teorema
Vol. XXXVIII/1, 2019, pp. 121-138
ISNN 0210-1602
[BIBLID 0210-1602 (2019) 38:1; pp. 121-138

From Queensberry Rules to Argumentative Theory: A Review of Mercier & Sperber's *The Engima of Reason*

Ira Noveck

RESUMEN

Esta recensión de *The Enigma of Reason* de Mercier y Sperber presenta partes de la historia reciente (y a veces personal) de la literatura sobre historia del Razonamiento con el objetivo de subrayar hasta qué punto es innovadora su teoría de la argumentación y proporcionar a la vez el telón de fondo para tres comentarios. El primer comentario aborda el punto de vista deflacionista de los autores sobre la realización de inferencias: se presentan las capacidades deductivas como algo tan común y corriente que no se pueden distinguir de otras intuiciones de orden inferior. Discrepo de esta caracterización y describo cinco líneas de investigación positiva para mostrar que la realización fundamental de inferencias ofrece regularidades en la conducta que las separa de otros géneros de inferencia discutidos en el contexto del Razonamiento. El segundo comentario se ocupa de la suerte de resultados experimentales que se predicen por parte de la Teoría de la Argumentación, lo que lleva a uno a preguntar, que es lo que haría a la teoría falsificable (específicamente ¿qué problemas de razonamiento se beneficiarían de la deliberación?). El tercer comentario aborda la noción de *explicaciones*. Los laboratorios de razonamiento han sido reticentes desde hace mucho tiempo a incorporar las explicaciones de los participantes a sus informes por razones filosóficas, empíricas e históricas. Los investigadores se beneficiarían si la teoría de la argumentación clarificase mejor qué tipo de explicaciones son beneficiosas para los propósitos experimentales.

PALABRAS CLAVE: *lógica mental, historia de la psicología, deducción, imaginería cerebral del Razonamiento, falsabilidad, predictibilidad.*

ABSTRACT

This review of Mercier and Sperber's *The Enigma of Reason* presents some recent (and at times personal) history of the Reasoning literature in order to underline how innovative their Argumentative Theory is and to provide the backdrop to three comments. The first comment addresses their deflationary view of deductive inference-making -- which presents deductive abilities as so run-of-the-mill that they are not differentiable from other lower-order intuitions. I take issue with this characterization and describe five strands of positive research showing that fundamental deductive inference-making affords regularities in behavior that sets it apart from other kinds of inference-making dis-

cussed in the context of Reasoning. The second comment is concerned with the sort of experimental outcomes that are predicted by Argumentative Theory, which leads one to ask, what would make the theory falsifiable (specifically, which reasoning problems would benefit from deliberation?). The third comment addresses the notion of *explanations*. Reasoning labs have long been reticent to incorporate participants' explanations into their accounts for philosophical, empirical and historical reasons. Researchers would benefit if the Argumentative Theory were to better clarify what sort of explanations are beneficial for experimental purposes.

KEYWORDS: *Mental Logic, History of Psychology, Deduction, Neuroimagery of Reasoning, Falsifiability, Predictability*.

As a graduate student in the 1980's, I was a member of one of the reasoning labs that Hugo Mercier and Dan Sperber mention early in the book. My labmates and I helped Marty Braine [see Braine & O'Brien, (1998)] develop his *Mental Logic* approach, which made the straightforward claim that there are content-free schemas of reasoning that are logical in nature. Lance Rips (1994) made a similar proposal. In both programs, *modus ponens* (*if p then q; p//therefore, q*) and *disjunction elimination* (*p or q; not-q//therefore, p*) were two of a handful of schemas that were proposed to automatically arise when the appropriate premise sets were made available. Overall, logical terms, such as *and*, *if* or *or*, were at the heart of each schema. The theory also featured a reasoning program that made predictions about the order in which certain inferences would be made (regardless of the order of the premises). The identified schemas were distinguished from other logical arguments. For example, *modus tollens* (*if p then q; not-q//therefore, not-p*) was not considered part of Braine's Mental Logic because it requires multiple steps (starting with *suppose p* as part of a *reductio ad absurdum*) and requires strategic thinking (making it more prone to error). This approach obviously held logical form as a feature of reasoning, even if it did not hold as lofty a place as it does in an Aristotelian program. Aside from Mental Logic there were two other major theories of reasoning vying for superiority at the time.

Adherents of the Mental Models approach (developed by Phil Johnson-Laird and Ruth Byrne) did not reject outright the import of logic in reasoning but placed a greater emphasis on the internal representations (the mental models) that allow reasoning to occur. Johnson-Laird and his colleagues [for a summary, see Johnson-Laird (2006)] argued that people use logical (as well as relational) terms to set up a mental model after which one can do deductive work; reasoners could then evaluate a conclusion by determining whether or not their internal model provides it.

While Phil Johnson-Laird and Marty Braine and their respective allies tangled, Jonathan Evans and his allies took a more pessimistic position by arguing that participants hardly paid attention to logical rules in reasoning tasks in lieu of non-logical features. According to Evans (1989), participants are primarily attracted to superficial features of a problem, such as the *matching* between a problem's premises and conclusion or the *believability* of a conclusion when evaluating syllogisms, rather than to the logical relations provided by the premises.

All three groups would box it out through outcomes on subtle variations (from a set) of reasoning tasks, many of which were originated by Peter Wason. The field of (the Psychology of) Reasoning, while plodding, was a lively if idiosyncratic corner of the cognitive world that inspired exceptional amounts of theory-making about rationality along with debates and, of course, data. Each group would provide its evidence to support its own approach and, when possible, each would present counterevidence against the other. In retrospect, none of the adversaries changed another's mind (although members of the next generation were arguably more flexible). Of course, the field did not remain focused on those three accounts of reasoning. As described in the *Enigma of Reason*, other theories — some of which were inspired by evolutionary concepts [e.g. see Cosmides (1989)] — would make prominent appearances.

No one in the Mental Logic and Mental Models camps, which were especially unreceptive to one another, anticipated that the folks defending superficial biases would eventually pull ahead. But that is exactly what happened. With the increasing influence of Kahenman and Tversky's approach in a neighboring discipline — i.e. probabilistic reasoning and decision making — along with their ever-growing catalogue of biases, Evan's heuristic and biases account of deductive reasoning — appeared to win out. The field of reasoning (viewed with a wide lens) increasingly emphasized, not how logical we are but, how susceptible we are to fallacies. Naturally, the view that logic, or other normative rules, can serve as a *source* of reasoning lost its exalted status in this corner of Psychology. To maintain any hope for rationality in reasoning, one would have to resort to a dual system in which one system is instinctively heuristical and at risk of violating normative reasoning and another is more reflective and potentially normative. This is arguably the most endorsed account of reasoning today [see Kahneman (2011)], which echoes those by Evans (2003) and Sloman (1996). This is the approach that Hugo and Dan find unsatisfactory and that leads them to develop their alternative.

It bears mentioning that, at least as far as my old lab was concerned, these debates generated scores, if not hundreds, of tasks investigating propositional reasoning, the Selection task, the THOG task, the 2-4-6 task, the Linda problem and many others (each of the tasks that I mention here can be found on the internet). I add that participants carried out these tasks by pen and paper while following the instruction to provide a one- or two-sentence-long justification for their response at the bottom of the page. Interestingly, we experimenters would look at these explanations and then summarily ignore them. Frankly, when I was a newcomer I didn't understand why we would do that when we were also aware that we would not report the justifications. The answer from the more senior members of the lab was that these justifications were requested so that subjects would take their task seriously and that is it. However, the better answer eventually became clearer to me, which was that it is common knowledge that justifications could not be counted on [Nisbett and Wilson (1977)]. In retrospect, this practice of dismissing justifications was at least partly due to a post-Structuralist suspicion of counting on subjects' immediate impressions. These sorts of data were disdained by Behaviorists and the new cognitive psychology was vigilantly avoiding it as well. The theory we were defending or confronting mattered most of all along with data that could be properly crunched. It also bears mentioning that we were aware, at least anecdotally, that one laboratory in Kansas had noticed that college students handled Wason's Selection Task better as members of a collaborative group than as individuals [see Moshman & Geil (1998)], but the potential of this curious fact was just that and also remained neglected.

The steady deflation of logic as an organizing (or prominent) feature of reasoning does not mean that deductive processes lost their luster among researchers. Interest in deduction certainly remains. In fact, some of the older debates have moved interestingly to neuroimaging, where mental logicians and mental modellers initially made different predictions about the neural structures that ought to support deductive reasoning. In my view, this initial debate has been fruitful for the neurological literature. In fact, much headway has been made with respect to both deductive reasoning *and* bias (perhaps not the *myside bias*) and the way these two features of reasoning interact in the brain. So, I disagree with some recent comments that Hugo and Dan made — during one of their recent interviews [Boyd (2018)] — in which they said that "the impact of neuroscientific methods has been much more limited so far" with respect to the study of reasoning. In fairness, Hugo and Dan were politely correct-

ing the interviewer's assumption that neuroscience was the only worthwhile form of empirical evidence to consider, but the upshot is that their comment does leave the impression that there are no reliable data to accrue from neuroscience, or more specifically neuroimaging. I will discuss these later. Deductive abilities are also topics for several developmental labs that highlight babies' abilities to detect logical relations; for example, one recently published study provides evidence showing how even preverbal infants (12-month-olds) can detect violations of disjunction elimination [see Cesana-Arotti et al. (2018)].

Why it is Relevant to Consider the Recent History of the Psychology of Reasoning

I share this extensive background for two reasons. One is that it helps explain the extent to which Hugo and Dan's theory (which I will not summarize here) aims to radically reshape the reasoning literature. By viewing reasoning through an evolutionary social lens while explaining how modules and representations figure into ***reason***ing, they have a) brought justifications or reasons out of the shadows and repurposed them and have; b) furthered the demise of the stance that views logical inference making as exceptional. While logical deductions may emerge through a modular inference generator, Hugo and Dan claim, these become prominent as a means of evaluating arguments. Logical deductions are not a feature of solitary "intellective" reasoning nor are they processed in a content-free manner.

The other reason I share this much background is that it provides a basis for my comments, which are threefold. First, unlike Hugo and Dan who despair of the lack of progress in reasoning [p. 48] and who see such a lack as a springboard for considering other approaches, I view the Reasoning literature more optimistically because it offers enough regularity so as to provide the basis for a solid cognitive science. The Reasoning literature has provided a plethora of robust effects — some showing that human beings are not normative reasoners and others showing an ability to carry out deductions — that need to be reconciled with theories and accounted for. Given the prominence attributed to the heuristics and failures of human reasoning (in the book and elsewhere), it does pay to emphasize what a content-free view of logical inference has yielded, which are robust findings showing the extent to which certain basic deductions are carried out automatically and reliably. This leads me to address Hugo and Dan's deflationary view of logical inferences, which sees

deductive inference-making as so non-unique that it is undifferentiable from multiple other reasoning processes. My second comment concerns ways to falsify the theory. While I am a fan (and in the interest of full disclosure, I have been privileged to see Hugo and Dan progressively develop this theory — albeit from afar — for about 10 years), I think it would behoove the Argumentative Theory to anticipate how the theory can be falsified, or delimited. The third comment addresses the notion of explanations, which is obviously central to their account. As I described earlier, labs have long been reticent to incorporate participants' explanations into their papers for deep philosophical, empirical and historical reasons, so it should not be surprising that reasoning researchers would remain cautious in treating explanations as an independent or dependent variable. I will describe how future investigations would benefit if the Argumentative Theory were to better clarify the borderline between explanations that are considered evaluable (the sort that are worth attending to, thus putting them at the center of deliberative ***reason***ing) and those explanations that do not count. While I think it is clear to Hugo and Dan how one would operationalize *explanations* as part of an experimental paradigm, it is less obvious to me (and I suspect to other old-time reasoning researchers), despite the book's numerous examples.

Preliminaries

Before starting on the more substantial portions of this commentary, I have to say that it is not an easy task to take a stand with respect to such an original theory. For one thing, while the theory addresses the Psychology of Reasoning, an area that has been around as part of modern cognitive psychology for 60 plus years (meaning there is a lot of ground to cover), the book points out that researchers tend to persist in their wrong-headed intuitions as part of a counter-productive (*myside*) bias. Now, if I defend a classical position (which I do in part here), I ostensibly become one of those people — described frequently in the book — who (refuses to change and) persists in sticking to some wrong-headed theory, making me "orthodox" (which sounds strange to me). I am aware that this charge can be easily applied to anyone who aims to defend at least some data that are aligned with traditional views. While this rhetorical tack is unfortunate because it puts me in an untenable position, I will charge ahead nevertheless.

Reports on the Death of Spontaneous Deductions Appear Exaggerated

One of the theory's claims is that inferences with logical import are so run-of-the-mill that they sit side-by-side inconspicuously with other low-level inferences. Hugo and Dan write [p. 166] that logic itself is "a heuristic tool" and that setting up a syllogism is a kind of idealization that accentuates logical relations and ignores all the features of thinking that we are typically contending with. Lower-intuitions, whether logical or not, inhabit our immediate impressions and it is only among higher-order intuitions, a kind of metacognitive ability, that logical features filter in. This approach, while very attractive, does not do justice to a lot of data, which show just how immediate and prominent logical relations jump out to participants. Just to be clear about how I approach their claims, I begin with a true story about wet floors that arose while I was preparing this piece.

One recent morning on a relatively dry day, I went out to our small courtyard and noticed that the floor was wet (to the point that there were a few banana-sized puddles) and that the water seemed to originate from the courtyard's faucet. Once on the ground, the water seemed to going down a slight incline to a drain two meters away. The courtyard's faucet was indeed dripping water but at a very slow pace, hardly fast enough to merit the large collection of water that I was looking at. I tried to resolve this quandary. Ultimately, I asked myself whether my son might have opened the faucet earlier in the morning (my wife was away so it could not have been her and my daughter is in University abroad, so she's no longer home). This was followed by my asking myself whether a stranger had opened the faucet (which would mean he or she entered through a locked door from the street). Before I knew it, I was making my way to my office to check on my computer (is it still there?). It was where I left it.

This reasoning "problem" was triggered by an event that I tried to resolve. This description fits with the Argumentative Theory because my quandary begins with intuitions that are generated by a conclusion (an excessively wet floor on an otherwise dry day). Note that of the four intuitions, glossed as (1a) through (1d) below, two (1a) and (1c) are deductive in character and the two others (1b) and (1d) are more inductive:

1. (a) If the ordinary slow-dripping faucet alone is the cause of the water on the floor, the puddles would not be so large.

   The puddles are indeed large.

   Therefore, the ordinary slow-dripping faucet alone is not the cause of the puddles.

(b) Someone must have opened the faucet (recently).

(c) It was my son or me.
I know it wasn't me.
Therefore, I'll ask my son (but I'm pretty sure it wasn't him).

(d) Did someone break in? Is there a sign of that?

Now, if I am understanding the theory and nomenclature properly, we make inferences (as readily as we breathe) and the contents of the inner speech (that I rendered explicit above) are intuitions that emerge through metacognitive activity, a "cognition about cognition", with respect to our capacity to evaluate our own mental state [page 65]. Importantly, these sorts of intuitions are not the result of a general processing mechanism but of a set of specific modular ones — such as a *protective papa* module that is on the prowl for predators and that is keenly aware of suspicious changes in the environment, a physics module (comprising knowledge about water and rates of evaporation) and, arguably, a mindreading module too. Importantly, the inferences in (1a) through (1d) at this point would not be considered reasoning. The series above becomes ***reason***ing when I use an intuitive reasoning module to justify my paranoia when I talk about my observations later. Until then, these reflections work side-by-side with other modules, on my representation. These remain just inferences in my head similar to those we use unconsciously all the time.

This is where I would like to pause. Note that the theory views logical inference-making as nothing more than perceptions and other immediate sensations. In light of the kind of work that has emerged from the Reasoning literature since the 1980's at least, this view strikes me as somewhat skewed. Can it be that disjunction elimination — implicitly carried out as I look solitarily at a few puddles of water — is so unremarkable in the panorama of human thought? Fundamental deductive inference-making, at least when part of comprehension and reasoning tasks, prompts much regularity and this ought to set it apart from other sorts of inference-making. Below I describe five regularities about deductive inference-making that make me doubt such a strong deflationary claim. All the examples below come from tasks — inspired in one way or another by the reasoning literature — that involve spontaneous deductive inference-making.

First of all, as far as comprehension goes, we — as researchers — can predict with confidence the rate at which logical forms are carried out correctly by ordinary adults regardless of language. The form in (1a) reflects *modus tollens*, which normally prompts correct performance at a rate of about 60% in straightforward propositional reasoning tasks and (1c) represents *disjunction elimination*, which more reliably prompts subjects to produce logical outcomes. Though not represented in my water-on-the-ground example, *modus ponens* — regardless of content — is probably the most reliable. While I am not familiar with work investigating a participant's confidence while evaluating various argument forms, it does appear that performance is linked to the cognitive effort each requires (as can be seen by the relative success in carrying out *modus ponens* compared to *modus tollens*).

Second, we know that, independently of specific content, participants generate logical inferences automatically as soon as two relevant premises are presented [Lea (1995); Bonnefond et al. (2013)]. For example, consider a case in which a participant reads about a story character who utters a conditional under his breath (2a) and, two sentences later, processes (2b):

2. (a) "If I can't find another shirt to wear on my date tonight," George thought, "then I had better find a needle.".

   (b) But George couldn't find another shirt to wear because all of his other shirts were dirty.

Under these circumstances, in which the pre-requisites for carrying out *modus ponens* are satisfied, associates to the word *needle* (e.g. *thread*) become activated more quickly compared to cases in which (2b) was replaced with (2b'), where the pre-requisites for *modus ponens* are no longer fulfilled:

2. (b') George realized that he had better figure out which shirt to wear soon because he was late.

This means that research has shown that deductive inferences emerge rather spontaneously (in tasks that capture subtle on-line processing).

Third, we can predict which collection of neural structures are reliably activated when carrying out deductions [for a meta-analysis, see Prado et al. (2011)]. This consists of a left fronto-parietal network that

prominently includes the left precentral gyrus, the left medial frontal gyrus, the left precuneus, and the left inferior frontal gyrus. By the way, one of those studies that supports this regularity was based on work that Dan and I started doing 20 years ago and that evolved only recently into an imaging study [Prado et al. (2015)].

Fourth, and relatedly, these neural areas that appear to support deductive inferences are impartial to logical form, meaning they are activated with propositional reasoning as well as with predicate reasoning [see Reverberi et al. (2012)]. The same holds for content. One study [Canessa et al. (2005)] provided two kinds of Selection Task (a social contract version and a more classical arbitrary version) and found that reasoning with both kinds of content prompted activity in the typical parietal-frontal areas when compared to control problems (with the social contract cases activating further areas, in the Right hemisphere). Deductions at one point were thought to be linked to language-support structures (due to the overlapping activity in the Inferior Frontal Gyrus), but some clever experimentation later showed that deductive activity can be distinguished from syntactic operations [Monti et al. (2009)].

Finally, deductive inferences can be distinguished from non-logical influences in neural studies (remember that logical inferences are assumed to operate side-by-side with non-logical inferences in the Argumentative Theory). We know that conditions that were designed to elicit biases among participants in a predicate reasoning task do not prompt patterns of activation that overlap with the reasoning network [Reverberi et al. (2012)]. We also know that right frontal areas (the Right Dorsal Lateral Prefrontal Cortex) are activated when there is a conflict between logical versus heuristic types of information. For example, Prado & Noveck (2007) used one of Evans's tasks to create cases where *matching* would come in direct conflict with the evaluation of a logical rule (consider the case in which the conditional rule *If there is an H then there is not a square* is tested with an exemplar consisting of an image showing an *H-in-a-circle*; the result is that: the two shapes do not match). This trial's true response is harder to derive than trials that show closer matches between the rule and exemplar (consider *If there is an H then there is a square* followed by the image of an *H-in-a-square*). The more mismatching the trial requires, the more one finds activity in frontal areas in the right hemisphere. This finding is consistent with other cases in which there are conflicts between deductive information and other low-level sorts of information, such as the belief bias [Goel & Dolan (2003)].

I have chosen these cases highlighting the automaticity and ubiquity of logical inferences along with their neural support, not only because they have become part of the corpus of work carried out by experts on the Psychology of Reasoning but, because they come from tasks that call for spontaneous deductions. Now, where do they fit in the Argumentative Theory? I think Hugo and Dan would agree that none of these cases reflect ***reason***ing (intuitions about reasons) as they described, which "consists in *attending to* reasons for adopting new conclusions" [p. 52]. If an experimenter asked participants why they chose "true" to an image of *H-in-a-circle* when provided *If H then not square*, they would probably be mystified by the question. It follows that these cases of deduction are more akin to *inference* because they involve the extraction of new information and across a wide range of circumstances. What module processes them? They arguably result from a linguistic comprehension mechanism (a "modular comprehension procedure that exploits, without representing them, relevance-based regularities in verbal comprehension", see p. 122). Note though that many of these cases are not your run-of-the-mill verbal situations because the relationships drawn by the experiment are arbitrary (see the last example linking letters and shapes). All of these examples involve some relatively low-level activity that hinges on the meaning of a logical word, such as *if*, and what is being observed as participants arrive at some answer.

Yet, these low-level inferences — which would have to be considered by Hugo and Dan as something akin to the perceptual inferences in illusions as they discussed — produce regularities with respect to outcomes in both behavioral and neuroimagery experiments. That is what makes it hard to view these deductions as run-of-the-mill inferences. They merit scientific attention on their own. They *afford* specific behaviors and reactions. This is why I hesitate to endorse the Argumentative Theory's strong deflationary view of logical inference.

On a related issue, I wonder what it would take to convince Dan and Hugo that these data argue against their deflationary account. It may be bracing to say that *x* (which could be anything from *modus ponens* to metaphor) is nothing special ("nothing to see here other than the mind at work, move on"), but how does one then ignore the overwhelming positive evidence indicating that *x* automatically prompts characteristic trains of thoughts. It's hard to know what criteria to adopt in order to accept this or any deflationary claim, unless relevant concomitant data can be accounted for in some other way. According to the more traditional view (that does assume that content-free disjunctions and condi-

tionals have status in human thinking), we could at least prove ourselves wrong. When we were defending Mental Logic, we emphasized that logical terms, such as *if*, *or, and* and *not*, are found in all languages and they appear to do the same thing universally. That has largely held up (even while keeping in mind the non-monotonic cases, which call for further explanation). At present, I see no reason to abandon these fundamental claims. Even if one treats the effects related to *if* as some form of conventional implicature [see p. 163], at the very least it appears to provide gains in information in a particular way that distinguishes it from other conventional implicatures (compare the word *but*, which prompts inferences that make it more than a mere conjunction, to the conditional which impacts on logical reasoning).

How Can the Theory be Wrong?

One of the strengths of the Argumentative theory is that its novel perspective provides new data and empirical predictions. As an experimentalist, this is what I find especially attractive. One class of prediction is that collaborative reasoning should generally lead to improved performance on Reasoning tasks and this has been demonstrated with, among other reasoning problems, Wason's Selection Task, the bat-and-ball problem, and the disjunctive reasoning task [see Trouche et al. (2014)]. In these experimental exercises that Hugo has pioneered, authors have been careful not to promise too much, which is a good thing. This can be seen in Trouche et al (2014), p. 1969, who wrote (about their set of experiments showing that collaborative reasoning led to improved performance):

> This is an important demonstration of sound argumentative competence, in line with the predictions of, for instance, the argumentative theory of reasoning [Mercier & Sperber (2011)]. However, it is clear as well that the conditions of demonstrability have to be well respected for this to be the case.

This is a good start, but it would be helpful, in my view, to go further. What are the general conditions for the theory to be predictive about deliberative processes in solving reasoning problems? Is it limited to a certain well-prescribed set of tasks? As far as I can tell, there seems to be a distinction to be made between, on the one hand, tasks in which heuristic biases seem to compete with normative information (like the Lawyer-Engineer problem), in which case the mere mention of one or the other (the base-rate or stereotype information) appears to prompt improved performance [see Obrecht & Cheney (2016)] and, on the other hand, tasks in

which there is a *working-out of* a normativizing rule (e.g. the Wason Selection Task). In the base-rate case, it seems to be a question of making relevant features salient and once a participant is reminded of, say, the importance of base-rates performance begins to improve. In cases like the Wason Selection Task or the Double Disjunctions riddle or the Bat-and-Ball problem, there is some computation that needs to be worked out, despite the fact that the relevant evidence might be hard to appreciate immediately. Not all reasoning problems are alike. Based on the distinction I just made, for example, one would expect deliberation to improve performance on the THOG task pretty readily because performance with this task (which asks participants to classify four arbitrary figures based on a rule concerning a figure's shape and color) would profit from having participants share, store and interact over some hard-to-track details. I am less clear what would happen on the Taxi-Cab problem, which involves recognizing base-rates while working out conditional probabilities. I am not sure what would happen on the Linda problem either; that problem requires participants to detect nested probabilities while also appreciating the problem-giver's intention. The book presented a Sudoku puzzle (a very difficult one that I was pleased to complete); are we to understand that (inexperienced) Sudoku players would not work better through collaboration because it involves discovering procedures? It would be useful to the theory (and the community of researchers interested in the theory) to make more precise predictions so that future experimentalists can prepare better experiments. This much would strengthen the theory.

When Do Explanations Become Reliable Sources of Data?

There is a rich strain of studies showing just how unreliable immediate explanations can be. My favorite examples of such phenomena come from split-brain studies where information is provided to eccentric areas of a participant's visual field so that the information is isolated as it is coded by each hemisphere. Consider work from Gazzaniga who would present two images — one at a time — to each hemisphere consecutively while asking patients to choose an associate — among four options — for each; such a patient's justifications incorporate the two chosen associates into a single (often humorous) post-hoc account. For example, in one trial described by Cooney and Gazzaniga (2003), the right hemisphere would be shown a snowed-in house and the left hand would choose (appropriately) a photo of a shovel; then the left hemisphere would be

shown a chicken claw and the right hand would choose (appropriately) a photo of a chicken shed. Once the two associated photos have been chosen (and are now available to both hemispheres), one split-brain patient justified her responses by saying that "the chicken goes with the claw and the shovel is to *clean out the chicken shed*" (my italics). The patient spontaneously invented a link between the two chosen pictures without realizing whence came the choice of shovel. Cooney and Gazzaniga (2003) write that "the human brain has a unique capacity to reflexively formulate causal theories about why events occur." I see this as support for the Argumentative Theory because it shows how scenes serve as conclusions and explanations are built (and often in an ad hoc manner), based on the limited resources applied.

With the advent of Argumentative Theory, one is asked, not to eschew justifications but, to take them seriously as variables. How is an experimentalist to manipulate these? I suppose one can view justifications as being on a spectrum, from reflexive to deliberative. Another way to appreciate justifications is more vertical, from lower-order intuitions to metacognitive higher-order intuitions before becoming the source of deliberation and proper ***reason***ing. While looking at the panorama of experiments that use explanations (in one form or another), it is not clear to me how to put proper sources of deliberation in play. Given a historical reluctance to rely on justifications in reasoning research, it would be helpful to current researchers to know the conditions under which the theory considers justifications acceptable (or not). One way to get a sense of good practices is through published work, but that does not provide clear guidance.

In one lovely paper that Hugo co-authored, Trouche et al. (2016) show that people are more vigilant to arguments that come from others as opposed to themselves. There, a normative logical argument (e.g. to recognize that "None of the apples are organic" entails "Some apples are not organic") becomes deliberative when the participant is confronted with that entailment as part of a reasoning task ("Apples are fruit, so if none of the apples (in a shop) are organic, at least some fruits (in that shop) are not organic") as having come from "a previous participant." This experiment operationalizes the idea that internal logical inferences (entailments) count as ***reason***ing when they are articulated as (or presented as a) part of a deliberative process in a second phase of the experiment. Does it follow that a justification does not count as ***reason***ing when the task simply asks a participant to provide one, as they are in the first phase of the Trouche et al. (2015) experiment? If ***reason***ing indeed

arises when a participant puts her requested explanation down on paper (or screen), we are back to the situation that reasoning researchers have been avoiding for decades because internal justifications are considered unreliable sources of information

Concerns about operationalizing justifications can be otherwise seen in the base-rate paper mentioned earlier from Obrecht & Chesny (2016). In that paper, "deliberation" is considered a factor when a condition presents a fictional character who presents an argument supporting the importance of base-rate information (in one condition) or of stereotyping information (in another) by adding a line of "explanation". This prompts two questions. First, does an explanation count as such because the task *refers to someone* who presents useful (or perhaps unuseful) information? To put it another way, would it no longer be a deliberative process if *task instructions* (which are essentially the words of an anonymous experimenter) provide an explanation? Is the providing of an interlocutor the crux of a deliberation manipulation? This leads to the second question, which is, "what counts as an explanation or argument in a reasoning experiment?" Does an explanation that merely reminds participants about a feature of the task count as an argument? That is essentially what is done in the Obrecht & Chesny paper. What if an explanation *indirectly* prompts a more normative resolution? In Baratgin & Noveck (2000), for example, we asked participants to rate the likelihood of whether a personality description corresponded with a Math teacher or a French literature teacher and we added cues to *complementarity*. Along with a standard control condition, we had one experimental condition in which we asked participants to provide two likelihood percentages per personality description — one for a judgement about the likelihood that the personality is a Math teacher and another for French literature teacher; in a second experimental condition we wrote "Please note that the sum of the two percentages must be 100%. For example, Anne is either a mathematics teacher or a French literature teacher". Each set of cues led to progressively higher rates of normative performance. Assuming we would get the same (or even more normative) results by introducing an interlocutor who presents these cues, would such indirect cues count as explanations?

I ask these questions so as to be clearer about what constitutes a deliberative process in a Reasoning experiment. As far as I can tell, the theory does not provide a researcher with the means to set up his or her own experiment that uses justifications as a variable. One experimenter's

(or one participant's) justification is potentially another's reflexive, dismissible causal theory.

Summary

Hugo Mercier and Dan Sperber's *Enigma of Reason* is refreshing. It gets the field of Reasoning out of an impasse brought on by dual systems accounts and it places reasoning more snugly and naturally in evolutionary theory. It keeps biases (that are now well known and accepted) in a prominent place while reconciling them with the emergence of normative thinking that, it is argued, arises through deliberation. It also underlines for all researchers the extent to which reasoning is a social activity. That said, I point out that the theory seems to throw the baby out with the bathwater when it portrays deductive inference-making as so low-level that it appears to ignore or downplay a swath of data showing that logical inference-making occurs with a regularity that can be captured both behaviorally and neurologically (as well as among babies). I think it is premature to claim that logical inferences are produced inconspicuously and unassumingly side-by-side with perceptual inferences. Furthermore, for the theory to get traction, I argue that it would benefit future researchers if the Argumentative Theory were to provide more specific predictions about the sort of reasoning problems whose solution would benefit from deliberation. Similarly, it would help practitioners of reasoning research if the theory were clearer about what counts exactly as a justification at each level of reasoning (from lower-order intuitions to higher-order ones to ***reason***ing). I suppose all this can be deliberated *viva voce* sometime soon.

Institut des Sciences Cognitives — Marc Jeannerod
CNRS — Université Lyon 1, UMR5304,
67 boulevard Pinel
69675 Bron CEDEX, France
E-mail: noveck@isc.cnrs.fr

ACKNOWLEDGMENTS

The author would like to thank Hugo Mercier for a lengthy conversation on Argumentative Theory and to Jean-Baptiste van der Henst and Jérôme Prado for their feedback on an earlier version.

REFERENCES

BARATGIN, J., & NOVECK, I. A. (2000), "Not Only Base Rates Are Neglected in the Engineer-Lawyer Problem: An Investigation of Reasoners' Underutilization of Complementarity; *Memory & cognition*, *28*(1), pp. 79-91.

BONNEFOND, M., NOVECK, I., BAILLET, S., CHEYLUS, A., DELPUECH, C., BERTRAND, O., FOURNERET, P. & VAN DER HENST, J-B. (2013), "What MEG Can Reveal about Reasoning: The Case of If… Then Sentences"; *Hum. Brain Mapp.* 34, pp. 684–697.

BOYD, B. (2018), "Interview with Hugo Mercier and Dan Sperber"; <*https://evolution-institute.org*>

BRAINE, M. D., & O'BRIEN, D. P. (eds.) (1998), *Mental Logic*; Psychology Press.

CANESSA, N., GORINI, A., CAPPA, S. F., PIATTELLI-PALMARINI, M., DANNA, M., FAZIO, F., & PERANI, D. (2005), "The Effect of Social Content on Deductive Reasoning: An Fmri Study"; *Human brain mapping*, *26*(1), pp. 30-43.

CESANA-AROTTI, N., MARTIN, A., TÉGLÄS, VOROBYOVA, L., CETNARSKI, R., BONATTI, L. (2018), "Precursors of Logical Reasoning in Preverbal Human Infants"; *Science, Volume* 359, Issue 6381, pp. 1263-1266.

COSMIDES, L. (1989), "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task"; *Cognition*, *31*(3), pp. 187-276.

EVANS, J. S. B. (1989), *Bias in Human Reasoning: Causes and Consequences*; Lawrence Erlbaum Associates, Inc.

— (2003), "In Two Minds: Dual-Process Accounts of Reasoning"; *Trends in cognitive sciences*, *7*(10), pp. 454-459.

GOEL, V., & DOLAN, R. J. (2003), "Explaining Modulation of Reasoning by Belief"; *Cognition*, *87*(1), pp. B11-B22.

JOHNSON-LAIRD, P. N. (2006), *Deductive Reasoning*; John Wiley & Sons, Ltd.

KAHNEMAN, D. (2011), *Thinking, Fast and Slow*; Macmillan.

LEA, R. B. (1995), "On-Line Evidence for Elaborative Logical Inferences in Text"; *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), p. 1469.

MERCIER, H., & SPERBER, D. (2011), "Why Do Humans Reason? Arguments for An Argumentative Theory"; *Behavioral and brain sciences*, *34*(2), pp. 57-74.

MONTI, M. M., PARSONS, L. M., & OSHERSON, D. N. (2009), "The Boundaries of Language and Thought in Deductive Inference"; *Proceedings of the National Academy of Sciences*, *106*(30), pp. 12554-12559.

MOSHMAN, D. & GEIL, M. (1998), "Collaborative Reasoning: Evidence for Collective Rationality"; *Thinking & Reasoning*, *4*(3), pp. 231-248.

NISBETT, R. E., & WILSON, T. D. (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes"; *Psychological review*, *84*(3), 231.

OBRECHT, N. A., & CHESNEY, D. L. (2016), "Prompting Deliberation Increases Base-Rate Use"; *Judgment and Decision making*, *11*(1), p. 1.

PRADO, J., CHADHA, A., & BOOTH, J. R. (2011), "The Brain Network for Deductive Reasoning: A Quantitative Meta-Analysis of 28 Neuroimaging Studies"; *Journal of cognitive neuroscience*, *23*(11), pp. 3483-3497.

PRADO, J., & NOVECK, I. A. (2007), "Overcoming Perceptual Features in Logical Reasoning: A Parametric Functional Magnetic Resonance Imaging Study"; *Journal of Cognitive Neuroscience*, *19*(4), pp. 642-657.

PRADO, J., SPOTORNO, N., KOUN, E., HEWITT, E., VAN DER HENST, J. B., SPERBER, D., & NOVECK, I. A. (2015), "Neural Interaction Between Logical Reasoning and Pragmatic Processing in Narrative Discourse" *Journal of Cognitive Neuroscience*, *27*(4), pp. 692-704.

REVERBERI, C., BONATTI, L. L., FRACKOWIAK, R. S., PAULESU, E., CHERUBINI, P., & MACALUSO, E. (2012), "Large Scale Brain Activations Predict Reasoning Profiles" *Neuroimage*, *59*(2), pp. 1752-1764.

RIPS, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. MIT Press.

SLOMAN, S. A. (1996), "The Empirical Case for Two Systems of Reasoning"; *Psychological bulletin*, *119*(1), p. 3.

TROUCHE, E., SANDER, E., & MERCIER, H. (2014), "Arguments, More Than Confidence, Explain the Good Performance of Reasoning Groups"; *Journal of experimental psychology. General*, *143*(5), pp. 1958-1971.

TROUCHE, E., JOHANSSON, P., HALL, L., & MERCIER, H. (2016), "The Selective Laziness of Reasoning"; *Cognitive Science*, *40*(8), pp. 2122-2136.