reason and action are not so much reworked as by-passed entirely.” (ibid., p. 126). But what is missing? Elsewhere he declares: “Rational behavior is, in some sense, behavior that is guided by, or sensitive to, reasons. Intuitively, this seems to involve some capacity to step back, and assess the options; to foresee the consequences, and to act accordingly.” (Clark 2003, p. 314) But this is surely only a necessary condition, and it is surely not denied by Gigerenzer’s program. Although he emphasizes that we should often rely on such heuristics, he does in no way think they are to be used blindly. Any reasoner must ask himself or herself, Am I in an environment where this heuristic works? If not, can I use another heuristic, and which one? That obviously involves the ability to step back, assess the options, and so on. We should also note in this context that Gigerenzer has explicitly distanced himself from characterizing heuristics in terms of the modularity thesis: While they are short-cuts that are tuned to specific domains, heuristics need not be viewed as implicit as, say, perceptual processes (see e.g. Gigerenzer 2007, p. 43f.).

Now, what about the argument in favor of Clark’s own position, viz. that rationality can better be understood by his extended mind thesis? Clark (2001, pp. 132ff.) points to the frequent close coupling of reasoning processes with artificial aids and scaffolds. But he actually expresses his thesis in two different ways: (1) We do understand rationality only if we see it as a capacity “tuned and applied to the very special domain of external and/or artificial cognitive aids” (2001, p. 131). (2) “human thought and reason is sculpted, enhanced, and ultimately transformed” by technology (ibid.). Certainly it is one thing to view rationality as a capacity tuned to work with certain external technologies, and quite another thing that to say that this capacity is (partly) constituted by technology. If our faculty of reason is fitted towards such aids, it should exist beforehand. Conversely, if it is “sculpted and transformed” by those technologies, it probably does not exist beforehand. Claims (1) and (2) do not easily go together, at least not without much further ado.

Perhaps here there is a mixing of rationality as a faculty and rational processes. That reasoning processes are causally related to such technological aids, or even constituted by them, does not show that rationality as a faculty is constituted by these interactions. Indeed, it cannot: What constitutes any mental faculty is commonly understood to be expressed in terms of laws that guide the behavior of the capacity—in the case of rationality, the usual suspects include rules of logic, probability theory and heuristics. To explain the constitution of a faculty by the constitution of certain processes is simply a category mistake.

Against this, Clark might bring up his view that reasoning involves feedback loops between the potential for refinement through reflection on its own basis (Clark 2001, p. 134). Fair enough. When someone gives us a reason for a belief or a decision, it is common to ask back “Why?”, requesting the thinker’s grounds and principles. But these may always be questioned too, of course. Moreover, using technological aids or developing new notations (in mathematics or logic, say) has been an essential part of these feedback loops. This has led to the repeated
renovation of the laws of proper reasoning. Assuming that the faculty of reason is constituted by some fixed laws is therefore problematic. If rationality possesses no immutable nature, it may seem that the distinction between the faculty and processes of reasoning cannot be used to undermine Clark’s position. By truly engaging in such processes, we are caught in those feedback loops, and create the faculty anew, with no end in sight.

The problem with this rejoinder is that it is inconsistent with Clark’s focus on a purely descriptive, “mechanical” account of rationality. Insofar as we ask for better modes and principles of reasoning, we engage in genuinely normative issues. BR doesn’t face this problem because, pace Clark’s reading, it does not commit to such a purely descriptive account. In contrast, one of the most interesting features of BR is that it constitutes a new attempt for bridging Is and Ought, at least for certain domains of human reasoning (Gigerenzer and Sturm 2012). This bridging is, in part, possible because BR does not subscribe to BCI in the way Clark thinks it does.

To sum up: While Clark thinks there is some good in the BR approach, he also thinks it is too much committed to BCI. This is largely a misunderstanding of the program. Clark’s alternative position is expressed by claims that tend to be inconsistent. Now, before getting back to its relation with BR, let’s take a closer look at CE.

Cognition Embedded, Scaffolded or Extended?
The cognitive externalist wants us to treat cases of complementary integration of internal and external resources as a whole, and not to frame cognition as the working of an organic system that is just causally embedded in external resources. One thing that critics object is that whatever the explanatory benefits of CE are, you can get the same from internalism, so it wins, if only for parsimony concerns. To address this charge, we need to spell out the precise statement of what CE proposes for the putative cases in which environmental resources are involved in cognitive processes. There are three basic options:

(a) **Embedded cognition** takes the domain of complementarity-motivated putative cases of external cognition as showing the ways in which cognition is causally embedded in features of the environment that surround and supplement real cognition, which remains still located within the organism. Ontologically speaking, it is a claim of mere causal dependence. On the methodological side, the claim is that cognitive sciences should be mostly concerned in studying processes that take place within the organism, and not outside (Sutton et. al. 2010).

(b) **Scaffolded cognition** is the idea that (at least some of) our cognitive capacities both depend on and have been transformed by our manipulations of environmental resources. The claim here is not about mere causal dependence but about integrative coupling between internal and external elements. Accordingly, cognitive science should study these processes as they appear distributed across organism and environment, instead of isolating the internal (Sterelny 2010).

(c) **Extended cognition** claims that sometimes cognitive processes and systems are literally extended, having regions of the environment as proper parts located outside the organism. In other words, sometimes manipulated elements and structures of the environment, material or otherwise, constitute part of the cognitive system. This claim of constitution is held hand in hand with the urge for a revisionary attitude in the cognitive sciences towards the study of such extended processes without isolating its biological parts.

These views can be considered as stretches in a continuum, each best covering a different range of putative cases (Sterelny 2010, Sutton 2010). Adjudicating between the three options might be more a matter of degree and preponderance than a ‘winner-takes-it-all’ situation. This is so because the range of cognitive phenomena that motivates CE is heterogeneous. Among other dimensions of variation (Sterelny 2010, Wilson and Clark 2009) we can distinguish between (a) Individual artifacts, such as notebooks or sensory substitution devices, that given certain conditions might call for a genuine extended cognition reading à la Clark; (b) Collective resources, in which the external cognitive resources are embedded in a collective activity which involves several coordinated individual agents, as in Hutchins’ (1995) case study of the distributed processes that enable ship navigation; and (c) public resources, like symbolic representational systems.

The relationship between agents and public resources is best seen as a process of cognitive niche construction in which humans sculpt their environment so that it affords novel cognitive possibilities (Sterelny 2010). The real issue, then, is not the current synchronic location of the elements that constitute the cognitive system, but the integration between internal and environmental resources. Thus, “resources can be extended in the relevant explanatory sense even when they are not literally external” (Sutton et. al. 2010, p. 535). It is the manipulation of such resources and the transformative effect they have on the individual cognitive profile that provides the explanatory cornerstone.

The common internalist strategy is to conflate embedded with scaffolded cognition and contrast them both with extended cognition. They stress the difference between claims of dependence and claims of constitution. By insisting that the putative examples of scaffolded cognition involve a claim of dependence, they see them as grist on the internalist’s mill.

The way to resist this move is to highlight that, beyond ontological qualms, the idea of scaffolded cognition moves cognitive science in practice in the same line as that of extended cognition (Sutton et. al. 2010). Despite the skepticism that the ontological claims of extended cognition can bring about, when it comes to the explanatory tasks of empirical research, we take the most significant divide to be that between embedded cognition on one side and scaffolded and extended cognition on the other. That is the choice between cognitive internalism and cognitive externalism. In the next section we will address BR from the perspective here sketched.
Unpacking Bounded Rationality

BR tells us that simple heuristics make us smart by exploiting the environment. But, what does that mean? In order to further assess BR, we need to unpack its vague appeals to the environment.

The Execution of Reasoning

Let us focus first on the role that the physical environment has in reasoning. Some of the heuristics discussed by the proponents of BR require for their execution that agents are in current sensorimotor interaction with the relevant physical environment, although it is doubtful that they would qualify, within the broad realm of problem solving, as instances of genuine reasoning. A model example of this kind of heuristics is the gaze heuristic that people use to catch a flying ball: Fix your gaze on the ball, start running, and adjust your running speed so that the angle of gaze remains constant. “A player who relies on the gaze heuristic can ignore all causal variables necessary to compute the trajectory of the ball (…) the player will end up where the ball comes down without computing the exact spot” (Gigerenzer and Brighton 2009, p. 110).

The execution of these heuristics requires that agents are in current physical contact with the relevant environment. The agent relies on the manipulation of the environment to solve the problem. The baseball player alters the relative position of the ball in the egocentric space. This subset of heuristics, although not distinctively characterized in the BR literature, is best seen as epistemic actions. By acting on the environment itself, agents dispense with the need of otherwise required complex internal representations. As they alter the physical and informational structure of the environment, these processes take the agent closer to the solution; thus they are part of the agent’s processing of the problem (Menary 2010).

However, just as much of human reasoning, many of the heuristics analyzed by the BR program do operate decoupled from the environment. What role does the environment play in those cases? Consider a much discussed cognitive task: estimating the relative size of two cities. Two suitable strategies provide a good illustration of the paradigmatic kind of heuristic that BR puts forth.

The first is the Recognition Heuristic: if one of the options is recognized and the others are not, infer that the recognized alternative has the higher value on the target criterion. That is, if we recognize one of the cities, we ought to infer that it is the larger. Research suggests this is indeed what people usually do (Todd and Gigerenzer 2007).

The second is Take the Best (TTB); “To infer which of two alternatives has the higher value (a) search through cues in order of validity, (b) stop search as soon as a cue discriminates, and (c) choose the alternative this cue favors” (Todd and Gigerenzer 2007, p. 168). That is, we consider the available cues (airports, tourism, industry, universities, etc.) sequentially, in order according to the degree to which they correlate with the population size, stop at the first cue that discriminates and infer that the city favored by that cue is the largest. There is ample empirical evidence from behavioral studies that precise models of such heuristics better predict the subjects behavior in different task settings than optimizing models (Gigerenzer and Brighton 2009).

These heuristics typically operate upon internal representations. Constrained as algorithmic process models, all that the algorithm requires is supplied by either the contents of our memory (i.e. cues) or by effects produced by their recall (i.e. perceived recognition). Thus, here the environment does not play any direct role in the execution of reasoning. So, apart from those peripheral cases of epistemic action, the bulk of BR’s account of the execution of reasoning can be accommodated by internalism.

The Assessment of reasoning

Defenders of BR do not merely claim that our reasoning often relies on heuristics; they also argue that this is frequently for the better. One of the major findings of BR is that often, for a given environment, fast and frugal heuristics outperform more costly strategies in terms of accuracy (Gigerenzer and Brighton 2009). Hence the claim that heuristics are rational; only that their rationality is ecological. Ecological rationality defines the rationality of heuristics by the match between internal processing and the environment. But what is meant by environment, here?

First, there are always some constraints on the strategy being applicable. In the case of the recognition heuristic, one of the items must be recognized while the others are not. If this obtains, there are further features of the environment that will determine the performance of the heuristic. Plainly, the recognition heuristic will perform well if and only if the target criterion correlates with the recognition of the item. This is called recognition validity, and it constitutes the relevant structure of the environment for the performance of the heuristic.

The necessary condition for the applicability of TTB is that we have different available cues and we can rank them in order of validity. Then, the main structural properties of the environment that determine its performance are that (i) the more correlated the available cues are, the less it pays to take them all into account (and accordingly, TTB performs as good as or better than optimizing strategies that demand taking all cues into account); (ii) the more the cue validities vary, the more it pays to use a strategy like TTB; and (iii) the smaller the learning sample, the better lexicographic strategies like TTB pay.

What is relevant and needs to be specified in order to evaluate reasoning strategies is the structure of the available information as defined by properties like cue redundancy, variability of cue validities, or size of learning samples. That is the environment that most heuristics exploit.

Reasoning performance is determined by the match between strategies and informational features of the environment that determine the strategy’s relative success. Thus, the locus of ecological rationality is not only the internally processed algorithm, but the internal-algorithm-in-specific-environment complex. In this sense, the
environment does play a normative role in reasoning according to the BR program.

At this point, the internalist might object, “Fine, agents implement reasoning algorithms that operate upon stored information, and then it turns out that the performance of these algorithms is dependent upon how things are in the world. So what? This doesn’t necessitate CE. There is no close internal-environmental resource integration here.” However, this is not the whole story. Some of the features of the environment that play a normative role are not simply encountered out there. They are the product of the agent’s ongoing coupling with the environment. Consider the role of uncertainty. As Gigerenzer has argued: “the degree of uncertainty reflects the environment (ontic uncertainty) as well as the mind’s limited understanding (epistemic uncertainty); hence, the degree of uncertainty is located in the system mind-environment … redundancy and variability of cues depend both on what is in the physical environment and on humans who select certain cues and not others, which can result in more or less redundant cues. Physicians, for instance, tend to look at redundant cues, but they could choose to analyze independent cues and thus in part create their environment” (Gigerenzer and Sturm 2012, p. 257).

A narrowly internalistic interpretation of BR takes heuristics as one-shot games. But in order to understand how rationality emerges from the use of such strategies, it is of much importance to consider the way in which the agents’ behavior shapes the environment that shape the performance of their available reasoning strategies. Complementarily, the relative performance of a heuristic does affect its actual occurrence, albeit in an unexplained way. Part of the selection of a strategy operates upon internal representations, as memory constrains which heuristics can be applied. A second factor driving the selection of strategies is reinforcement by feedback. Still, beyond memory constrains and in absence of reinforcement, there is evidence that people rely on heuristics when they face those environments in which doing so pays off and not otherwise –i.e. relying on TTB when the validity of the available cues varies highly but not otherwise (Gigerenzer and Brighton 2009). This suggests that people are sensitive to the structure of the environment that determines the performance of heuristics.

The internalist can account for part of this story. It can explain how the heuristic strategy that the agent happens to apply is derived from a given structure of mental representations. It can also evaluate the performance of a strategy given a particular environment. It can even make inventories of heuristic-environment successful pairings. But, when it comes to deeper issues, such as why and how the agent chooses right and correctly applies a heuristic, or how and why certain environments have come to afford simple strategies, or how we have come to have such proficient heuristic tools, internalism seems to fall short. It faces the threat of falling into the “just happens” stance.

The emergence of normativity that results from the interaction between agent and environment is hardly the result of (only) internal computations. Whatever a much needed further investigation into these frontiers of theorizing delivers, we contend that, at least for the large domain of uncertain reasoning, it is highly plausible that normativity is due to the ongoing process of back-and-forth manipulation (of the environment) and transformation (by the environment) that CE aims to unravel.

The Enhancement of Reasoning

Next to the descriptive and the normative, BR also has a prescriptive dimension. Insofar as reasoning performance has to be assessed by the match between internal processing and environment (for some limits, see Sturm 2012), a dual perspective opens with regard to the enhancement of reasoning: we can either change what goes on within our heads, or change the environment. The development of both these prescriptive stances highlights a different point of connection between BR and CE.

A good example of the first line of prescription is fast and frugal decision trees for medical diagnoses. These are sequential trees designed for a very specific situation, like deciding whether a patient in the ER requires immediate attention facing a heart attack. They work as enacted protocols that the physicians must blindly follow. Interestingly, fast and frugal trees do not directly modify doctor’s internal capacity of processing. From the doctor’s point of view, they are an environmental resource, a cultural artifact they engage with in repeatable cognitive practices. It is part of the setting of the ER, upon which coordination between doctors is optimized. And indeed it has a transformative effect in doctor’s cognitive behavior –i.e. leading their attention to certain cues while ignoring others.

The second path for reasoning enhancement that BR pursues is typically illustrated by drawing attention to the alarming lack of competence of physicians to assess risk probabilities (Gigerenzer 2008). Several studies show that when given information in percentages, physicians often perform poorly. They could be further instructed not to neglect base rates. But, as Gigerenzer puts it, “in the ecological view, thinking does not happen simply in the mind, but in interaction between the mind and its environment. This opens up a second and more efficient way to solve the problem: to change the environment” (2008, p. 17). The proposed prescription is to provide information in natural non-normalized frequencies (that is, in terms of “n out of every 1,000 people have disease x. Of these n people, m will have a positive test, etc.”).

Interestingly, “the relevant part of the environment is the representation of the information, because the representation does part of the Bayesian computation” (Gigerenzer 2008, p. 18). The object of manipulation of these change-the-environment proposals is neither the material environment that plays a role in the execution of heuristics nor the information-structural properties of the environment that play a pervasive role in assessing the performance of heuristics. Instead, they point to symbolic and representational resources, which are thereby assumed to play a computational role. That is, cognitive practices carried upon the manipulation of symbolic structures. This

**Conclusion**

Although Clark’s characterization of BR is misguided, it is still true that BR’s account of reasoning does not fit much with the old-school notion of extended cognition. Its processes and mechanisms do not involve agents in stable and durable couplings with technological resources they trust. But, pace Clark, this does not mean BR is a dead end for CE, since CE needs not be bounded to those cases in which the environmental resources constitute a proper part of cognition. BR’s environmentalism does not consist in any ontological claim about the location and boundaries of the mind. It is a claim about explanatory variables and about normative criteria for understanding and improving human rationality. As such, it is best accommodated by the notion of scaffolded cognition.

From this moderate take on CE, we have shed light on the appeals to the environment made throughout the BR literature. We distinguished the role played by the physical environment in online heuristics qua epistemic actions, where we find instances of reasoning carried upon processes that functionally span through certain physical elements of the environment. But we also acknowledged that the execution of heuristics is typically decoupled from the real, mind-independent environment, and the account offered here by BR still fits with what a traditional internalist account would claim. Then we focused on the normative role played by the informational structure of the environment in determining heuristic performance, and on the role played by symbolic representations in the enhancement of reasoning performance. We believe that this conceptual clarification is quite a beneficial outcome of the whole debate for the development of the BR program.

Beyond that, we also pointed to some blind spots of internalism. The internalist can accommodate the actual processing of the heuristic, but it can hardly sustain the picture of rationality that emerges from the use of heuristics. Its narrow scope leaves important parts of the story on the dark, such as the emergence of the so-called “informational environment” and the co-enhancement process between agents and environments by which we can explain why we use the strategies we use and why is it that these happen to succeed beyond luck.

Those issues require further research that should be focused on the ongoing dynamic of interaction between agents and environments, and we contend that a moderate, scaffold externalism is the most fertile framework for this. In turn, we take these considerations to vindicate our take on CE, which acknowledges and incorporates the heterogeneity of cognitive engagements with the environment, and needs not be too much drawn by ontological claims, but rather to prove its explanatory value for cognitive science.

**References**


