Notations in logic: countering belief bias with extended cognition

Abstract

In an much-quoted passage from the preface of the *Begriffsschrift*, Frege seems to be claiming that human agents are prone to letting prior beliefs and tacit presuppositions interfere unnoticed in reasoning processes, and that a suitable notation can act as a counterbalance to this tendency whenever required, in particular in the process of formulating mathematical proofs. These are empirical claims which are not treated as such by Frege, but recent results in psychology and cognitive science allow for an empirical assessment of these claims. On the basis of work by Stenning, and Landy & Goldstone, I shall argue for the thesis that, viewed from the point of view of extended cognition (in Clark's sense), notations in mathematical and logical contexts incite a form of bodily engagement that relies on perceptual and motor functions. Moreover, I shall argue that it is precisely this *externalization* of reasoning processes that offers a counterbalance to some of our spontaneous reasoning patterns, belief bias in particular (as investigated by Evans and Stanovich, among others). Thus, I conclude that Frege's observations may be seen as vindicated by these empirical results.

Personal overture

I first met Martin Stokhof as a stressed master of logic student at the ILLC, in 1999/2000. I was struggling to keep up with the highly technical courses I was taking, and it was extremely important to take Martin's now legendary course 'Hermeneutics and forms of life' in order to recover confidence in my academic capabilities ("Ok, so I *am* good at *this*" – 'this' being philosophy). It was during one of his classes that I had an eureka moment and decided to write my master thesis on medieval logic (a decision with long-lasting consequences!). After I completed my master degree and went on to do a PhD in Leiden, we weren't much in touch, but I'm pretty sure he was keeping track of my progress. Then, back in 2005, when I was thinking about putting together a research project about formal languages (essentially a spin-off of the last chapter of my PhD dissertation), I corresponded with Martin briefly about a paper he was writing ('Hand or hammer?'), and he asked to see some of my material. When he later sent me a draft version of his paper, I saw that it was full of references to my own draft, and that was another eureka moment: I was going to work on a project on formal languages with Martin in Amsterdam! (Fortunately, NWO (our big academic sugar mamma in the Netherlands) also thought it was a good idea.)

Thus, Martin became my 'boss' during the four years (2007-2011) of my VENI-research on formal languages at the ILLC. He was very surprised when he first heard that I referred to him as 'my boss', and this says a lot about Martin's personality and mentoring style; he is modest and gentle, and more importantly, he never tells us 'pupils' what we should do. He wants us to discover for ourselves where we should be going, but with the best guidance and support one can hope for. He never tries to impose his own interests or viewpoints, but rather compels us to find our own (through a Wittgensteinian-Socratic method of constant questioning!) and to follow the path that is best for us (which may not

be in his own best interest). In short, he works with his apprentices from within, not from without.

With the paper below, I want to showcase Martin's 'therapeutic' supervising style. Here's the story: initially, the plan was to adopt a 'Wittgensteinian' approach in my analysis of uses of formal languages in logic, and this was very much in sync with Martin's own interests. But then, more or less a year down the road, I stumbled upon the psychology of reasoning literature, and fell madly in love with it. It became clear to me that an empirically-informed approach was precisely what was required for me to find answers to the questions haunting me: Why and how do formal methods 'work'? What's the magic? Martin was of course already sympathetic to empirically-informed approaches in philosophy generally speaking, but for a while I was a bit concerned that he might think I was 'betraying' the original plan of using mostly a Wittgensteinian framework as the background for my research. So I asked him to have a meeting to discuss this change of plans. With hindsight, it is clear that I had nothing to fear, but I was still a little worried that he might think I was letting him down (I'm sure there is a lame Freudian explanation for that!). I see that meeting as an absolutely crucial turning point in my career; he convinced me that I was not a 'one-philosopher' kind of person, that the questions I was interested in were essentially systematic, and that I should pursue them. Moreover, he was very supportive of the idea that an empirically-informed perspective was the right way to move forward (in fact, it *is* a Wittgensteinian perspective, at least on Martin's reading of Wittgenstein). So I was ready to set sail and go explore uncharted areas, with Martin's gentle guidance on the background. The paper below offers a summary of my main findings in this period, and illustrates my path starting from a philosophical question, venturing into psychology and cognitive science, and going back to philosophy to formulate tentative answers. None of this would have been possible had Martin said 'No' at that meeting – although a possible world where he says 'No' at such a meeting is highly unlikely, as is a world where I don't do exactly what I set my mind on doing! Still, his approval was very important.

All in all, my interaction with Martin was certainly one of the very best things of my VENI years in Amsterdam. It was always ok for me to drop by and give him a short summary of my latest adventures and accomplishments. While he and I most definitely have very different personalities (I'm more of the loud, pushy kind), I was always under the impression that he actually rather enjoyed my 'upbeatness', and I'm thankful to him for his patience. I've learned an awful lot of philosophy from Martin, and more importantly, I've become a better philosopher through his gentle guidance. The main thing, however, is Martin's caring, empathetic, ethically-inspiring way to go about in life, which sets an example for us all. I wish I could be his apprentice forever.

In the preface of the *Begriffsschrift*, Frege offers the following motivation for the introduction of his 'ideography':

> To prevent anything intuitive from penetrating here unnoticed, I had to bend every effort to keep the chain of inferences free of gaps. In attempting to comply with this requirement in the strictest possible way I found the inadequacy of language to be an obstacle; no matter how unwieldy the expressions I was ready to accept, I was less and less able, as the relations became more and more complex, to attain the precision that my purpose required. This deficiency led me to the idea of the present ideography. Its first purpose, therefore, is to provide us with the most reliable test of the validity of a chain of inferences and to point out every presupposition that tries to sneak in unnoticed, so that its origin can be investigated. (Frege 1879/1977, 5-6)

He seems to be making two crucial claims in this passage: i) that human agents are prone to letting prior beliefs and tacit presuppositions interfere unnoticed in reasoning processes, in particular due to features of 'ordinary' language, and ii) that a suitable notation can act as a counterbalance to this tendency whenever required, in particular in the process of formulating mathematical proofs. These are empirical claims which are not treated as such by Frege, which is not surprising given the scientific and intellectual background at the time. Now, however, recent results in psychology and cognitive science allow for an empirical assessment of these claims, and this is the purpose of the present contribution.

Beyond Frege's specific goals with his ideography, the present analysis concerns the crucial but burgeoning topic of the cognitive impact of external devices such as notations when used in reasoning processes. I shall argue that Frege's two claims can both be shown to be justified in light of empirical results. As for i), the literature on belief bias and related concepts supports the idea that human agents do have a strong tendency towards relying on prior, external beliefs in reasoning processes, thus systematically letting presuppositions 'sneak in'.[1] As for ii), recent work in cognitive science suggests that notations in mathematical and logical contexts incite a form of bodily engagement that relies heavily on sensorimotor functions. I shall argue that it is precisely this *externalization* of reasoning processes (in the sense of extended cognition and the reliance on external devices for reasoning) that offers a counterbalance to some of our more spontaneous reasoning patterns, belief bias in particular. As suggested by Frege, notations deeply transform the process of logical inference.

---

[1] But I shall not discuss here whether this is mainly due to features of everyday life language, as Frege suggests. Moreover, notice that, for Frege, the concern with letting presuppositions sneak in does not primarily regard possible reasoning mistakes (as discussed in the belief-bias literature): for him, the intuitive moves that mathematicians make in the course of reasoning are usually correct. It is his *logicism* that leads him to worry about intuitive inferential moves. (I owe this point to D. Macbeth.)

1. Belief bias

In the last decades, an important topic in the psychology of reasoning has been the so-called 'reasoning biases'. The terminology is not entirely felicitous, as it suggests the notions of error and cognitive vices, but it is not clear which normative standards can and should be used to characterize certain reasoning tendencies as 'mistakes'. At any rate, subjects typically deviate from the 'normative responses' (that is, according to the traditional deductive canons) in experiments with deductive tasks (Evans 2002); but the 'mistakes' they make are not random, rather they seem to reflect systematic patterns, and these became known as 'reasoning biases'.

A very pervasive reasoning pattern can be described as the tendency to reason towards confirmation of the beliefs one already holds. This is observed for example in experiments where subjects are asked to evaluate the correctness of (syllogistic) arguments; typically, they "will tend to endorse arguments whose conclusions they believe and reject arguments whose conclusions they disbelieve, irrespective of their actual validity" (Evans et al. 1983, 295). This phenomenon is referred to as 'belief bias', and the experimental results are very robust, having been replicated several times. In (Evans et al. 1983), for example, participants were presented with syllogistic arguments and asked to evaluate their validity. The arguments were of four kinds: valid arguments with believable conclusions, valid arguments with unbelievable conclusions, invalid arguments with believable conclusions and invalid arguments with unbelievable conclusions. The general results of the experiment were the following (percentage of arguments accepted as valid):

	Believable conclusion	Unbelievable conclusion
Valid	89	56
Invalid	71	10

Notice in particular that *in*valid arguments with believable conclusions are more often endorsed than valid arguments with *un*believable conclusions. Similar results have been obtained in conclusion production tasks as well (Oakhill and Johnson-Laird 1985). Several competing theories of reasoning (e.g. mental models theory, dual processing theory) have offered accounts of the phenomenon, but they all agree on the ubiquity of this cognitive tendency in human agents. It seems to be a particular case of what K. Stanovich (2003) has described as a 'fundamental computational bias' in human cognition: "the tendency to automatically bring prior knowledge[2] to bear when solving problems. […] this tendency toward contextualizing problems with prior knowledge is so

[2] Technically, it is not a matter of *knowledge* as philosophers understand it, i.e. as involving factuality, but rather a matter of *belief*.

ubiquitous that it cannot easily be turned off" (Stanovich 2003, 292/3) (see also (Goel & Dolan 2003)).

Thus, the question naturally arises as to how, if at all, our tendency to bring prior belief to bear can be compensated for, even if only temporarily. It is clear that, at least in *some* reasoning contexts, the deployment of prior belief can be disadvantageous. This is particularly the case in purely deductive settings such as mathematics and logic; one of the hallmarks of deductive reasoning is precisely that all external, implicit information must be kept out of the reasoning process. When reasoning deductively, one should only rely on information that is explicitly on the table, the premises; but as well described by Frege, the pull towards what is intuitive and implicitly assumed is very strong.

Some recent studies on the inhibition of reasoning biases, belief bias in particular (Moutier et al. 2006, De Neys & van Gelder 2009, Markovits & Schroyens 2007) suggest that e.g. basic instruction in logic can weaken the effect. To my knowledge, however, the possible de-biasing effect of using specific notational systems such as formal languages or algebraic formalism has not yet been sufficiently investigated.[3] One study (Sá et al. 1999) suggests that, when subjects are given material that is to some extent 'meaningless' to them, the belief-bias effect is weaker. In this study, subjects were first given an invalid syllogism with familiar content, and in fact with a believable conclusion:

> All living things need water.
> Roses need water.
> Roses are living things.

As could have been anticipated, only 32% of the subjects gave the logically 'correct' response when evaluating this syllogism, i.e. that it is invalid. They were then given a little scenario of a different planet, involving an imaginary species, wampets, and an imaginary class, hudon, and subsequently were asked to evaluate the following syllogism:

> All animals of the hudon class are ferocious.
> Wampets are ferocious.
> Wampets are animals of the hudon class.

Interestingly, 78% of the very same subjects whose great majority had failed to give the 'logically correct' response in the previous task gave the 'logically correct' response here, i.e. that the syllogism is invalid. Even more significantly, the two syllogisms have the exact same mood, presumably AAA-2 (the universal quantifiers are unstated in the second premise and in the conclusion). So while they had failed to 'see' the invalidity of

[3] Some of the few studies on the cognitive improvement afforded by notations in reasoning processes are (Koedinger 1991) and

the first syllogism, arguably in virtue of the familiar content and the believability of the conclusion, in the second case the unfamiliar content apparently made it so that prior beliefs did not interfere to the same extent in the subjects' reasoning.

Thus, it can be hypothesized that reasoning with symbols which do not immediately elicit meaning connections, as in the case of special notations, might be a way to circumvent, at least partially and temporarily, the tendency to elicit prior belief in reasoning processes. Now, when such an engrained reasoning pattern is not called upon, other reasoning strategies may take its place, which in turn may enlist different cognitive systems. This is indeed (at least part of) what seems to happen when an agent reasons aided by these special notations, as I shall argue now.

2. Notations and extended cognition

What exactly is the role of notations when an agent is doing logic, mathematics or any other intellectual activity that relies heavily on these devices? Prima facie, there seem to be (at least) two plausible but incompatible positions on the matter: i) notations merely express (internal) cognitive processes that take place prior to and independently from their expression; ii) notations have an active cognitive function to play in the very reasoning processes in question. According to the first account, "[s]ymbolic reasoning is proposed to depend on internal structural rules, which do not relate to explicit external forms" (Landy & Goldstone 2007a, 720). By contrast, according to the second account, notations actually *embody* the cognitive processes themselves, which take place through the 'paper-and-pencil' manipulation of the symbols. These two positions are represented in the relevant literature (see (Landy & Goldstone 2007a, 720) and (Landy & Goldstone 2009) for references), but it is fair to say that the first position, which stresses the irrelevance of (most) perceptual properties of notations for reasoning processes, is still more widely endorsed than the second one.

Here, I argue for two claims: a) the second approach is a more accurate picture of actual reasoning processes and the use we make of notations; b) it is precisely because it incites a form of perceptual, bodily engagement from the reasoner that reasoning with notations is a fundamentally different cognitive process, and thus offers an alternative to more spontaneous modes of reasoning such as those related to the belief bias phenomenon. Therefore, reasoning with notations can, and in fact often does, counter our tendency towards relying on prior beliefs when reasoning.

That portions of the environment can play an active role in cognitive processes is an idea that has been extensively discussed in recent years, often under the heading of 'extended cognition', having A. Clark (2008) and D. Kirsh (2010) as some of its main proponents. The range of portions of the environment that can have this function is very wide, including electronic devices, notebooks (as in the classical example of Otto in (Clark & Chalmers 1998)), diagrams, maps, objects, and many more. Here, we are interested in the

cognitive impact of specific kinds of inscriptions, namely special notations and especially the formal languages used in logic. Indeed, the concept of extended cognition seems particularly suitable to investigate the import of notations in these cases.

Empirical evidence in support of the extended cognition approach to notations can be found in the works of Landy and Goldstone (2007a, 2007b, 2009) and Stenning (2002, especially chap. 2). Stenning reports on a study comparing the effects of logic instruction using two kinds of formal languages: traditional 'sentential' language and an innovative diagrammatic language (Barwise and Etchemendy's 'Hyperproof' system). In particular, he was interested in investigating the transfer of learning from logic to other kinds of reasoning tasks. Clearly, even though the two systems are formally equivalent, their perceptual properties are quite different. So if learning logic were simply a matter of acquiring and developing the appropriate "internal structural rules", one should expect there to be no significant differences in results between the two approaches. What Stenning's results show, however, is that students being exposed to each of the two languages improve their skills at different cognitive tasks (on the basis of pre- and post-instruction Graduate Record Exam (GRE) tasks): students taking the traditional course had a more dramatic improvement of their performance on verbal tasks, while students taking the Hyperproof course had a more dramatic improvement of their performance on analytical tasks. More surprisingly, the results show that individual differences interact in unexpected ways with exposure to different formal languages (Stenning 2002, 67). What can be concluded is that two notational systems with different perceptual properties have had a significantly dissimilar cognitive impact on the students' learning process and transfer of learning, thus reinforcing the idea that learning logic is not just a matter of developing the appropriate "internal structural rules".[4]

Landy and Goldstone's work focuses on how the reasoner's engagement with the notation exploits sensorimotor systems. In one study (Landy & Goldstone 2007b), they analyzed the results of self-generated productions in the domains of handwritten arithmetic expressions and typewritten statements in formal logic. In both tasks, they claim to have "found substantial evidence for spatial representational schemes even in these highly symbolic domains" (Landy & Goldstone 2007, 2033), in particular concerning the systematic introduction of spacing having no logical/formal function to play. In another study (Landy & Goldstone 2009), they asked participants to solve simple linear equations with one variable, displayed against a background that moved rightward or leftward. They observed that solving the equation was facilitated when the background motion moved in the direction of the numeric transposition required to determine the unknown variable, thus suggesting the involvement of motor processing. In other words, with an equation such as

[4] For the record, all students passed the course, so there is no doubt that they did learn logic, at least at the introductory level aimed at.

$$2x + 4 = 7$$

solving was facilitated by a moving background from left to right, that is in the same direction as 'bringing' 4 over to the other side of the equation.

$$2x = 7 – 4$$

It is important to notice that the rule being applied is (in theory at least) that one can subtract the same amount from each side of an equation:

$$2x + 4 – 4 = 7– 4$$

But in practice, subjects were apparently transposing the digit '4' from the left to the right side of the equation.

These results suggest that, rather than mere auxiliaries for the performance of internal cognitive processes based primarily on internal schemata and internal structural rules, notations in logic (and arguably also in mathematics), formal languages in particular, are *constitutive* of the very cognitive processes that we refer to as 'doing logic'. Of course, this does not mean that one cannot 'do logic' without the explicit engagement of notations, but when notations are being extensively used, it would seem that they are constitutive of the cognitive processes, in a strong sense of 'constitutive'. So it seems appropriate to speak of an *externalization* of the reasoning process: when notations are involved, thinking takes place on the paper (or computer screen or what have you) as much as in the head.

The externalization of the reasoning process in turn allows the reasoner to counter, or at least mitigate, her own reasoning biases; when sufficiently developed, the notation provides specific instructions on how to proceed (i.e. which 'moves' are allowed) by means of its well-defined syntax. In particular, in deductive settings, an appropriate notation requires that all premises be made entirely explicit, and the transformational steps allowed for given premises are entirely determined by the rules of transformation within the notational system. In effect, Landy and Goldstone's work suggests that these transformations substantially engage sensorimotor systems to be carried out: "elements of the problem are 'picked up' and 'moved' across the equation line" (Landy & Goldstone 2009, 1). The agent literally 'moves' bits and pieces of the notation around to perform these transformations.[5]

[5] In recent work, D. Macbeth has been arguing that this is exactly how one should operate with the formal system presented in Frege's *Begriffsschrift*. However, strict rule-following and the engagement of sensorimotor systems are two independent features; I am claiming here that most notational systems have both features, but it is possible that a given notation has one feature without having the other.

Thus, reasoning with notations typically (though perhaps not always) has two features that seem to contribute to the possibility of, when appropriate (e.g. in mathematical contexts), countering spontaneous reasoning mechanisms, such as those described in section 1. One of them is the engagement of sensorimotor systems, providing a form of 'physical grounding' that differs from the grounding of e.g. conceptual metaphors (Lakoff & Nuñez 2000). The second feature is the fact that, in developed notational systems such as formal languages/systems, the rules of formation and transformation are explicitly and exactly formulated, thus constraining the moves available to the reasoner. She cannot let her mind wander 'at will', which would likely lead to some well-entrenched patterns (such as those described as 'reasoning biases'); she must instead strictly follow the instructions contained in the notation. In both cases, the externalization of reasoning processes plays a crucial role.[6]

3. Conclusion

Reasoning with notations offers the possibility of circumventing certain reasoning tendencies such as calling upon prior belief, precisely because it is a fundamentally different kind of cognitive task; by means of the (partial) externalization of the reasoning processes, it allows these processes to run 'on a different software', as it were.[7] One of the 'built-in' softwares that we seem to come equipped with (described by Stanovich as a 'fundamental computational bias') is the tendency to rely on external, prior belief when reasoning – and for good reasons, as discussed by Stanovich (2003) himself. A formal language/system offers a different 'software' to operate with, which is arguably implemented not by the internalization of its rules, as one might think, but at least to some extent by the sensorimotor manipulations of the notation itself.

Thus, I conclude that Frege's two claims in the passage above can be seen as vindicated by recent experimental results on belief bias and on the involvement of perceptual-motor systems in our uses of special notations.[8] Unbeknownst to him, he seems to have

---

[6] In 'extended cognition' terms, the account proposed here is in the spirit of 'second-wave extended mind' (Sutton 2010) and 'cognitive integration' (Menary 2010) rather than in the spirit of Clark's parity principle (Clark & Chalmers 1998): I emphasize the transformative effect of reasoning with external devices such as formal languages, rather than the similarities between processes involving external devices and processes not involving them.

[7] But notice that the use of the software metaphor should not be read as an endorsement of the computational theory of mind. It is above all intended as an illustrative metaphor. Moreover, the claim that these are different cognitive tasks should have important implications for the 'parity argument', which is central to the 'extended mind' thesis. But again, this is a topic for further research.

[8] This is not intended as a claim pertaining to an exegetical and historical analysis of Frege's work. In fact, it would seem that Frege's thoughts on notation reveal a tension between the idea that notations express cognitive processes that are *prior* to their expression, and the idea that the cognitive processes *themselves* take place with the paper-and-pencil activity. One way to describe this tension is in terms of his hesitation between a Kantian and a Leibnizian conception of the role of notation in reasoning, which are diametrically opposed; but this will remain a topic for future work.

anticipated a promising way of viewing the cognitive impact of notations, i.e. from the standpoint of extended cognition.

References

Clark A. (2008). *Supersizing the Mind*. Oxford, OUP.

Clark A., Chalmers D. (1998). 'The extended mind'. *Analysis* 58:10-23.

De Neys W., Van Gelder E. (2009). 'Logic and belief across the lifespan: The rise and fall of belief inhibition during syllogistic reasoning'. *Developmental Science* 12, 123-130.

Evans J. (2002). 'Logic and human reasoning: an assessment of the deduction paradigm'. *Psychological Bulletin*, 128(6), 978-996.

Evans J., Barston J., & Pollard P. (1983). 'On the conflict between logic and belief in syllogistic reasoning'. *Memory and Cognition*, 11, 295-306.

Frege G. (1879). 'Begriffsschrift'. In van Heijenoort J. (1977). *From Frege to Godel – A source book in Mathematical Logic, 1879-1931*. Cambridge, Mass., Harvard University Press, third printing. 1-82.

Goel, V., & Dolan, R.J. (2003). 'Explaining modulation of reasoning by belief'. *Cognition*, 87, 11-22.

Kirsh D. (2010). 'Thinking with external representations'. *AI and Society*, forthcoming.

Koedinger K.R. (1991). 'On the design of novel notations and actions to facilitate thinking and learning'. In Proceedings of the *International Conference on the Learning Sciences*, (pp. 266-273). Charlottesville, VA: Association for the Advancement of Computing in Education.

Lakoff G., Nuñez R. (2000). *Where mathematics comes from*. Basic Books.

Landy, D., Goldstone, R. L. (2007a). 'How abstract is symbolic thought?' *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **33**(4), 720-733.

Landy, D., Goldstone, R. L. (2007b). 'Formal notations are diagrams: evidence from a production task.' *Memory and Cognition*, **35**(8), 2033-2040.

Landy, D., & Goldstone, R. L. (2009). Pushing Symbols. *The 31st Annual Conference of the Cognitive Science Society*, Amsterdam. Available at
https://facultystaff.richmond.edu/~dlandy/Publications/Landy_Goldstone_2009.pdf

Menary, R. (2007). *Cognitive Integration: mind and cognition unbounded*. Basingstoke: Palgrave Macmillan.

Moutier, S., Plagne-Cayeux, S., Melot, A., & Houdé, O. (2006). 'Syllogistic reasoning and belief-bias inhibition in school children: evidence from a negative priming paradigm'. *Developmental Science* 9(2), 166-172.

Markovits, H., Schroyens, W. (2007). 'A curious belief-bias effect: Reasoning with false premises and inhibition of real-life information'. *Experimental Psychology* 54, 38-43.

Oakhill J.V., Johnson-Laird P.N. (1985). 'The effect of belief on the production of syllogistic conclusions'. *Quarterly Journal of Experimental Psychology* 37A, 553–569.

Sá, W., West, R. F., & Stanovich, K. E. (1999). 'The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill'. *Journal of Educational Psychology*, 91, 497-510.

Stanovich, K. E. (2003). 'The fundamental computational biases of human cognition: Heuristics that (sometimes) impair decision making and problem solving'. In J. E. Davidson & R. J. Sternberg (eds.), The psychology of problem solving (pp. 291-342). New York, CUP.

Stenning, K. (2002). *Seeing Reason*. Oxford, OUP.

Sutton, J. (2010). 'Exograms and Interdisciplinarity: history, the extended mind, and the civilizing process'. In Menary, R. (2010a). *The Extended Mind*. Cambridge (MA): MIT Press, 189-226.