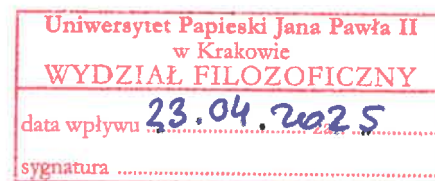


Dr hab. Paweł Gładziejewski, prof. NCU
Nicolaus Copernicus University in Toruń
Department of Cognitive Science
Gagarina 39, 87-100 Toruń
pawel.gla@umk.pl

Toruń, 7th April 2025



Review of Hajo Greif's habilitation application based on an evaluation of his academic contributions to the discipline of philosophy

1. Introduction

My aim here is to evaluate Dr Hajo Greif's habilitation application by assessing the novelty and academic quality of his contributions to philosophy, in particular to the philosophy of mind and cognitive science. Dr Greif has provided a summary of his academic achievements, highlighting the scholarly works that will form the main basis of my evaluation. These works include the book *Environments of Intelligence: From Natural Information to Artificial Interaction* (Routledge, 2017), accompanied by four thematically related journal articles published in *Biology & Philosophy*, *Synthese*, *Studies in History and Philosophy of Science*, and *Avant*.

Let me state at the outset that my overall conclusion is positive. I will argue that Dr Greif has achieved three theoretical goals, the originality and intellectual quality of which fully justify awarding him habilitation. These three achievements include: (1) successfully supplementing Gibson's ecological theory of perception with a developed concept of information grounded in Dretske's work; (2) proposing an account of perceptual illusions that extends Gibson's theory of perception; (3) providing an original, evolutionary-oriented interpretation of the idea of an extended mind, resulting in a formulation that is better positioned to address classical criticisms raised against this idea. As I will explain below, some other achievements stated in Dr Greif's self-report and application, though valuable, fail to stand as developed enough to merit the status of habilitation-level contributions. Nevertheless, these shortcomings do not undermine the overall positive assessment of Dr Greif's achievements.

In Section 2 of this review, I will focus on justifying why the three achievements mentioned above constitute original and high-quality academic contributions. In Section 3, I will briefly summarize and evaluate Dr Greif's additional research-related, teaching-related, and organizational achievements.

2. Evaluation of the main academic contributions to philosophy

Dr. Greif's philosophical outlook combines influences from Dewey and classical pragmatism with the analytic tradition and is significantly shaped by evolutionary theory, though on numerous occasions he also makes reference to the tradition of German philosophical anthropology. The body of work that comprises Dr Greif's main academic contribution is broad in its thematic scope but internally coherent and consistent. The ideas developed by the candidate revolve around a core notion that human intelligence is tightly bound up with and shaped by extra-cranial environments. Working within a broadly evolutionary approach (inspired by developmental systems theory or niche construction approaches), Dr Greif construes the environments that shape and co-constitute cognition as informational in nature. The way Dr Greif views the way in which humans rely on informational environments is, in turn, rooted in the Gibsonian approach to perception. Given those theoretical building blocks, the candidate develops a theory of cognitive artifacts that co-constitute cognition by modifying humans' informational environments. As such, the project is directed at tackling core and hotly debated questions connected to the very theoretical underpinnings of cognitive science. Considering the breadth and ambition of the aims behind Greif's academic achievement, the proposal being reviewed is appropriate for what should be expected from a habilitation-level contribution.

Although Dr Greif's proposal comprises tightly interconnected ideas that, to a large extent, form a relatively seamless whole, I will proceed by singling out three particular proposals that I regard as unequivocally novel and valuable contributions to the philosophy of cognitive science. I discuss each of those contributions in turn.

- Contribution 1: Applying Dretske's concept of information to clarify Gibsonian ecological theory of perception, as developed in:
 - Greif, H. (2017). *Environments of Intelligence* (Chapters 2, 4). Routledge.

- Greif, H. (2019). Affording illusions? Natural information and the problem of Misperception. *Avant* 10, 3/2019, 1–21.

Gibson's ecological theory of perception notoriously raises a number of questions regarding its conceptual foundations. What does it mean that perception relies on "picking up" of information that is "directly" present in the optic flow (i.e., the structure of light projected onto the retina in response to the movement of the observer)? What is the notion of information at use? What is the ontological status – "objective" vs. "subjective" vs. "relational" – of affordances? It is, therefore, a very worthwhile project on the part of the candidate to try to elucidate some of those issues by relying on classic work from philosophy of mind in the analytic tradition. Specifically, Dr Greif attempts to combine Gibsonian approach to perception with Dretske's account of information.

This sort of move may initially strike someone as implausible: Dretske's analysis of information is conducted in the context of constructing a theory of representation (and knowledge), while Gibson is famously a staunch antirepresentationalist. However, Dr Greif executes this theoretical move in a way that is perfectly legitimate. In Gibson's theory, information directly "specifies" the environment in a way that allows the observer to guide behavior with respect to it. For Dretske, information is a matter of non-semantic indication relation, such that a signal r indicates (carries information about) state S if the probability of S being the case, given r , is 1. Thus construed, informational relations are objective: they exist "out there", in virtue of causal-nomological relations that hold independently of observers. At the same time, information can be seen as a resource that an organism could begin to rely on in its behavior, for example through an evolved mechanism. For example, motion parallax or texture gradient can be regarded as carrying spatial information "objectively", but since this information is such that it can turn out useful for navigation, it will create pressure to evolve perceptual mechanisms that appropriately exploit this preexisting information. Mechanisms of this kind are, simply, perceptual mechanisms.

Dr Greif's elucidation of Gibson through Dretske is largely well-motivated and successful. It clarifies the notion of "information" at use in the ecological theory of perception, as well as aptly navigates between "objectivist" and "subjectivist" interpretations of the theory. I wonder, however, if this sort of view carries with it a

somewhat surprising consequence that most Gibsonians would vehemently oppose (and I suspect such a conclusion might also be rejected by Dr Greif). Note that Dretske famously argued that information becomes semantic – that is, it gives rise to satisfaction or accuracy conditions – when it starts to perform a function for an organism. But does this not suggest that on a Dretskean interpretation of Gibson, once objective information becomes something that an organism relies on in its practical dealings with the environment, then it becomes semantic, i.e. representational? For example, could we say that what Gibson treats as a misperception of affordance is equivalent to what Dretske would call a misrepresentation?

Despite my positive appraisal of this contribution, I want to notice an open question it raises. As the candidate himself notices, Dretske's restrictive, nomological notion of information has been subject to numerous criticisms and initiated the move towards more consumer-oriented and less conceptually restrictive versions of teleosemantics (Millikan, Skyrms, Shea). A natural question to be raised in this context is this: why use Dretske in the first place? For example, on the face of it, Millikan's theory seems significantly more in line with the candidate's overarching philosophical aims. First, Millikan's theory is not purely consumer-oriented, as it stresses the role of "isomorphism" between states of organisms and external states of affairs (the famous "input" condition). Millikan does so *without* relying on the excessively restrictive condition regarding the probability of a signal given the state signaled as being 1. As such, Millikan leaves plenty of room to explain how information could "specify" the environment in a way that is not burdened with problems that accompany Dretske's theory. At the same time, Millikan's teleosemantic theory has a number of features that look almost custom-made to make her theory compatible with Dr Greif's general approach (and with Gibson): it is explicitly stated in evolutionary terms and is more extensively reliant on the idea of information as something that is affordance-like, i.e. as being exploitable by an organism for action guidance. In any case, while I do appreciate the value of Dr Greif's contribution here, the very theoretical choice of using Dretske seems to me somewhat suboptimal and undermotivated in light of the candidate's theoretical aims.

- Contribution 2: Accounting for perceptual illusions within the assumptions of the framework of Gibsonian theory of perception, as developed in:

- Greif, H. (2017). *Environments of Intelligence* (Chapter 3). Routledge.
- Greif, H. (2019). Affording illusions? Natural information and the problem of Misperception. *Avant*, 10, 3/2019, 1–21.

Another way in which Dr Greif supplements the Gibsonian account of perception is by addressing the following problem. If perception is a matter of direct interaction with the immediate environment, how are we to account for the possibility of perception sometimes being wrong? In addressing this problem, Dr Greif makes use of a distinction between misperception and perceptual illusion. In misperception, an object is mistaken for another object (a plastic replica is mistaken for a tomato due to possessing perceptual features of redness, roundness, etc.). In Gibsonian terms, this could be explained in terms of “misjudgment” of affordances. For example, a person might perceive stairs as climbable stairs despite them not being so (here let me note again a deep conceptual affinity between this sort of account and more traditional representationalist views that treat perceptual states as possessing accuracy conditions). Illusions, in turn, include cases where the organism gets the very perceptual *features* wrong. For example, the Müller-Lyer illusion and the Ponzo illusion involve misspecification of the spatial features of the objects in the environment.

With the illusion/misperception distinction in hand, Dr Greif focuses on accounting for the former. To that end, he uses the "Empirical Strategy" approach in perception, as developed by Purves and his collaborators. Instead of focusing on the relationship between the environmental triggers and the perceptual state, the focus here is on how perceptual states guide successful action. In particular, the way organisms perceive the environment is treated as having been shaped by how perception guided successful action in the past (again, as a side note, it seems that Millikanian evolutionary-oriented consumer semantics is more naturally suited to this approach than Dretske's theory). From this perspective, illusions may be construed as side-effects of properly functioning perceptual mechanisms that were shaped by past success. A visual system that is well-adapted to ecologically normal conditions (in the sense of corresponding to environments in which the visual system evolved) will almost inevitably generate illusions in circumstances that differ from such conditions. For example, if a shaded object has the same luminance as a well-lit

object, this usually means that the latter is objectively darker than the former. A perceptual system that is well-adapted to this pattern will also naturally tend to be susceptible to the famous checkerboard illusion. In this sense, illusions do not count as misperceptions in the traditional sense after all, and the challenge they pose to the ecological approach is neutralized.

I think that linking Gibson and the Empirical Strategy in this way is a theoretically sound move and a valuable development when it comes to accounting for misperception within Gibson's conceptual framework. This approach is also reminiscent of how Gerd Gigerenzer accounts for the ecological rationality of heuristics that people rely on in thinking and problem-solving. This suggests that Empirical Strategy is an example of a broader approach that scales up beyond the domain of sensory perception.

- Contribution 3: Proposing a novel, evolutionary interpretation of the extended mind theory, as discussed in:
 - Greif, H. (2017). *Environments of Intelligence* (Chapters 5, 6, 7). Routledge.
 - Greif, H. (2017). What is the extension of the extended mind? *Synthese*, 194, 4311–4336.

What I take to be Dr Greif's most significant contribution is his proposal regarding the nature of extended cognition. Since the publication of the seminal paper by Clark and Chalmers, there has been an ongoing debate in the literature about what it takes for cognition to extend "beyond the skull" in a non-trivial sense. So it is no mean feat that the candidate has succeeded, as I think he has, in an endeavor to add genuinely novel and worthwhile ideas to this ongoing debate.

There are three main components to Dr Greif's account of the extended mind. First, he construes the environments that extend cognition as primarily informational. That is, cognitive extensions do not simply build upon objective, organism-independent physical surroundings, nor even on ecological environments, that is, on the subset of the physical world that is relevant to an organism's fitness or its ability to fulfill its goals. Instead, the idea is that extensions involve those parts of the ecological environments that the organism can selectively track by registering the relevant information.

Second, Dr Greif proposes an evolutionary approach that considers the brain-environments couplings as functional units that emerged through evolutionary (selective) processes – in an inclusive sense that may involve cultural evolution. That is, the extended mind is not only a matter of current interactions spanning the brain/organism and environment but is constituted by the history of how those sorts of interactions came about. In proposing this solution, Dr Greif is largely inspired by such approaches in evolutionary theory as niche construction or the developmental systems theory.

Third, Dr Greif proposes a distinction of two different ways in which parts of the environment can be constitutive of cognitive functions. A cognitive extension is “constitutive_w” if part of an external environment performs a function that can also be carried out by mechanisms internal to the organism. An extension is “constitutive_s” if it is necessary for a given cognitive/biological function to exist in the first place. That is, when an extension is constitutive in this latter sense, there is no alternative purely “internal” mechanism that could give rise to this function. Dr Greif discusses language as an example of this latter, stronger type of constitution. That is, language can be regarded as an external resource that is necessary for certain human cognitive functions to exist. It is a cognitive equivalent of a beaver dam (which is necessary for beaver populations to exist) rather than a tool used by a chimpanzee (which affords the achievement of goals that are also achievable without reliance on tools).

As Dr Greif argues (persuasively, I think), this sort of formulation of extended cognition can serve as a basis on which proponents of this view could reply to a number of classical challenges to it. By highlighting the role of etiological/evolutionary factors, it provides a principled way to distinguish genuine cognitive extensions (that is, couplings whose existence has been generated and stabilized by way of processes of selection) from mere external causes/scaffolds of cognitive processes. On clearly principled grounds, this view drops the requirement that an extension must be persistently present in the cognitive system, which neutralizes the famous challenge raised by Robert Rupert. Regardless of many other nuances that could be discussed in this context, I want to reiterate that Dr Greif’s work on extended cognition is a genuinely novel and valuable contribution to the literature.

- Evaluation of other proposed contributions

Before closing this section, I want to briefly discuss two other contributions of Dr Greif's which I consider to be slightly too underdeveloped to be considered habilitation-level achievements. The discussion of likeness-making and the markings from the Lower Paleolithic (Greif, H. (2022). Likeness-making and the evolution of cognition. *Biology & Philosophy*, 37) is speculative and tentative to the degree that I think it should be regarded as an initial sketch of an idea rather than a fully developed proposal. More importantly, I think that the candidate's central idea regarding the nature of cognitive artifacts, which he develops in the closing chapters of *Environments of Intelligence*, would require some further refinement and clarification to fulfill its potential. The very distinction between convergence-based artifacts (which make accessible to the organism information that is unavailable to its natural sensory apparatus) and isomorphism-based artifacts (that rely on the construction of completely new, virtual informational environments) seems useful and promising. However, the discussion of concrete examples lacks sufficient detail to persuasively establish that the conceptual framework introduced by Dr Greif really illuminates the nature of those artifacts. It is not always clear how the concepts introduced (for example, constitutive_w/constitutive_s, convergence/isomorphism) connect to some of the examples (for example, social robots). Despite the existence of extensive empirical literature on how language reconfigures and expands human cognition, the discussion of language as a cognitive extension is very limited and sketchy.

Regardless of those reservations, let me reiterate that the three other proposals discussed above are fully sufficient to warrant a decisively positive overall assessment of Dr Greif's habilitation application.

3. Evaluation of other academic, didactic and organizational achievements

Dr Greif is a researcher whose interests are not confined to a narrow, highly specialized subject area, but instead span a relatively wide and varied range of issues. Nevertheless, his diverse academic pursuits remain interconnected and form a coherent, unified research profile. Outside of the strictly habilitation-related publications already mentioned, the candidate has published on the history of cognitive science (the work of Alan Turing and W. Ross Ashby, as well as their connections), philosophy of AI (opacity/transparency of AI model, exploratory models in AI), history of Darwinian theory (Darwinism in light of romantic *Naturphilosophie*

versus mechanistic natural philosophy) and history of philosophy of science (in particular, Ludwik Fleck's influence on Kuhn). To evaluate those contributions, I will briefly discuss four papers selected from Dr Greif's overall academic output, each corresponding to one of the areas mentioned. I regard them as high-quality papers and I benefitted from reading each of them.

In "Turing's biological philosophy: Morphogenesis, mechanisms and organicism" (co-authored with Adam Kubiak and Piotr Stacewicz, published in *Philosophies*, 2023), Dr Greif and co-authors dive into Alan Turing's lesser-known foray into theoretical biology. In particular, they discuss Turing's theory of morphogenesis, which aimed to mathematically explain how complex patterns in living organisms emerge from simple diffusion reactions between idealized biochemical substances or "morphogens" (reactions that could, in turn, be described computationally). The authors argue that Turing's work wasn't just a side project, but an extension of his mechanistic worldview. They also investigate the relation between Turing's mechanicism and "organicism" ideas about the nature and emergence of biological forms, often championed by his contemporaries.

Dr Greif's "Models, algorithms, and the subjects of transparency" (a chapter in *Philosophy and Theory of Artificial Intelligence*, Springer, ed. V.C. Müller, 2022) is devoted to AI's so-called "black box problem" by asking what we really mean when we talk about opacity and transparency in AI. The candidate argues that the issue isn't just whether AI systems are hard to understand, but how different epistemic subjects – scientists, engineers, users – relate to these systems and models. A distinction is made between opacity related to complexity (opacity related to the computational tractability of a process) and universality (uncertainty related to the representational properties of a computational model of some phenomenon or system). Dr Greif argues that achieving transparency in AI models is a challenge that extends beyond the technical problem of computational complexity; it also concerns fundamental differences between the nature of representations in computational models and the expectations and capacities of human epistemic agents.

In "The Darwinian tension: Romantic science and the causal laws of nature" (published in *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2015), Dr Greif investigates the intellectual roots of Darwin's theory of evolution. In particular, he focuses on the interplay between two intellectual traditions that influenced Darwin: the mechanistic, causal approach of early Victorian natural

science (exemplified by Herschel, Lyell, and Malthus), and the holistic, aesthetic, and teleological perspective of German Romantic science (represented by Humboldt and Goethe). Dr Greif argues that Darwin's work synthesizes these traditions, stemming from a productive tension between them. While Darwin adopted Humboldt's holistic view and refined observational practices, he also employed causal explanations characteristic of Herschel's scientific method. This synthesis enabled Darwin to apply causal laws to "animate" nature without abandoning the Romantic vision of nature as a harmonious, teleologically structured whole.

In "The 'Aristotle experience' revisited: Thomas Kuhn meets Ludwik Fleck on the road to *Structure*" (co-authored with Paweł Jarnicki, published in *Archiv für Geschichte der Philosophie*, 2022), Dr Greif and Paweł Jarnicki take a fresh look at Thomas Kuhn's famous "Aristotle experience" narrative, arguing that the story about Kuhn's sudden epiphany in 1947 downplays the influence of Ludwik Fleck. Fleck's ideas on thought styles and scientific collectives strikingly anticipated Kuhnian paradigms. The authors trace how Kuhn's narrative of intellectual self-discovery became more dramatic over time, especially after Fleck's work was translated into English, and suggest Kuhn may have been more indebted to Fleck than he was aware of – or cared to admit. They explore several hypotheses about how Fleck's theory could have informed Kuhn's *The Structure of Scientific Revolutions*, including one on which Kuhn was implicitly and unconsciously (in a slightly cryptoamnesiac manner) affected by his understanding of Fleck's work, which was limited by the language barrier. The paper constitutes an interesting take on the "origin story" of Kuhn's pivotal work in philosophy and the history of science.

Let me now turn to the assessment of the candidate's organizational and teaching-related achievements. Dr Greif spent a significant part of his academic career in Germany and Austria. He conducted both research and teaching duties at well-regarded institutions such as the Technical University of Darmstadt (where he completed his PhD under a DFG-funded fellowship), Institute for Advanced Studies on Science, Technology and Society in Graz, Austria (research fellowship), University of Klagenfurt (assistant professorship and lectureship) and Technical University of Munich (first, as an FWF Erwin Schrödinger Fellow and later as a senior researcher and lecturer). The candidate also visited the Science Studies Unit at the University of Edinburgh as a part of a DFG-funded research project. After his appointment at the Warsaw University of Technology, he was awarded the Opus 19 grant from NCN,

acting as a Principal Investigator. Considering all this activity spanning multiple academic institutions, Dr Greif decisively meets the criterion of carrying out his research in more than one university, particularly abroad.

The candidate is also an active participant in academic life in an organizing role. He holds board memberships at the International Association for Computing and Philosophy (IACAP), the Commission for the History and Philosophy of Computing (HaPoC), and the European Philosophy of Science Association (EPSA). He participated in organizing multiple IACAP and HaPoC conferences, as well as many other conferences and symposia. He gave 19 keynote or invited talks at conferences, seminars and symposia. Dr Greif also acts as a reviewer for multiple academic journals, including ones that are crucial venues for the areas of his expertise (for example, *Synthese*, *Biology & Philosophy*, *Minds & Machines*).

Lastly, it should be noted that dr Greif has amassed significant teaching experience during his academic career, giving courses at multiple academic institutions, on subjects that span the philosophy of science and technology, philosophy of biology, and general philosophy of science. The candidate also supervised two doctoral theses and one MA thesis.

In light of all these achievements, I am confident that Dr. Greif's academic output (beyond the habilitation accomplishment itself), along with his organizational and teaching activities, fully satisfies the criteria typically expected of candidates for the habilitation.

4. Conclusion

In light of my positive evaluation of Hajo Greif's academic contributions, and in recognition of his organizational and teaching activities, I am pleased to recommend awarding him the habilitation. I am confident that his accomplishments and commitment to scholarly work fully justify this distinction.



dr hab. Paweł Gładziejewski, prof. NCU