Studies in History and Philosophy of Science 107 (2024) 82–91

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Theoretical concepts as goal-derived concepts

Matteo De Benedetto

Ruhr-University of Bochum, Department of Philosophy II, Universitätsstrasse 150, 44780 Bochum, Germany

ARTICLE INFO

Keywords:
Scientific concepts
Theoretical concepts
Goal-derived concepts
Goals in science
Pragmatist philosophy of science

ABSTRACT

In this paper, I will focus on the nature of theoretical concepts, i.e., the psychological entities related to theoretical terms in science. I will first argue that the standard picture of theoretical concepts in twentieth-century philosophy of science understood them as representation-oriented common taxonomic concepts. However, I will show how, in light of recent pragmatist approaches to scientific laws and theories, several important theoretical concepts in science do not seem to fit such picture. I will then argue that these theoretical concepts should be understood instead as goal-derived concepts, since their construction and use exhibit the typical characteristics that cognitive scientists assign to goal-derived concepts. I will furthermore argue that the existence of theoretical concepts that are goal-derived concepts represents yet another example of the central role that human goals play in science.

1. Introduction

The problem of theoretical terms has been a central issue in philosophy of science for much of the last century. The semantics of terms such as ‘force’ and ‘gene’ has been, in fact, a central preoccupation of logical empiricists and structuralists for decades. Even when both statement and non-statement views of scientific theories fell out of fashion, the ontology and the denotation of theoretical terms has been heavily discussed in debates over scientific realism, scientific objectivity, and scientific progress. In all these debates, many aspects of theoretical terms have been analyzed, including their semantics, their pragmatics, their epistemological status, and their ontological import.

An aspect of theoretical terms that has not been much discussed is the nature of the cognitive relata of these terms, i.e., (what I call) theoretical concepts.[1] In this paper, I will address precisely this aspect. Specifically, I will focus on the kind of concepts that theoretical concepts are. I will first argue that the received, implicit view of theoretical concepts has conceptualized all of them as what cognitive scientists call common concepts (or taxonomic concepts), i.e., concepts like DOG or TREE.[2] I will show how this standard picture of theoretical concepts is implied by the standard representation-oriented view of theoretical terms commonly assumed in twentieth-century philosophy of science. However, I will show how more recent pragmatist approaches in philosophy of science depict several important theoretical terms in science as inherently contextual and functional entities. I will then argue that this pragmatist understanding of theoretical terms is incompatible with the standard picture of theoretical concepts that wants them to be all common concepts. Building upon recent cognitive science, I will propose that these theoretical concepts that do not fit into the standard picture should be understood as goal-derived concepts, i.e., concepts like THINGS TO TAKE ON VACATION or CLOTHES FOR THE WINTER. I will show how this assumption of theoretical concepts as goal-derived concepts adequately accounts for how, according to contemporary philosophy of science, scientists construct and use several important theoretical concepts, such as FORCE, GENE, and HARDNESS. I will furthermore demonstrate how my proposal that several important theoretical concepts in science are goal-derived concepts has some important consequences for our general understanding of science, in that it exemplifies the centrality of human goals and values for all aspects of scientific inquiry, even the most abstract ones. We will see how this centrality of human goals in the construction and use of some the most abstract scientific concepts calls for a re-orientation of our intuitive expectations about what science aims for.

The aim of this paper is then two-fold. The first aim is to argue against the implicit received view that assumes that all theoretical concepts are common concepts, showing instead how several important theoretical concepts in science are goal-derived concepts. The second, more general aim is to highlight how the hitherto under-discussed cognitive dimension of theoretical terms in science is very relevant for philosophy of science, in that it shows us how human goals are

E-mail address: matteo.debenedetto@rub.de.

[1] For some work discussing the cognitive dimension of scientific concepts, see Andersen, Barker, and Chen (2006), Giere (1988), Nersessian (2010), Thagard (1992).

[2] In this paper, I will use the somewhat standard notation of putting concepts in small caps lock and terms in quotation marks. Thus, DOG refers to the concept of dog, while ‘dog’ refers to the related term.

https://doi.org/10.1016/j.shpsa.2024.08.004
Received 26 September 2023; Received in revised form 30 July 2024

central to even the most abstract aspects of scientific practice, such as the construction and use of theoretical concepts.

In Section 2, I will discuss the standard representation-oriented view of theoretical terms in twentieth-century philosophy of science and I will argue that this view implies an implicit standard picture of theoretical concepts as common concepts. In Section 3, I will describe the problems of this standard view of theoretical terms, focusing on how pragmatist philosophy of science questioned this view and stressed instead the inherent contextual and functional nature of several important theoretical terms. I will then argue that this different view of theoretical terms in science casts doubts on the standard picture of theoretical concepts as common concepts. In Section 4, I will present the idea of goal-derived concepts, as conceptualized by recent cognitive science. In Section 5, I will then argue that the theoretical concepts that do not fit the standard picture are arguably goal-derived concepts. I will first argue that this proposal adequately fits with how, according to our best philosophy of science, scientists construct and use several important theoretical concepts and, then, I will show the more general implication of this view of theoretical concepts for our ideal of the ultimate goal of scientific inquiry. Section 6 concludes.

2. The standard picture of theoretical concepts

In this section, I will describe the standard picture of theoretical concepts in twentieth-century philosophy of science. We will see that such a standard picture considers concepts like GENE and FORCE in science to be cognitively alike to common taxonomic concepts like DOG and TREE. More specifically, we will see that such a standard picture is an implicit consequence of the standard representation-oriented view of theoretical terms in science.[3]

Before starting our discussion, two important clarifications are needed. The first one concerns my use of 'concepts'. Throughout this work, I will use 'concepts' to denote the psychological entity involved in conceptual talk (i.e., the mental representation), while I will use 'term' to denote the related linguistic entity, and 'category' (or kind) to denote the abstract or worldly entity denoted by conceptual talk. I will not assume any specific psychological theory of conceptual structure (e.g., prototype, exemplar, theory-like, atomistic, etc.), but I will assume the standard view in cognitive science that our concept-talk refers primarily to the abstract bodies of knowledge, stored in the long-term memory, with which we perform several higher-cognitive tasks such as categorization, abstraction, and conceptual inferences (cf. Machery 2009, Margolis and Laurence 1999, Murphy 2002). This focus on the cognitive aspect of scientific concepts is consistent with general naturalistic analyses of scientific development (e.g., Andersen et al. 2006, Giere 1988, Nersessian 2010), but it does not presuppose any specific view of science. It just focuses on an under-discussed aspect of theoretical terms, namely, the nature of their cognitive relata. So, from now on, when I speak about concepts, I mean to refer exclusively to the cognitive item involved in concept talk. The second clarification concerns instead my talk of theoretical terms (and consequently of theoretical concepts). I do not assume a sharp theory-independent distinction between theoretical terms and observational terms, nor I assume a specific theory about theoretical terms. I just take, pragmatically, theoretical terms to be the most abstract, non-logical, theory-dependent terms of a theory. This is compatible with virtually all the traditional accounts of theoretical terms (cf. Andreas 2021), such as the semantic-based ones (e.g., Carnap 1966) or the measurement-based ones (e.g., Balzer, Moulines, and Sneed 1987, pp. 47-78). Yet, if someone is completely skeptic about the very notion of a theoretical term, the whole paper can be read as focusing on certain abstract scientific concepts. The paper is about the concepts that are related to the terms traditionally considered theoretical, not about the terms themselves.

Twentieth-century philosophy of science did not care much about the cognitive dimension of concepts. The mental representations related to our scientific terms were not considered of great philosophical interest by many philosophers of science, who discarded them in favor of their linguistic counterparts. A paradigmatic example of this anti-psychologistic approach to scientific concepts is the philosophy of science of the logical empiricists (e.g., Carnap 1966, Hempel 1952), in which the psychological dimension of our scientific theories and concepts is purged away in the process of rational reconstruction. The output of this process, a syntactic or semantic structure in a formal language, was supposed to adequately represent the rational constituents of a given scientific theory and, as such, the philosophically-relevant part of the theory. Even when logical empiricism fell out of fashion, the centrality of semantic considerations did not diminish. The answer to the incommensurability threats allegedly embodied by Kuhn's (Kuhn, 1970) and Feyerabend's (Feyerabend, 1962) philosophies was, in fact, again a semantic one, consisting either in refined model-theoretic rational reconstructions (e.g., Balzer et al. 1987) or in reference-oriented externalist accounts of scientific kinds (e.g., Kitcher 1995, Putnam 1973, 1975). Either way, concepts and the related cognitive dimension of our scientific talk did not play a major role in the philosophy of science of that time.[4]

Despite this lack of explicit focus on concepts, some semantic and ontological assumptions commonly made by philosophers on theoretical terms implicitly project a somehow standard picture of their psychological relata, i.e., theoretical concepts. More specifically, we can identify three assumptions on the semantic and the ontology of theoretical terms that are particularly significant for understanding the implicit standard picture of theoretical concepts: the intra-theory stability, the property-based content, and the representational role of theoretical terms.

The intra-theory stability assumption wants the meaning of theoretical terms to be stable within the theory to which they belong. This means, for instance, that the same theoretical term should have exactly the same meaning in the different domains to which the theory is applied. Such a stability in meaning was assumed by virtually all contenders in twentieth-century debates over theoretical terms. A paradigmatic example of this assumption can be found in the logical empiricists' reconstruction of scientific theories (cf. Carnap 1966, Hempel 1952), where the meaning of theoretical terms was holistically determined by the most general laws of a theory (together with the related correspondence rules). Since such laws and correspondence rules were thought to be universally valid within the domain of the theory, this holistic determination of meaning assured full intra-theory stability. Even opponents of logical empiricism, such as Kuhn (1970) and Feyerabend (1962), held the intra-theory stability assumption. As a matter of fact, the assumption that the meaning of theoretical terms is the same within the whole theory was one of the main arguments

[3] Note that what I am calling the standard picture of theoretical concepts and the related standard view of theoretical terms denote two related sets of assumptions on, respectively, theoretical concepts and theoretical terms that were default assumptions in twentieth-century philosophy of science. As such, this view and picture should not be thought as shared by every philosopher of science of that time in exactly the same terms and degree, but instead as shared philosophical paradigms, the specific instantiations of which might change from philosopher to philosopher.

[4] An important exception is Kuhn's philosophy of science, the early formulations of which took the cognitive dimension of our scientific theories and concepts into serious considerations (cf. Kuhn 1970, 1974). Yet, as already stressed in Kuhn scholarship (Andersen, Barker, & Chen, 1996; Bird, 2002; Shan, 2020a), this cognitive dimension of Kuhn's philosophy did not have much impact in twentieth-century philosophy of science, being overshadowed by semantic considerations. In recent years, the philosophical significance of this cognitive dimension of Kuhn's work has been appreciated by cognitively-minded philosophers of science (cf. Andersen et al. 2006, Bird 2002, Nersessian 2010).

through which they contended the existence of incommensurability in scientific revolutions. Since the meaning of theoretical terms is determined by the theory as a whole (including the theory, the methods, and the other methodological components), they argued, radical changes of theories (and therefore of laws and methods) must imply also radical changes in the meaning of theoretical terms.

The property-based content assumption held that the extensional content of a theoretical term is determined thanks to the specific properties that its instances satisfy. Thus, something belongs to the extension of a given theoretical term because of the properties that it possesses. The exact specification of how the extensional content of theoretical terms is determined varies greatly depending on the specific semantics under consideration. Thus, for instance, in the descriptivist semantics favored by the logical empiricists, the extension of a theoretical term is determined by the theoretical and material postulates that define such term. This property-based content assumption can be paradigmatically seen at work in Carnap's (Carnap, 1961; Psillos, 2000) epsilon-based definition of theoretical terms, where a theoretical term is an arbitrary object in the domain that satisfies the conjunction of the related theoretical and material postulates. A completely different determination of the extension of a theoretical term is theorized by externalist semantics that want it to be determined mostly externally, i.e., by the properties that the instances of the kind causally referred to by the term actually share, downsizing the role of descriptive components to mere reference-fixing (e.g., Psillos 1999, Putnam 1973, 1975). Either way, whether one favors an internalist or an externalist semantics, the extensional content of a theoretical term is heavily dependent on the properties that its instances satisfy.

The representational role of theoretical terms refers to the assumption that the main purpose of theoretical terms is to represent entities. Specifically, theoretical terms are usually meant to represent hidden entities, i.e., entities that are not directly or indirectly observable. Since they cannot be observed, the existence of such hidden entities needs to be inferred from our best scientific theories. The theoretical terms of a theory constitute its ontological import, that is, they denote the entities the existence of which is postulated by the theory. The ontological status of the entities denoted by theoretical terms varies, of course, greatly depending on which attitude one takes towards questions of scientific realism. Full-blown scientific realists (e.g., Kitcher 1995, Psillos 1999), for instance, consider the entities referred by theoretical term natural kinds that carve the world at its joints. Anti-realists (e.g., van Fraassen 1980) consider instead these entities as useful abstractions that allow us to categorize phenomena in advantageous ways. What is common to both camps is that they both assume that the main function of these terms is to refer to some entity, the (fictional or real) existence of which is implied by the theory in which the term figures.

The combination of these three assumptions on theoretical terms, i.e., the intra-theory stability, the property-based content, and the representational role assumption, jointly sketch what I called the standard view of theoretical terms. The meaning of theoretical terms is stable across the theory to which they belong, since it is determined by the most abstract and general part of a scientific theory. The main purpose of theoretical terms is to refer to abstract entities, the existence of which is postulated by the laws of the theory, and to encode certain core properties that the instances of these entities all share.

This standard view of theoretical terms projects a related picture of the concepts connected to these terms, i.e., what I called theoretical concepts. In order to adequately relate with the intra-theory stability of meaning that theoretical terms exhibit, theoretical concepts must be relatively stable in content. Similarly, because the extensional content of theoretical term is property-based and because the main purpose of theoretical terms is representational, theoretical concepts should be made for representing common properties of scientific kinds. According to this picture, scientific concepts such as FORCE and GENE are general and stable, that is, they do not vary greatly from context to context. Theoretical concepts are moreover learned and used with the aim of representing scientific kinds and their common properties. Therefore, concept learning, identification, categorization, and use of such concepts is centered around the representation of central features of scientific kinds. These characteristics (i.e., stability of content, features centrality, and representational orientation) that theoretical concepts, according to what I am calling the standard picture of theoretical concepts, possess are typical of what cognitive scientists call common or taxonomic concepts (cf., Rosch and Mervis 1975). These concepts, i.e., concepts that represent classes of everyday things such as DOG, RED, TREE, and the like, are very stable in content, they are learned, identified, and used thanks to the common properties that they encode, and they are oriented towards faithful representation of the related categories. Thus, we could re-state the standard picture of theoretical concepts as saying that theoretical concepts work like common concepts.

This standard picture of theoretical concepts, with its focus on content stability and on the representational function of scientific concepts is not peculiar to philosophy of science, but it is arguably a part of a more general default attitude towards concepts and conceptual change that pervades analytic philosophy. This attitude wants concepts to be passive proxies of properties of things in the world that can be descriptively analyzed by philosophers and characterized by exact definitions. I take such an attitude to be what Wilson (Wilson, 2006) calls the "classical picture of concepts" and what Kindi (Kindi, 2012) calls the "concepts as vessel" view. The effects of such an attitude can be recognized in many different places in analytic philosophy, such as, in addition to the present case of the standard picture of theoretical concepts, the lasting popularity of the definitional theory of concepts in philosophy of mind (Murphy, 2002), the long neglect of conceptual change in philosophy of science (Arabatzis & Kindi, 2008; Kindi, 2012), the popularity of metaphysical realism over properties and attributes (Wilson, 2006), and the dominant status of conceptual analysis as a philosophical methodology (Cappelen, 2018).[5] Although the interconnections between all these parts and components of this default attitude are extremely interesting from a philosophical point of view, they would bring us far away from our present topic. As such, in what follows, I will focus exclusively on the strength and weaknesses of the standard picture of theoretical concepts as an independent philosophical thesis.

3. Problems with the standard view: Contexts and goals in science

We saw how the implicit standard picture of theoretical concepts in philosophy of science wants them to work like common concepts. Specifically, we saw how standard assumptions about the intra-theory stability, the property-based content, and the representational role of theoretical terms projected a related picture of theoretical concepts as stable in content, centered around features that they encode, and oriented towards representation.

However, in the last decades, the received views of scientific theories and laws in philosophy of science have changed. As a consequence of the so-called practice turn in philosophy of science, philosophers have questioned the received views about scientific theories, laws, and theoretical terms. The increasing attention dedicated by philosophers to the practical and pragmatic aspects of science made them reconsider some central assumptions of the semantic-centered orthodox view of theoretical terms that we saw in the last section. We can distinguish two main strands of (groups of) pragmatist critiques to the standard view of theoretical terms: a first group of critiques that question the intra-theory stability of theoretical terms, i.e., what could be deemed contextualist critiques, and a second group of critiques that question

[5] Many more examples of the effects of such an attitude can be found in Wilson's (Wilson, 2006, pp. 139–146) list of "Chief Theses" of the classical picture of concepts.

instead the property-based content and the representational role of theoretical terms, i.e., what could be deemed functionalist critiques. Let us look at these two groups of critiques, in turn.

What we can call contextualist critiques of scientific laws and theoretical terms encompass a group of pragmatist approaches in philosophy of science that contested the alleged universality and generality of scientific laws (e.g., Batterman 2001, Cartwright 1983, Giere 1988, Hacking 1983, Mitchell 1997, Wilson 2006).[6] A paradigmatic example of such critiques is the work of Cartwright (Cartwright, 1983, 1999), who highlighted the hidden contextuality of scientific laws in their applications to specific scientific problems.[7] In order to successfully apply a scientific law to a given problem in a given domain, Cartwright (Cartwright 1983, pp. 21–73, Cartwright 1999, pp. 23–74, 179–233) argued through a series of examples from physics and economics, scientists often drastically modify the laws and their ontological import, up to the point that different applications of the same law can differ so much as to describe the world in incompatible ways. The received view in philosophy of science of scientific theories that wanted them structured around a group of general and universal laws, easily applicable to any phenomenon in the domain of the theory via suitable restrictions or additional parameters, is then just a philosophical myth. According to contextualist critiques, scientific practice involves complex contextual adjustments and modifications of scientific theories to the specific problem and domain under focus. Our best scientific theories are not structured as a neat hierarchy of laws of growing generality, but as a patchwork of laws, locally-valid in a specific domain of application and related to each other in a complex way (cf., Batterman 2013, Cartwright 1999, Wilson 2006, 2017).

If scientific laws have to be often contextually adjusted in scientific practice, the same holds, according to contextualist critiques, for theoretical terms. The inferences afforded by a given scientific term often vary from application to application or from scale to scale (cf., Batterman 2001, 2013, Wilson 2006, 2017). The same theoretical term can have different referents, different meanings, different inferential consequences in different parts of the same scientific theory. This is the main idea behind so-called patchwork approaches to scientific concepts (e.g., De Benedetto 2021, Haueis 2021b, Novick and Haueis 2023, Wilson 2006), i.e., a cluster of pragmatic theories that wants scientific concepts to be structured as a complex cluster of partially-connected local domains of usages.[8][9] In recent years, several important theoretical terms have been shown to exhibit a patchwork structure, including the concept of force (Wilson, 2006, 2017), hardness (Wilson, 2006), species (Novick & Doolittle, 2021), neural column (Haueis, 2021a), attention (Taylor, 2023), and homology (Novick, 2018). In all these examples of patchwork structures, theoretical terms do not exhibit the intra-theory stability that the standard view assumed them to have. This is because of the aforementioned contextuality of scientific laws that change their form consistently with the demands of particular applications and domains. By modifying scientific laws, scientist often modify also the meaning, the reference, and the inferential import of the theoretical terms that appear in that law. Just like scientific laws, then, many of our theoretical terms seem to organize themselves as a complex cluster of contextually adjusted localized patches of usage.

The second strand of pragmatist critiques that we are going to focus on is what we can call functionalist critiques.[10] These pragmatic critiques contested the traditional representation-centric conception of scientific laws and theoretical terms, stressing instead the important functional role that scientific laws and theoretical terms play in solving problems and in pursuing goals of the related scientific communities (e.g., Brigandt 2010, Chang 2004, Feest 2010, Hacking 1983, Laudan 1978, Nersessian 2010, Wilson 2006). Traditionally, the goals pursued by scientists and the specific problems faced by a scientific community did not play a significant role in the philosophical image of a scientific theory. The pragmatic dimension of scientific practice, just like its cognitive dimension (cf. Section 2), was overshadowed by the semantic dimension, which was considered sharply detached from it. Through a series of case studies, functionalist critiques of scientific theories have argued that several central episodes of scientific change, and the related scientific theories and concepts involved in them, can only be understood if we take the pragmatic dimension into consideration. Often, in fact, the specific problems and contexts faced by scientists guide the emergence and dynamics of scientific theories (cf., Nersessian 2010, Wilson 2006). Indeed, episodes of scientific change can only be assessed as constituting progress if one takes into consideration the specific problem-situations that originated them (cf., Kuhn 1970, Laudan 1978, Shan 2019). Moreover, scientific goals do not encompass only epistemic goals (Brigandt, 2010, 2012), but involve also practical design-oriented goals that constrain the construction of scientific concepts (Wilson, 2006, 2017).

This kind of pragmatic encroachment of goals and problems into the semantics of science extended also to the most general theories and laws in science, and to the theoretical terms and concepts related to them. The history of many theoretical terms is, in fact, according to functionalist critiques, heavily dependent on the goals and the problems that scientists had to face. Indeed, for several important theoretical terms, such as ‘gene’ (Brigandt, 2010; MacLeod, 2012), ‘electron’ (Arabatzis, 2006, 2012), ‘electro-magnetic field’ (Nersessian, 1984, 2010), ‘temperature’ (Chang, 2004), and ‘magnetic pole’ (Steinle, 2012), the goals of the scientific community seem to constitute the very stability criterion of the term itself. If the representational and semantic features of these terms changed many times, the epistemic or pragmatic goal that they were constructed to achieve remained stable throughout their whole history. Goals do not only provide stability to theoretical concepts, but also shape their emergence. As recalled in several of the aforementioned case studies, the historical emergence of many theoretical terms seem to be the end-product of several tentative solutions of a specific scientific problem (cf., Nersessian 2010, Wilson 2006). Furthermore, according to functionalist critiques, many theoretical terms seem to play several important functions in scientific theorizing, functions that are equally (if not more) important than their representational one. As such, functionalist critiques stand in stark contrast with the traditional view of theoretical terms that assumed that their main function was always representational. Against this representation-centric background, functionalist critiques argued that in many episodes from the history of science the non-representational functions of a theoretical term were arguably more fundamental for the construction and use of a concept than its ability to represent worldly

[6] Note that what I call contextualist critiques is a rather heterogeneous group of writers, who, in different times, different contexts, and for different reasons, contested the alleged stability of scientific laws. Nonetheless they all share a common polemic aim and, as such, I pragmatically group them together here.

[7] Note that Cartwright’s critique of the received view of scientific laws is more general, and perhaps more radical, than the contextualist component on which I focus here, in that it argues against any semantic view of laws and theories (cf. Cartwright 1983).

[8] Note that patchwork approaches usually speak of scientific concepts and not of theoretical terms. Yet, their point is mainly a semantic one about the meaning of theoretical terms. Thus, since in this paper I reserved the term concept for the cognitive relata of our scientific terms, I will consider patchwork approaches to be focusing mainly on theoretical terms.

[9] There are also alternative accounts of scientific concepts that conceptualize this complexity in a different way, e.g., Novick (2023), Taylor and Vickers (2015). For the purpose of this paper, the differences between these accounts are not relevant.

[10] Just like what I call contextualist critiques (and perhaps even more), what I call functionalist critiques denotes an heterogeneous group of writers that contested, in different times, in different ways, and for different reasons, the representation-centric depiction of scientific laws and theoretical terms.

phenomena. Examples of these non-representational functions include boosting exploratory experiments (Feest, 2010, 2012), allowing richer measurement procedures (Chang, 2004), driving analogies (Nersessian, 2010), transferring inferential techniques across domains (Wilson, 2017), connecting different scales of scientific description (Bursten, 2018; Wilson, 2017), and many others.

Taken together, the contextualist and functionalist critiques to the received views of scientific laws and theoretical terms completely subvert the three assumptions on theoretical terms that we focused on in the last section, i.e., the intra-theory stability, the property-based content, and the representational role assumption. Theoretical terms are not always stable across the whole domain of application of a theory, but they often change drastically according to the contextual need of a specific domain of application. Moreover, the extensional content of theoretical terms is not always determined by the properties that their instances possess, since these properties may vary depending on which application of a theory one considers or on which diachronic version of a theory one focuses on. What is kept fixed throughout the applications and the life of a theoretical term is not necessarily its conceptual content nor its representational role, but it is instead often some other epistemic or pragmatic non-representational goal(s) that the term is supposed to perform. Consistently with this centrality of goals in the determination of the content of these theoretical terms, the semantics and pragmatics of these terms is equally not centered around their representational role, but it is instead centered around a different non-representational purpose that these terms serve within a theory, such as experimentation, abductive reasoning, scale management, and classification.

Such a different view of theoretical terms, depicted by pragmatist approaches to philosophy of science, implies also a different picture of theoretical concepts. All the three major characteristics that the standard view of theoretical terms in science projected on their cognitive counterparts, i.e., stability of content, property identification, and representational orientation, contrast in fact with the pragmatic picture of theoretical terms. Theoretical concepts such as FORCE and GENE cannot be that stable in content, since the meaning and reference of their related theoretical terms change with the evolution of the related theories and with their application to different domains. Due to these changes, many theoretical concepts seem to be often not identifiable via the properties that they encode, since these properties change frequently, but by the goal or purpose that they serve. Finally, just like in the case of theoretical terms, it is a mistake to hold representation as the main purpose of all theoretical concepts, since many of them arguably perform a vast range of functions in scientific inquiry, just like their linguistic counterparts. Such a change into the projected picture of theoretical concepts in science makes many theoretical concepts no longer so much alike to the common concepts studied by cognitive scientists. In fact, the content of concepts like DOG, TREE, RED, and the like, does not change much. From application to application, these concepts encode stable properties and they serve, mainly, to faithfully represent stable categories of our world. But, then, if not all theoretical concepts are like common concepts, which kind of concepts are these non-common theoretical concepts? In order to answer this question, we need to take a look at some recent perspectives on concepts in cognitive science. This will be the task of the next section.

4. Goal-derived concepts

In this section, we will focus on a particular class of concepts, i.e., what cognitive scientists call goal-derived concepts. We will see that these concepts have specific characteristics that, as I will argue in the next section, make them suitable candidates for being the cognitive relata of the theoretical terms that do not fit into the standard picture.

A superficial look at the psychological literature on concepts might cause the impression that we only possess the kind of concepts that I called common concepts. In fact, for decades most of psychological research (e.g., Margolis and Laurence 1999, Murphy 2002) on concepts has focused on these stable, representation-oriented bodies of knowledge that are identifiable via the properties that their instances (are likely to) possess. This is understandable, since common concepts (also known as taxonomic concepts) constitute the backbone of our conceptual development and of our intuitive picture of the world. Yet, in the last decades, empirical evidence shows that not all of our concepts are common concepts. An important class of non-common concepts is the class of goal-derived concepts.

Starting from the seminal work of Barsalou and his colleagues in the eighties (cf. Barsalou 1983, 1985, 1987, 1991, 2010, 2021, Ratneshwar, Barsalou, Pechmann, and Moore 2001), the properties of goal-derived concepts have been the subject of several empirical and theoretical studies in psychology. Goal-derived concepts, also known as role-governed concepts (Goldwater, Markman, & Stilwell, 2011; Markman & Stilwell, 2001) or ad-hoc concepts (Barsalou, 2010), are concepts that are constructed for achieving a specific goal, such as THINGS TO TAKE ON VACATION or CLOTHES FOR THE WINTER.[11] Barsalou and his colleagues have studied for years the properties of goal-derived concepts, also in connection with Barsalou's neo-empiricism about concepts (Barsalou, 1985, 1999). According to Barsalou (1983, 1985, 1987, 1991, 2021), goal-derived concepts have particular properties related to their purely functional role in our cognition that distinguish them from common concepts.

Perhaps the most important difference between goal-derived concepts and common concepts involves the nature of their graded structure (Barsalou, 1983, 1985, 1991; Voorspoels, Storms, & Vanpaemel, 2015), that is, the internal structure of their instances. Just like common concepts, goal-derived concepts are not flat, i.e., their instances differ with respect to their typicality. So that, when asked about instances of goal-derived concepts such as THINGS TO TAKE ON VACATION, people robustly judge instances to be more or less typical instances of the concept, just like they do for common concepts. Yet, there is an important difference in the relationship between the most typical instances of goal-derived concepts and their analogs in common concepts. The graded structure of common concepts is property-based, since it is mostly determined by the central tendency of a category (cf. Rosch and Mervis 1975), i.e., the features that instances of the concepts are very likely to exhibit. Thus, all typical instances of a common concept (e.g., BIRD) possess the same typical properties (e.g., flying, having feathers). The same is not true for goal-derived concepts, since their graded structure is instead mostly determined by ideal tendency (Barsalou, 1985, 1987; Voorspoels et al., 2015), i.e., the degree to which an instance is likely to be an ideal solution of the goal related to the concept. Thus, the typicality of instances of a goal-derived concept (e.g., THINGS TO TAKE ON VACATION) is not a function of the properties that they possess, but of the function they perform with respect to the goal related to the concept. This means that very typical instances of a goal-derived concept do not need to share many properties, because their typicality is not based on the properties that they encode, but on the function that they perform. This can be easily seen when looking at concept like THINGS TO TAKE ON VACATION: what are typical instances of such a concept? Very different objects, such as, perhaps, beach towel, sun cream, ski, books, passport,

[11] It should be noted that ad hoc concepts are sometimes considered a proper subset of goal-derived concepts, i.e. the contextual goal-derived concepts connected with unexpected goals such as THINGS TO TAKE FROM THE HOUSE IN CASE OF A FIRE. Moreover, some authors make goal-derived concepts a proper subset of role-concepts (cf. Markman and Stilwell 2001). Yet, the distinction between ad hoc concepts and goal-derived concepts is, as Barsalou stresses several times, mostly a pragmatic one and, as such, not so sharp. And the same is held by Markman and Stilwell (2001) for the distinction between goal-derived concepts and role-governed concepts. I will thus blur these distinctions in this paper, focusing on goal-derived concepts and on the fundamental distinction between these concepts and common concepts.

sunglasses, and so on. All these objects are very typical instances of THINGS TO TAKE ON VACATION, but they have almost no property in common. What they have in common is the role that they typically perform, namely, to be taken on vacation. This difference in the graded structure between goal-derived concepts and common concepts reflect their different main purpose in a conceptual system (cf. Barsalou 2003, 2021): while common concepts are the building block of our ontology and, as such, their main job is to faithfully represent basic categories, goal-derived concepts are constructed to help agents to achieve their goals and, as such, their main job is to maximize the likelihood of achieving a certain goal.

Another important difference between common concepts and goal-derived concepts concerns the way in which these concepts are constructed. In fact, while common concepts are usually acquired by exemplar learning and general abstraction (cf., Murphy 2002), goal-derived concepts are mostly acquired by contextual modifications of other concepts (Barsalou, 1991). That is, whereas common concepts are usually acquired by abstracting a certain common structure from a set of exemplars of the concepts (either a set of prototypical features, an intuitive theory, a set of paradigmatic exemplars, depending on the specific theory of concept learning one favors, cf., Murphy 2002), goal-derived concepts are acquired by modification of other concepts in the context of optimizing conceptual resources to achieve a certain goal. More specifically, according to Barsalou (1991, 2021), goal-derived concepts are built in two steps. First, agents construct a frame (Barsalou, 1992; Barsalou & Hale, 1993) for a given event or situation that they want to pursue (e.g., a vacation). Then, agents focus on a certain attribute of this event-related frame (e.g., things to take), conceptualizing potential values for such attribute from instances of related concepts (e.g., beach towel, passport, sunglasses, …). In the second step, these potential values, i.e., the soon-to-be instances of the new goal-derived concept, get organized with respect to how well they satisfy the ideal goals and constraints related to the attribute (e.g., minimize space in the luggage, minimize weight, maximize usefulness on vacation, maximize weather conditions appropriateness, …). At the end of this process, a new goal-derived concept gets established in the memory (i.e., THINGS TO TAKE ON VACATION), together with its graded structure centered around the ideal tendency related to the goal under focus (cf., Barsalou 1991, 2021, Ratneshwar et al. 2001).[12]

Two further noticeable differences between common concepts and goal-derived concepts relate to the degree of stability of these concepts and to the homogeneity of the categories that they denote. With regards to the degree of stability of these two kinds of concepts, goal-derived concepts are far less stable than common concepts (Barsalou, 1987, 2010, 2021). This is because of the contextuality of their purpose and the related flexibility of their graded structure which, since it is organized around ideal tendency, changes consistently with the unfolding of agents' plans and actions. For this reason, goal-derived concepts exhibit far more radical contextual effects in their instantiations than common concepts, changing depending on the specific optimizations and constraints that the agents' plans exhibit (cf., Barsalou 1987, Ratneshwar et al. 2001).[13] Finally, the categories represented by goal-derived concepts are often very heterogeneous clusters of things. This is because of the mainly non-representational purpose that goal-derived concepts serve in our conceptual system. In fact, while common concepts, whose purpose is mainly representational, are organized around central tendency and, therefore, they group together things that share many properties, goal-derived concepts need to maximize the relevancy of entities to human goals. Thus, goal-derived concepts often group together very different things that share only functional properties in relation to our contextual goal (e.g., the many different things that we take on vacation). As such, while common concepts usually denote homogeneous categories that respect environmental features (cf., Rosch and Mervis 1975), goal-derived concepts usually denote heterogeneous categories that often cross-cut physical and environmental distinctions (and thus related natural kinds).

Table 1
The main differences between common concepts and goal-derived concepts.

	Common concepts	Goal-derived concepts
Graded structure	Central tendency	Ideal Tendency
Main purpose	Representation	Contextual goal-achievement
Mode of construction	Exemplar learning	Conceptual modification
Stability	High	Low
Categories	Homogeneous kinds	Heterogeneous kinds

We saw then how goal-derived concepts represent a subclass of concepts that has very different properties than the ones exhibited by common concepts. Specifically, we saw how goal-derived concepts differ from common concepts in the organization of their graded structure, their main purpose, their mode of construction, their stability, and their related categories. These difference between common concepts and goal-derived concepts are summarized in the following table (cf. Table 1):

5. Theoretical concepts as goal-derived concepts

It is now time to put the pieces together. We saw in Section 3 that pragmatic understandings of theoretical terms in science arguably showed that some important theoretical concepts do not fit into the standard picture that understood them as common concepts. In the previous section, we looked at how recent cognitive science conceptualizes a specific class of concepts, i.e., goal-derived concepts, that possess properties that make them very different in their nature and use from common concepts. In this section, I will argue that the theoretical concepts that do not fit into the standard picture should be considered goal-derived concepts. More specifically, we will see that understanding these theoretical concepts as goal-derived concepts adequately accounts for how, according to our best philosophical pictures of scientific activity, several important theoretical concepts are constructed and used in scientific practice.

We concluded Section 3 by stressing that, differently from what was commonly assumed by much of twentieth-century philosophy of science, several important theoretical concepts in science appear not so stable in content, they are mostly identifiable by the goal that they perform (and not by the features that they encode), and their main function seem to be to help the scientific community to achieve some pragmatic non-representational goal rather than to faithfully represent hidden entities. These characteristics make these theoretical concepts different from common concepts like DOG and TREE, since common concepts are very stable in content, they are identifiable by the properties that they encode, and their main purpose is to faithfully represent entities. In Section 4, we saw then how these properties that make these theoretical concepts different from common concepts (e.g., contextual instability, goal-centrality, and primarily non-representational purpose) are among the characteristics that, according to cognitive scientists, make goal-derived concepts different from common concepts. Given these common characteristics, I propose to understand the theoretical concepts in science that do not fit into the standard picture as goal-derived concepts. In order to substantiate my thesis, let us look at whether and how we can make sense of the construction and the use of theoretical concepts in science as the construction and use of goal-derived concepts. We highlighted in the last section five specific characteristics of

[12] Note that, despite Barsalou's model of the construction of goal-derived concepts uses a specific theory of conceptual representation (i.e., frames), it has been modeled also in other psychological theories of concepts, such as ideal dimensional theories (Voorspoels, Vanpaemel, & Storms, 2011) and conceptual spaces (Coraci, 2022).

[13] It should be noted that also the content of common concepts might change from context to context. Nevertheless, contextual effects are far more radical in the case of goal-derived concepts (cf. Barsalou 1987, 2021).

the construction and use of goal-derived concepts. Specifically, we saw that goal-derived concepts have a graded structure mostly determined by ideal-tendency, their main purpose is contextual goal-achievement, they are often constructed via conceptual modification, they are not so stable across different contexts, and they often denote heterogeneous categories. Let us see whether and how theoretical concepts, *qua* goal-derived concepts, exhibit these properties.

Graded structure The graded structure of goal-derived concepts is mostly determined by ideal-tendency and not, as in the case of common concepts, by central-tendency (cf. Barsalou 1983, 1985). This means that the degree of typicality of an instance of a goal-derived concept is not based on the features that such an instance possesses, but on the function it performs (i.e., how much it can contribute to achieving the goal related to the concept). An important byproduct of such a function-based graded structure is that very typical instances often do not have many properties in common (recall the examples of typical instances of THINGS TO TAKE ON VACATION). Indeed, several important theoretical concepts in science seem to exhibit precisely this kind of function-based graded structure and the related differences between typical instances. Important differences between typical instances of theoretical concepts were in fact the reason why pragmatist approaches started to conceptualize theoretical concepts as patchworks (cf. De Benedetto 2021, Haueis 2021b, Novick and Haueis 2023, Wilson 2006, 2017). Behind the apparent uniformity of a single theoretical concept such as FORCE or HARDNESS, pragmatist critiques uncovered an heterogeneous set of instances. So that, the FORCE concept arguably denotes an heterogeneous set of instances in classical mechanics (cf., Wilson 2006, pp. 157–165, 175–182), lumping together real forces with different entities such as net losses and gains of momentum of fluid particles (Wilson, 2006, pp. 58–159), frictional net effects (Wilson, 2006, p. 175), and measures of internal stress of elastic bodies (Wilson, 2006, p. 176). Even more striking are the differences between typical instances of HARDNESS in material science (cf. Wilson 2006, pp. 335–355). A typical instance of HARDNESS can be, in fact, depending on the patch under consideration, an automobile tire, a nylon, a window glass, a diamond, or a chrome knife (cf. Wilson 2006, p. 338). What makes all these different entities typical instances of the theoretical concept HARDNESS is then not the properties that they share (which are very few, if any), but the function that they perform (namely, to perform well within the contextual hardness test suitable to the material under consideration, cf. Wilson 2006, p. 336). In cognitive terms, we can say that the graded structure of theoretical concepts like HARDNESS and FORCE seems then centered around ideal-tendency and not around central-tendency, just like goal-derived concepts and unlike common concepts.

Main purpose Differently from common concepts, whose main purpose is to represent environmentally faithful homogeneous categories, the main purpose of goal-derived concepts is to help agents to achieve whatever contextual goal prompted the creation of the concept (cf. Barsalou 1991, 2021, Ratneshwar et al. 2001). Thus, the purpose of goal-derived concepts is contextual in nature and it heavily depends on the actions and the goals of the agents that created them. Such a contextuality of purpose is also stressed by pragmatist pictures of science in their description of the different functions performed by theoretical concepts in scientific inquiry. Different theoretical concepts serve, in fact, different purposes, depending on the specific problem and the specific situations that prompted scientists to create them. Moreover, these contextual purposes are often constitutive of the concept itself, in that they represent the most stable feature in the dynamic life of a theoretical concept (Arabatzis, 2006; Brigandt, 2010, 2012). We can see such a pivotal role of contextual purposes in the life of a theoretical concept by looking at some of the examples that we briefly mentioned in Section 3. For instance, the main purpose of the concept of GENE in evolutionary biology was (and still is) to function as a unit of heredity and as a causal agent for biological traits (Brigandt, 2010; MacLeod, 2012), while the main purpose of the concept of IMPLICIT MEMORY in cognitive neuropsychology was to support experimental operation on dissociations in memory tests (Feest, 2010). Another example of a fundamental non-representational purpose in the life of a theoretical concept is given by the concept of MAGNETIC POLES, whose main purpose was to function as a source and reference frame for magnetic forces in modern physics (Steinle, 2012). In all these different examples, the construction and use of a theoretical concept is mostly governed by a non-representational purpose, while the function of representing entities is always of secondary importance.

Mode of construction Goal-derived concepts are usually constructed via contextual modifications of other concepts (cf., Barsalou 1991, 2021). Such contextual conceptual modification is intertwined with planning and practical problem-solving. When faced with a specific problem or goal, we often modify our concept to tailor them to the task at issue, thus creating a related goal-derived concept. As we saw in Section 4, Barsalou and colleagues (Barsalou, 1991, 2021; Ratneshwar et al., 2001) developed detailed models of such construction that understand the creation of goal-derived concepts as a specific kind of modification of conceptual frames (i.e., Barsalou's preferred system of knowledge representation, cf. Barsalou 1992, Barsalou and Hale 1993) where the values and the attributes related to a given concept are modified to maximize relevance to the specific goal that the agent sets herself. For instance, in deriving the goal-derived concept CLOTHES FOR THE WINTER, we modify our general concept of clothing by focusing on the attributes that are most relevant for winter time, such as warmth and ability to protect from rain and wind. Such a goal-relevance maximizing contextual modification of our concepts is also at work, according to contemporary philosophy of science, behind the scientists' construction of theoretical concepts. A paradigmatic example of this modeling of conceptual creativity in science as goal-derived contextual modification of previously acquired concepts is Nersessian's (Nersessian, 2010) fine-grained bootstrapping model of concept creation. In her detailed case study on Maxwell's creation of the concept of FIELD, Nersessian (1984, 2010) argues that behind the construction of this theoretical concept lies an iterated process of modification of our concepts, where concepts from different domains (i.e., fluid mechanics, machine mechanics, and continuum mechanics) get recombined in new ways (i.e., vortex fluid, vortex idle wheel, and elastic vortex idle wheel models), through analogies and abductive reasoning, in order to solve a given problem (i.e., to give a unified account of electric and magnetic forces). This centrality of epistemic and pragmatic goals for understanding how scientists construct new theoretical concepts from old ones is shared by several other models of conceptual change in science (e.g., Andersen et al. 2006, Brigandt 2010, 2012, Kitcher 1995, Laudan 1984). Indeed, as we already saw in the previous paragraph of this section, the construction of several important theoretical concepts in the history of science such as GENE, FIELD, IMPLICIT MEMORY, and MAGNETIC POLES, is arguably guided by the desire of the related scientific community to achieve a certain specific goal.

Stability Because of their entanglement with agents' goals, goal-derived concepts are not very stable, changing often quite radically from one context of use to another one (cf. Barsalou 1987, 2021). Thus, the extension of a goal-derived concept such as THINGS TO TAKE ON VACATION is heavily affected by the specific context for which we retrieve the concept (Which type of vacation is it? With whom are we going? What does the weather forecasting say? ...). Analogous contextual effects have been observed also in theoretical concepts in science. As a matter of fact, as we recalled in Section 3, the hidden contextuality of theoretical terms (and thus of the related concepts) was one of the main issues that fueled pragmatist critiques to the standard view of scientific laws and theoretical terms. Behind a single theoretical concept, such as FORCE or GENE, argued pragmatist critiques, often lies a patchwork of different localized usage (cf. Cartwright 1983, 1999, Wilson 2006, 2017). So that, in classical mechanics, one finds that the concepts of

FORCE is instantiated very different in different parts of the theory. The contextuality of theoretical concepts in science can moreover not only be seen in a synchronic way, by comparing different parts of a theory, but also diachronically, by comparing how theoretical concepts change with the evolution of the theory to which they belong. Examples of this diachronic contextuality can be found in several of the theoretical concepts we discussed so far, including, for instance, the many different modifications of the concepts of MAGNETIC POLES (Steinle, 2012), the troubled life of the ELECTRON concept (Arabatzis, 2006), the changes in the concept of DOMINANCE in the study of heredity (Shan, 2020b), or the goal-oriented modifications of the concept of GENE (Brigandt, 2010; MacLeod, 2012).

Categories Goal-derived concepts, since they are created to support goal-derived action and, thus, they do not perform mainly a representational role, denote often heterogeneous categories that cross-cut environmental distinctions and related natural kinds (cf. Barsalou 2003, 2021). Thus, the categories denoted by goal-derived concepts often grouped together different things that do not share many properties other than the function that they perform. Although the naturalness of scientific kinds is an extremely controversial topic in the metaphysics of science, we saw in Section 3 how several important theoretical concepts such as FORCE, GENE, HOMOLOGY, HARDNESS, NEURAL COLUMN, SPECIES, and ATTENTION are increasingly thought to represent functional categories that group together multiple homogeneous categories into one patchwork. Whatever one's theory of natural kinds might be, we can see that there is an increasing trend in philosophy of science towards conceptualizing the categories denoted by theoretical concepts as organized in virtue of the function that their instances perform, rather than the properties shared by these instances. This is yet another aspect with respect to which the characteristics that several important theoretical concepts, according to contemporary philosophy of science, exhibit are the characteristics of goal-derived concepts.

We saw then how the construction and the use of several important theoretical concepts in science arguably exhibits the specific characteristics of goal-derived concepts. More specifically, we saw that the construction and use of theoretical concepts such as FORCE, GENE, HARDNESS, IMPLICIT MEMORY, and FIELD, is structured around the contextual achievement of certain specific goals of the related scientific community. This analysis substantiated my thesis that the theoretical concepts that do not fit the standard picture are best understood as goal-derived concepts.

It is important to stress that the exact extension of this group of theoretical concepts that are best understood as goal-derived concepts is unclear. In fact, as we saw in this section, the identification of a theoretical concept as a goal-derived concept, rather than a common concept, is dependent on an adequate historical analysis of the construction and use of the concept, together with a philosophical interpretation of the relevant conceptual features (i.e., graded structure, main purpose, mode of construction, stability, and categories). As such, the identification of a given theoretical concept as goal-derived (or as a common concept) can only proceed in a case-by-case manner, comparing different cases and pragmatically weighting the significance of the relevant features of the concept. The examples analyzed in this paper, since they spawn across many different sciences and different centuries, suggest that many theoretical concepts might be goal-derived concepts. At the same time, however, a quick look at some of our best scientific theories reveals some theoretical concepts that appear perfectly described as common concepts, such as, for instance, ENERGY, ACID, TEMPERATURE, NEUTRINO, and many others. Given this situation, a methodological consequence of the present analysis is that we should be wary of any unrestricted philosophical generalization about theoretical concepts.[14]

The identification of some important theoretical concepts with goal-derived concepts is not just important for understanding the cognitive dimension of our theoretical terms in science (and, in turn, of scientific laws and theories), but it has important implications for our general understanding of science. In fact, goal-derived concepts and common concepts are not just different classes of concepts that we construct and use in different ways, but they instantiate two opposite ideals towards which conceptual systems can be thought as being oriented. In Barsalou's own words:

> At the one extreme, people are intuitive taxonomists. Their goal is to discover the categorical structure of the world, develop taxonomic systems that represent this structure, and establish background theories that frame these taxonomies. At the other extreme, people are goal achievers who organize knowledge to support situated action. On this view, the primary organization of the conceptual system supports executing actions effectively in the environment, with taxonomic hierarchies constituting a secondary-level of organization that supports this activity. (...) The presence of categories that arise specifically to achieve goals intimates the importance of goal achievement in organizing the conceptual system (Barsalou, 2003, p. 546).

These two different ideals towards which conceptual system can be said to be oriented, that is, the intuitive taxonomist ideal and the goal achievement ideal, are, according to Barsalou, closely related with the kind of concepts that compose the conceptual system. The more a conceptual system is made out of common concepts, the more it naturally leans towards the intuitive taxonomist ideal, whereas a conceptual system in which many concepts are goal-derived concepts leans instead towards the ideal of goal achievement. This is how the above identification of several important theoretical concepts in science with goal-derived concepts tells us something important for our image of the most general aims of scientific inquiry. If, in fact, several theoretical concepts within our best scientific theories are goal-derived concepts, as the above analysis has arguably showed, then our best scientific theories are structured for maximizing action-supporting goal achievement. This means that, from a conceptual point of view, science does not seem to be perfectly structured for representing the world in the most faithful way, but rather for helping us to achieve our human goals. Of course, these two ideals are not necessarily contrasting with each other, but they underlie very different views of science. The goal-achievement ideal is closely connect with the human-centered functionalist view of scientific inquiry which is common to several broadly pragmatist views of scientific progress, scientific kinds, and scientific rationality. These views include, for instance, functionalist theories of scientific progress (Kuhn, 1970; Laudan, 1978; Shan, 2019), pragmatic accounts of scientific kinds (Chang, 2004, 2012; Hacking, 2007), and value-laden theories of scientific rationality (Douglas, 2009; Laudan, 1984; Longino, 1990; Mitchell, 2009). What all these views have in common is the ideal that science is an inherently human activity built upon human goals and values. Such an ideal is then maximized by adding to our scientific theories some central goal-derived concepts. This is how the present proposal fits with other pragmatist-functionalist views of concepts, laws, terms, kinds, progress, and rationality in science. By focusing on the hitherto under-discussed cognitive dimension of theoretical concepts, we saw how several important theoretical concepts used by our best scientific theories are goal-derived concepts. Such a specific thesis provides independent cognitive evidence for viewing scientific inquiry, even in its most abstract elements, as oriented towards the satisfaction of human goals. This centrality of human goals in all the parts of scientific inquiry implies also a re-orientation of some of our most general intuitive expectations about science and scientific concepts. In particular, the image of science as the domain of neat taxonomies and clear-cut concepts appears, despite all its intuitive appeal, yet another popular philosophical myth that obscures the very human character of scientific activity.

[14] Such methodological carefulness is also advisable when dealing with theoretical terms and their (meta)semantics. See De Benedetto (2024) for an argument against common philosophical generalizations about theoretical terms.

6. Conclusion

Let us recap the main steps of the present work. I started by inquiring into an hitherto under-discussed aspect of theoretical terms in science, namely, the nature of their cognitive relata, i.e., what I call theoretical concepts. I argued that the implicit picture of theoretical concepts of much of twentieth-century philosophy of science understood them as common concepts. Specifically, we saw how the standard view of theoretical terms in twentieth-century philosophy of science projected a standard picture of theoretical concepts as stable in content, centered around the features that their instances possess, whose purpose is mainly representational. Then, recalling the pragmatist critiques to the standard view of scientific laws and theoretical terms, I argued that contemporary philosophy of science gives us a different picture of several important theoretical concepts, stressing instead their inherently contextual and functional nature. Building upon recent cognitive science, I thus argued that these theoretical concepts that do not fit into the standard picture are best described as goal-derived concepts. Specifically, I showed how the construction and the use of several important theoretical concepts in science, as described by our best pictures of scientific activity, arguably exhibit the typical characteristics of goal-derived concepts construction and use, in that it is fundamentally structured around the contextual achievement of some specific goals of the related scientific community.

This identification of several important theoretical concepts in science with goal-derived concepts is then another example of the centrality of human goals to scientific inquiry. By focusing on the cognitive dimension of theoretical terms, we found independent support for a human-centered functionalist conception of scientific inquiry that is common to several, related views on scientific progress, scientific kinds, and scientific rationality. The exact relationships between these positions represent promising directions for future work. Another natural continuation of the present work would be to extend the scope of its analysis, by searching for other goal-derived theoretical concepts across science. Finally, a third possible direction for future work would be to extend the present account of several important scientific concepts as goal-derived concepts by connecting its insights with the ones of other kinds of critiques of the default attitude towards concepts and conceptual change in analytic philosophy (see Section 2), such as contextualist theories of concepts in philosophy of mind, deflationist theories of properties in metaphysics, and programs of conceptual engineering in metaphilosophy.

CRediT authorship contribution statement

Matteo De Benedetto: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

I am indebted to Michele Luchetti, Yafeng Shan, all the members of the FPTB group in Bochum, and two anonymous referees for helpful comments on earlier drafts of this paper. I am also indebted to audiences in Belgrade, Turin, and Bochum for constructive feedback.

Funding

My work on this paper was funded by the Emmy-Noether project "From Perception to Belief and Back Again", Deutsche Forschungsgemeinschaft (BR 5210/1-1).

References

Andersen, H., Barker, P., & Chen, X. (1996). Kuhn's mature philosophy of science and cognitive psychology. *Philosophical Psychology*, *9*(3), 347–363.
Andersen, H., Barker, P., & Chen, X. (2006). *The cognitive structure of scientific revolutions*. Cambridge: Cambridge University Press.
Andreas, H. (2021). Theoretical terms in science. In Edward N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2021 Edition). < https://plato.stanford.edu/archives/fall2021/entries/theoretical-terms-science/ >.
Arabatzis, T. (2006). *Representing electrons: a biographical approach to theoretical entities*. Chicago.: University of Chicago Press.
Arabatzis, T. (2012). Experimentation and the meaning of scientific concepts. In U. Feest, & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 156–173). De Gruyter.
Arabatzis, T., & Kindi, V. (2008). The problem of conceptual change in the philosophy and history of science. In S. Vosniadu (Ed.), *International handbook of research on conceptual change* (pp. 345–373). Routledge.
Balzer, W., Moulines, C. U., & Sneed, J. (1987). *An architectonic for science, the structuralist program*. Dordrecht: D. Reidel Publishing Company.
Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.
Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 629–654.
Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge: Cambridge University Press.
Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory*: *vol. 27*, (pp. 1–64). San Diego, CA: Academic Press.
Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer, & E. Feder Kittay (Eds.), *Frames, fields, and contrasts* (pp. 21–74). London: Routledge.
Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.
Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*, 513–562.
Barsalou, L. W. (2010). Ad hoc categories. In P. C. Hogan (Ed.), *The Cambridge encyclopedia of the language sciences* (pp. 87–88). Cambridge University Press.
Barsalou, L. W. (2021). Categories at the interface of cognition and action. In C. Mauri, I. Fiorentini, & E. Goria (Eds.), *Building categories in interaction: linguistic resources at work* (pp. 35–73). John Benjamins.
Barsalou, L. W., & Hale, C. (1993). Components of conceptual representation: from feature lists to recursive frames. In I. Van Mechelen, & et al. (Eds.), *Categories and concepts: theoretical views and inductive data analysis* (pp. 97–144). Cambridge (MA): Academic Press.
Batterman, R. (2001). *The devil in the details: asymptotic reasoning in explanation, reduction, and emergence*. Oxford: Oxford University Press.
Batterman, R. (2013). The tyranny of scales. In *The oxford handbook of philosophy of physics* (pp. 255–286). Oxford: Oxford University Press.
Bird, A. (2002). Kuhn's wrong turning. *Studies in History and Philosophy of Science*, *33*, 443–463.
Brigandt, I. (2010). The epistemic goal of a concept: Accounting for the rationality of semantic change and variation. *Synthese*, *177*(1), 19–40.
Brigandt, I. (2012). The dynamics of scientific concepts: The relevance of epistemic aims and values. In U. Feest, & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 75–104). De Gruyter.
Bursten, J. R. (2018). Smaller than a breadbox: Scale and natural kinds. *British Journal for Philosophy of Science*, *69*(1), 1–23.
Cappelen, H. (2018). *Fixing language: an essay on conceptual engineering*. Oxford: Oxford University Press.
Carnap, R. (1961). On the use of Hilbert's ϵ-operator in scientific theories. In Y. Bar-Hillel, & et al. (Eds.), *Essays in the foundation of mathematics* (pp. 154–164). Jerusalem: Magnes Press.
Carnap, R. (1966). *Philosophical foundations of physics: an introduction to the philosophy of science*. New York: Basic Books.
Cartwright, N. (1983). *How the laws of physics Lie*. Oxford: Oxford University Press.
Cartwright, N. (1999). *The dappled world: a study of the boundaries of science*. Cambridge: Cambridge University Press.
Chang, H. (2004). *Inventing temperature: measurement and scientific progress*. New York: Oxford University Press.
Chang, H. (2012). *Is water H2O? evidence, realism and pluralism*. New York: Springer.
Coraci, D. (2022). A unified model of ad hoc concepts in conceptual spaces. *Minds and Machines*, *32*(2), 289–309.

De Benedetto, M. (2021). Taming conceptual wanderings: Wilson-structuralism. *Synthese*, *199*, 13225–13246.
De Benedetto, M. (2024). What conceptual engineering can learn from the history of philosophy of science: Healthy externalism and metasemantic plasticity. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *14*(1), 1–24.
Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press.
Feest, U. (2010). Concepts as tools in the experimental generation of knowledge in cognitive neuropsychology. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, *4*, 173–190.
Feest, U. (2012). Exploratory experiments, concept formation, and theory construction in psychology. In U. Feest, & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 167–190). de Gruyter.
Feyerabend, P. (1962). Explanation, reduction and empiricism, In H. Feigl, & G. Maxwell (Eds.), *Scientific explanation, space, and time, minnesota studies in the philosophy of science*, (vol. III). (pp. 28–97). Minneapolis: University of Minnesota Press.
Giere, R. (1988). *Explaining science: a cognitive approach*. Chicago: University of Chicago Press.
Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition*, *118*(3), 359–376.
Hacking, I. (1983). *Representing and intervening: introductory topics in the philosophy of natural science*. Cambridge: Cambridge University Press.
Hacking, I. (2007). Natural kinds: Rosy dawn, scholastic twilight. *Royal Institute of Philosophy Supplement*, *61*, 203–239.
Haueis, P. (2021a). The death of the cortical column? Patchwork structure and conceptual retirement in neuroscientific practice. *Studies in History and Philosophy of Science Part A*, *85*, 101–113.
Haueis, P. (2021b). A generalized patchwork approach to scientific concepts. *The British Journal for the Philosophy of Science*, https://www.journals.uchicago.edu/doi/10.1086/716179.
Hempel, C. (1952). *Fundamentals of concept formation in empirical science*. Chicago: University of Chicago Press.
Kindi, V. (2012). Concept as vessel and concept as use. In U. Feest, & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 23–46). De Gruyter.
Kitcher, P. (1995). *The advancement of science: science without legend, objectivity without illusion*. Oxford: Oxford University Press.
Kuhn, T. S. (1970). *The structure of scientific revolution* (2nd ed.). Chicago: University of Chicago Press.
Kuhn, T. S. (1974). Second thoughts on paradigms. In F. Suppe (Ed.), *The structure of scientific theories* (pp. 459–482). Chicago: University of Illinois Press.
Laudan, L. (1978). *Progress and its problems: towards a theory of scientific growth*. Berkeley: University of California Press.
Laudan, L. (1984). *Science and values*. Berkeley: University of California Press.
Longino, H. E. (1990). *Science as social knowledge: values and objectivity in scientific inquiry*. Princeton (NJ): Princeton University Press.
Machery, E. (2009). *Doing without concepts*. New York.: Oxford University Press.
MacLeod, M. (2012). Rethinking scientific concepts for research contexts: The case of the classical gene. In U. Feest, & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 47–74). De Gruyter.
Margolis, E., & Laurence, S. (1999). *Concepts: core readings*. Cambridge (MA): MIT Press.
Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, *13*(4), 329–358.
Mitchell, S. (1997). Pragmatic laws. *Philosophy of Science*, *64*(4), 468–479.
Mitchell, S. (2009). *Unsimple truths: science, complexity, and policy*. Chicago (IL): University of Chicago Press.
Murphy, G. (2002). *The big book of concepts*. Cambridge (MA): MIT Press.
Nersessian, N. J. (1984). *Faraday to Einstein: constructing meaning in scientific theories*. Drodrecht: Springer.
Nersessian, N. J. (2010). *Creating scientific concepts*. Cambridge (MA): MIT Press.
Novick, A. (2018). The fine structure of 'homology'. *Biology and Philosophy*, *33*(1.2), http://dx.doi.org/10.1007/s10539-018-9617-3.
Novick, R. (2023). The neutral theory of conceptual complexity. *Philosophy of Science*, 1–17. http://dx.doi.org/10.1017/psa.2023.25.
Novick, A., & Doolittle, W. F. (2021). Species without species. *Studies in History and Philosophy of Science Part A*, *87*, 72–80.
Novick, R., & Haueis, P. (2023). Patchworks and operations. *European Journal for Philosophy of Science*, *13*(15), http://dx.doi.org/10.1007/s13194-023-00515-y.
Psillos, S. (1999). *Scientific realism: how science tracks truth*. London: Routledge.
Psillos, S. (2000). Rudolf Carnap's 'theoretical concepts in science'. *Studies in History and Philosophy of Science*, *31*, 151–172.
Putnam, H. (1973). Explanation and reference. In *Reprinted in philosophical papers vol. 2: mind, language and reality* (pp. 196–214). Cambridge (MA): Cambridge University Press, (1995).
Putnam, H. (1975). The meaning of 'meaning'. In *Reprinted in philosophical papers 2: mind, language and reality* (pp. 215–271). Cambridge (MA): Cambridge University Press, (1995).
Ratneshwar, S., Barsalou, L. W., Pechmann, C., & Moore, M. (2001). Goal derived categories: The role of personal and situational goals in category representation. *Journal of Consumer Psychology*, *10*, 147–157.
Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structures of categories. *Cognitive Psychology*, *7*, 573–605.
Shan, Y. (2019). A new functional approach to scientific progress. *Philosophy of Science*, *86*(4), 739–758.
Shan, Y. (2020a). Kuhn's "wrong turning' and legacy today. *Synthese*, *197*(1), 381–406.
Shan, Y. (2020b). *Doing integrated history and philosophy of science: A case study of the origin of genetics*. Cham: Springer.
Steinle, F. (2012). Goals and fates of concepts: The case of magnetic poles. In U. Feest, & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 105–126). De Gruyter.
Taylor, H. (2023). Attention as a patchwork concept. *European Journal for Philosophy of Science*, *13*(3), 1–25.
Taylor, H., & Vickers, P. (2015). Conceptual fragmentation and the rise of eliminativism. *European Journal for Philosophy of Science*, *7*(1), 17–40.
Thagard, P. (1992). *Conceptual revolutions*. Princeton (NJ): Princeton University Press.
van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.
Voorspoels, W., Storms, G., & Vanpaemel, W. (2015). Idealness and similarity in goal-derived categories: A computational examination. *Memory & Cognition*, *41*, 312–327.
Voorspoels, W., Vanpaemel, W., & Storms, G. (2011). A formal ideal-based account of typicality. *Psychonomic Bulletin & Review*, *18*, 1006–1014.
Wilson, M. (2006). *Wandering significance: an essay on conceptual behavior*. Oxford: Clarendon Press.
Wilson, M. (2017). *Physics avoidance: and other essays in conceptual strategy*. Oxford: Oxford University Press.