May 22nd, 9:00 AM - May 25th, 5:00 PM

Commentary on: Mark Battersby and Sharon Bailin's "Critical thinking and cognitive biases"

Frank Zenker
Lund University, Department of Philosophy and Cognitive Science

Commentary on: Mark Battersby and Sharon Bailin's "Critical thinking and cognitive biases"

FRANK ZENKER

Department of Philosophy and Cognitive Science
Lund University
Kungshuset, Lundagård, 22 222 Lund
Sweden
frank.zenker@fil.lu.se

1. INTRODUCTION

Battersby and Bailin's paper provides an overview and a discussion of a range of fallacies, or reasoning errors (owed to various "biases"), identified in experimental cognitive psychology. The main claim, if I understand it correctly, is that this psychological research *can* be of pedagogical relevance to those who instruct students in critical thinking (CT). Specifically, this research may assist in developing strategies that allow avoiding the kinds of errors regularly reproduced in experimental settings both in the lab and outside (aka "reasoning in the wild").

Battersby and Bailin provide examples of fallacies/errors that are relatively new by the standards of "classical" critical thinking instruction (e.g., loss aversion, framing, anchoring, over-confidence). Moreover, they assert, cognitive psychology appears to explain—at least to some extent, and in some sense of 'explain'—not only why such errors are reliably reproducible in experimental settings, but also why they are *persuasive*: "these errors are grounded in natural reasoning processes" (p. 3). Their paper evidences a current trend which seeks to appropriate research results originating in *a particular strand of* cognitive psychology—the *Heuristics and Biases* program, significantly driven by the research of Amos Tversky and Daniel Kahneman—into argumentation studies and critical thinking instruction.

This commentary will likewise stress both the importance of this research and its relevance for CT instructors and argumentation theorists. If an error in reasoning and argumentation can be assumed to be "owed to" a wide-spread and "natural" reasoning process (in the sense of "first nature"), then there is *prima facie* reason to believe that such errors can be avoided through some form of prescriptive intervention. And there is reason to believe that offering an explanation why these errors are persuasive might help students to "put the brakes on our tendency to rush to inference under certain circumstances" (p. 7). Yet, this is at best part of the full story. To speak meaningfully of biases, or errors, some *normative standard* must be assumed as correct. That assumption, as Sect. 2 argues, is not so innocent.

Further, I have a mixed attitude about the current trend, which appears (to me) to run the risk of adopting a one-sided, and a somewhat "dumbed-down"—or popularized—version of the *full breadth* of the relevant psychological research. Empirical results and more conceptual considerations originating in the "*ecological rationality*"-research program (Gigerenzer & Sturm, 2012) cast doubt upon the validity status of some of the claims, assumptions, and methodologies endorsed on

Mohammed, D., & Lewiński, M. (Eds.). *Virtues of Argumentation. Proceedings of the 10th International Conference of the Ontario Society for the Study of Argumentation (OSSA), 22-26 May 2013.* Windsor, ON: OSSA, pp. 1-6.

the Kahneman-side of the Heuristics and Biases program (see particularly Samuels et al., 2002). Battersby and Bailin's paper stays clear of increasing this risk but, as Sect. 3 argues, does not reduce it.

Finally, it is less than perfectly clear which *de-biasing strategies*, or generally which forms of critical thinking pedagogy, are optimal or even sufficient in order to reliably overcome biases *in the contexts in which they in fact result in errors*. The thesis in Sect 4 is that our perhaps most ubiquitous form of instruction—informally known as the technique of "show and tell"—may not only fall short, but may even have adverse effects.

1.1 Disclaimer

To be clear, and because the below may be misinterpreted, I see no disagreement between Battersby and Bailin's position and my own. The effects they cite have been well-established in experimental studies, and are by now widely known (though perhaps not within "classical" argumentation theory) (see, e.g., Politzer, 2004); I do not doubt that these results were obtained, I take issue with their interpretation. Further, the idea that such effects are ubiquitous since they result from natural reasoning processes—and so are *persuasive because natural*—appears to be entirely plausible; yet I doubt that this constitutes an *informative* explanation. Finally, I also agree that the challenge for instructors consists in "help[ing] students to see [a slow mode of] thinking critically as being worth the mental effort" (p. 8); yet I doubt that cognitive psychology currently has much to offer in terms of tested strategies that could assist CT instructors.

I hope that the following will be understood not as an attempt to downplay the importance of Battersby and Bailin's paper, but as an attempt to bring out the complexities often played-over when cognitive psychologists, and these days also argumentation scholars, engage with (the literature on) cognitive biases.

2. NORMATIVE STANDARD

It goes almost without saying that *some* normative standard of correctness is required to render the term 'error' meaningful. Take mathematical reasoning, for instance. By the standard of Peano arithmetic: "1 and 1 is 2"; by the standard of binary arithmetic: "1 and 1 is 0"; and by the standard which my nice endorsed around age five: "1 and 1 is 11." So, to claim that human are systematically prone to err in reasoning (and in being persuaded by arguments grounded in such reasoning) must presuppose that *exactly one* standard is normatively correct in some context. By and large, the normative standard is that which subjects' performance reliably *deviates from* when they solve experimental reasoning tasks of the kind typically put to them by experimental cognitive psychologists.

As Cohen (1981) and several later authors (e.g., Evans 2008) have pointed out, for any reasoning task in experimental psychology, sound methodology presupposes the experimenter to have *previously* selected the correct normative standard. Moreover, the correct normative standard, S, chosen for such tasks tends to be co-extensive with that endorsed in recognized theories of decision making.

Normally part of our best current *expert knowledge*, then, S is embedded within a theoretical framework under idealization and simplification. Such theories, and the standards they recommend, we may say, are formulated against the backdrop of an idealized environment, or an idealized context.

Natural environments or contexts, in contrast, are *uncertain* rather than risky. And they are populated by *bounded* agents, i.e., agents who lack full insight, information, time, etc. Bounded or not, moreover, agents may *contingently* not yet be acquainted with whichever normative standard experimenters call "confirmed" (or "instantiated") when they observe subjects exhibiting the "rational response" to (or the "rational choice-behavior *vis-à-vis*") a reasoning task, T.

In fact, in some contexts similar but not identical to the lab setting, there can be good reasons to *inhibit* deploying standard S, although S remains the correct standard *vis-à-vis* T in context C. Much here depends on the phrase "in some contexts." It is a placeholder for the conditions (some, but not others, of which are by now better known) under which our "natural reasoning and decision making-inclination" (aka System 1 reasoning; see Evans 2008 for a critique of current assumptions), *vis-à-vis* task T, leads to a result, R*, that is *at least as good*—and sometimes, in a special sense, *better*—than R, where R is the results one would obtain by deploying the experimenter-chosen S *outside the lab* (see below).

When much depends on such conditions, it is *irresponsible* to assume—and to teach to others—that psychological experiments demonstrate, come what may, that a majority of subjects err *whenever* they choose in ways that are not licensed by the normative standard selected as correct. Rather, whenever subjects do deploy an alternative standard—let's assume that *some* standard is always deployed—, then they behave incorrectly *in the experimental setting*. But that says rather little!

One would, or so I presume in the following, like it to come out as analytical that *errors of reasoning* are committed, *if and only if* deploying the reasoning standard S* *vis-à-vis* a reasoning task T in context C produces a result R* (aka a decision) that deviates significantly from a result R, where R—ultimately obtained via expert knowledge—is licensed by standard S as the correct result *vis-à-vis* T in C.

3. ECOLOGICAL RATIONALITY

The discussion on whether subjects do in fact understand the experimenter's instructions exclusively concerns the correct identification of the task, T. Kahneman, for instance, does on occasion express slight bewilderment at subjects applying, say, a representative heuristics, although the task explicitly uses the term 'probability' (and avoids the term 'representativeness')—catchphrase: "they solve the easier task." It stands to reason that such results too often fall back upon the experimenter (Cohen, 1981; Kuhn, 1962).

This may for instance be the case with the infamous *Linda problem* (the presentation of which always comes backed up by reports that subjects agree with the normative standard selected as correct, *once it has been explained to them*). Strangely, however, subjects seem to become better—they appear more inclined to "avoid the error" (perhaps by inhibiting deployment of S*)—whenever a task apparently equivalent to the Linda problem is formulated in terms of frequency.

Similar things hold for tasks that mirror the logical structure of Wason's selection task, but employ a social setting (so called cheater detection task, see Cosmides, 1989, Cosmides & Tooby, 1992). And similar things again hold for the Bayes rate fallacy (Cosmides & Tooby, 1996). So, such cases are not straightforwardly errors unless subjects in fact understand the task instructions in ways that make *only* T, but not T*, the identified reasoning task to be solved.

It appears doubtful (to me) to assume that T—for which, recall, *vis-à-vis* context C deploying standard S is assumed to be correct—*can* be understood unequivocally by subjects, unless (the deployment of) standard S is *in fact* at their disposal. If so, then the deployment of the correct normative standard S to T *vis-à-vis* C presupposes the correct identification of T by subjects. If *ought* shall imply *can*, however, then any hypothesis on the frequency of subjects' expected deployments of S (*vis-à-vis* T in C) that stipulates a *non-zero* frequency must likewise presuppose S to be deploy*able*; so S must be accessible to subjects.

Moreover, whenever deploying S* *vis-à-vis* T leads to a result R*, but if R* does not (significantly) deviate from R, then deploying S* should not count as an error either. At best, deploying S might here be a process-mistake. The important difference is simply that S* did *not* deliver an erroneous result.

But when a choice between two different reasoning standards (which, respectively, adhere to standards S and S*) is *neutral* with respect to the result obtained—*vis-à-vis* T in C—, then it is *prima facie* plausible to assume that our best frequency measure of 'deployments of normative standards (such as S or S* *vis-à-vis* T or T*) in a C context' (i.e., in the lab) will register values not far from those for such deployments *in non-C contexts,* i.e., in "the wild". One may now perhaps more readily appreciate the significance of the frequency distribution of T and T* tasks, and of the deployments of S and S* standards, *in the wild*. By and large, in the wild it is difficult to find equivalent ones for the kinds of tasks put to subjects in the experimental context, C. Put differently, experimental tasks suffer from a lack of naturalness *in most contexts but* C.

Given the above, the perhaps most important contribution for a pedagogy of critical thinking may be the insight that deploying S* may be rational when contextual constraints keep S from being applicable. In Battersby and Bailin's paper, this insight can be discerned in the phrase "put the brakes on our tendency to rush to inferences *under certain circumstances*" (p. 7; *italics added*). Effectively, when S is not applicable, then task T cannot be meaningfully raised, making task T (what might be called) *S-non-solvable* in such contexts or environments. In such cases, the deployment of S* to T* (instead of S to T) *is* ecologically rational, and it is rational *simpliciter* when R* is at least as good a result as R, but when deploying S* is also less costly in terms of resources than deploying S.

As an immediate consequence, it would be misleading (to say the least) to teach "psychological insights into human biases" without stressing contextual constraints. Whenever some reasoning experiments demonstrate that, in context C, S* is predominantly applied *vis-à-vis* (instructions formulated to convey task) T, then one might ask how such results bear on non-T tasks in non-C contexts, especially on non-T tasks relevantly similar to T *vis-à-vis* non-C contexts relevantly similar to C.

To apply the above to an example, consider *anchoring*. The term describes agents' tendency, exposed in experiment, to be influenced by information most recently (and in some case initially) received, e.g., in the form of task instructions, *although* experimenters select this information so so that it is irrelevant for *this* task in *this* context. In such cases, the allegedly "rational" response amounts simply to discount the evidence, i.e., to choice-behave as if *no evidence* had been received (in our parlance: to *inhibit* the deployment of S*)—whatever this means.

One can now inquire into the frequency of non-C contexts in which such discounting is ecologically rational. In a bar-context, for instance—where forms of "bullshitting," or "pulling each other's leg" are normal—, subject's might frequently discount evidence. But what reason is there to suspect that, in the experimental context C, subjects will (with significant frequency) behave in ways one would far more readily expect in a bar context? There seem to be few grounds to expect as much. After all, in non-C contexts that bear out a greater similarity to the experimental than to the bar context, information received can normally be discounted only for good reasons.

Each *observed* distribution of subjects' choices (in response to T *vis-à-vis* C) remains of course informative, because it can always be compared to the entire range of possible distributions. But, although measuring the frequency of deployments of S* to T *vis-á-vis* C is thus informative, this also yields a measure of the dis-similarity of a C context (the experiment) from non-C contexts (such as the bar) in which S* is ecologically rational. So the explanation of subjects' behavior would be that subjects did perceive C as being *too similar* to a context in which to deploy S* is an ecologically rational response.

This, then, may finally serve to explain (to some extent) why certain "biases," or "reasoning errors" are persuasive, i.e., have acquired a certain naturalness: they exhibit the ecological rationality of deploying S* to T* in non-C contexts. Note, however, that this explanation arises from comparing experimental and natural contexts. Moreover, the explanation depends on C being an ecologically rational choice in similar contexts, and emphatically not an error. It will, I suppose, be difficult to find this message stated very clearly in the work of those who have been inspired by Tversky, and perhaps more so Kahneman.

Finally, it appears (to me) to be a rather non-informative explanation to stipulate that deploying S* is persuasive because it is natural to deploy S*.

4. DE-BIASING STRATEGIES

The implications of the foregoing for the development of de-biasing strategies and critical thinking pedagogy may now be discussed perhaps more briefly. When it depends on contextual information whether a reasoning error is what the term suggests—information which one *must* "factor into" a fallacy-judgment—, then we cannot hope to improve learners' reasoning abilities, *much less so in general* (i.e., across all contexts), unless we can point out *very specific reasons* why it is S that should be deployed *vis-à-vis* T in C, but it is S* that *should* (or *might as well*) be deployed *vis-à-vis* T* in C*. Normally, what speaks for S *vis-à-vis* T in C (the lab) is that S is the *only* correct solution for T in C.

Leave the experimental setting, and things are immediately less clear. Unless we instill in our students the requisite sensitivity towards contextual constraints, we may expect that they will develop the same attitude towards biases as they normally develop towards fallacies. Their names are seemingly too frequently used for the sole purpose of "throwing" them in the face of adversaries in order to gain a perceived advantage in the kinds of contexts in which humans typically want to appear smart, to others and to themselves.

To avoid misunderstanding, *there are* errors of reasoning in the sense that the following necessary conditions must obtain: (i) solving T requires deploying S in C; (ii) T is recognized by the agent, and (iii) S is at the subject's disposal; (iv) S is not deployed, but S* is; (v) the result R* obtained by deploying S* deviates significantly from R. To teach anything less complex is, in my opinion, intellectually irresponsible.

REFERENCES

Battersby, M., & Bailin, S. (forthcoming). Critical thinking and cognitive biases. In D. Mohammed, & M. Lewiński (eds.), *Virtues of Argumentation*. Proceedings of the 10th International Conference of the Ontario Society for the Study of Argumentation (OSSA), 22-26 May 2013 (pp. 1–9). Windsor, ON: OSSA.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, *4*, 317–370.

Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped how humans reason? Studies with Wason Selection Task. *Cognition*, *31*, 187–276.

Cosmides, L., & Tooby, J. (1992) Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides & J. Tooby. (eds), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford University Press: New York.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Evans, J. St. B. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*, 255–78.

Gigerenzer, G. & Sturm, T. (2012). How (far) can rationality be naturalized? *Synthese*, *187*, 243–268.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago, Ill.: CUP.

Politzer, G. (2004). Rationality and the experimental study of reasoning. In P. Bourgine & J.-P. Nadal (eds.), *Cognitive economics: an interdisciplinary approach* (pp. 79–93). Berlin: Springer-Verlag.

Samuels, R., Stich, S., & Bishop, M. (2002). Ending the rationality wars. In E. Renee (ed.), *Reasoning and Rationality* (pp. 260–268). New York: Oxford University Press.