Volume XIII ◇ Issue 1

2022

Logos & Episteme

an international journal of epistemology

Romanian Academy
Iasi Branch
◆◆◆
"Gheorghe Zane" Institute for Economic and Social Research

Contact address:

Institutul de Cercetări
Economice şi Sociale „Gh.Zane”

Iaşi, str.T.Codrescu, nr.2, cod 700481

Tel/Fax: 004 0332 408922

Email: logosandepisteme@yahoo.com

http://logos-and-episteme.acadiasi.ro/

https://www.pdcnet.org/pdc/bvdb.nsf/journal?openform&journal=pdc_logos-episteme

ISSN 2069-053

TABLE OF CONTENTS

RESEARCH ARTICLES

STRUCTURAL-EPISTEMIC INTERDISCIPLINARITY AND THE NATURE OF INTERDISCIPLINARY CHALLENGES

Cătălin BĂRBOIANU

ABSTRACT: Research on interdisciplinarity has been concentrated on the methodological and educational aspects of this complex phenomenon and less on its theoretical nature. Within a theoretical framework specific to the philosophy of science, I propose a structural scheme of how interdisciplinary processes go, focusing on the concepts of availability of the methods, concept linking, and theoretical modeling. In this model, the challenges interdisciplinarity is claimed to pose to its practitioners are of the same *nature* as the challenges scientists encounter within the evolution of their own disciplines.

KEYWORDS: interdisciplinarity, transdisciplinarity, disciplinarity, structure of knowledge, theoretical modelling

Introduction

In the last few decades, the concept of interdisciplinarity has become central for the metascientific analyses of both scientific practice and higher education in the sciences. Researchers have focused on these two aspects – methodological and educational – as determinant and in many views exhaustive for the nature of interdisciplinarity. Motivations and goals of interdisciplinary practice have also been associated with these two aspects and as such have been reduced to solving complex problems and enhancing education.

How have we referred to interdisciplinarity since it was acknowledged? We have observed and participated in the *phenomenon* of interdisciplinarity in scientific practice, and we have emphasized the good track record of its results. Such recognition has encouraged us to search for and practice interdisciplinarity whenever possible, despite the various theoretical and methodological problems it certainly poses. The phenomenon gained such importance that within scientometry, various formulas have been proposed for indicators measuring the interdisciplinary approach and content of journals.[1] Labeling with the attribute of 'interdisciplinary'

[1] See for instance (Leydesdorff & Rafols 2011) for a comparative analysis of the indicators used to measure interdisciplinarity of the journals.

has become almost a "fashion trend:" researchers are claiming their proposed projects as interdisciplinary, organizations are advertising that they hire interdisciplinary teams, and candidates are presenting themselves in their résumés as interdisciplinary-oriented. Of course, there is nothing wrong in that, since interdisciplinarity is a real phenomenon, and searching for interdisciplinarity is entirely justified epistemologically. But is interdisciplinarity really a virtue "higher" than the virtues of the scientific practice performed within one discipline, or of a special nature? The virtue of interdisciplinarity that made us "advertise" it with such persistence and shape it as the new orthodoxy of current scientific practice is actually established by the complex challenges it poses to its practitioners. If scientists overcame these special challenges regarding special problems (and as such, contributed to the good track record of the practice), we were justified in claiming this special status of interdisciplinarity. However, inquiring deeper into the nature of interdisciplinarity, we may find that the answer to the question of "higher virtue" is not straightforward.

Literature on interdisciplinarity has presented it as an ongoing, growing phenomenon focusing more on the problems, barriers, and challenges it poses from methodological, educational, and research-community-related perspectives, and less on its theoretical nature. The research has not yet reached a crystallized theory of interdisciplinarity by which to account theoretically for those problems, to provide criteria of adequacy and norms of optimization for the practice of interdisciplinarity, and to make predictions for the developments of the disciplines themselves, as well as to provide norms for the interdisciplinary education. At most, research is in the stage of constituting a conceptual framework for such theory, still concentrating on typologies, and revealing so *many* apparently independent aspects of interdisciplinarity as phenomenon, that prospects of establishing such a framework are dim. If this is the case, we put the cart before the horse when focusing on practical problems and developing programs of management rather than clarifying the nature of interdisciplinarity? Overall, what discipline or disciplines would be entitled to deal with such a theory – philosophy of science, cognitive sciences, educational sciences, or another? Note that all these disciplines are already interdisciplinary, and if we propose that one or all of them contribute, we have an interdisciplinary methodology for investigating interdisciplinarity, which poses an issue of epistemic circularity.

A definition of interdisciplinarity which reflects all its essential aspects has not yet been provided. In the attempts toward a definition, key constituent concepts of interdisciplinarity have been assigned names reflecting in ordinary language concepts of an overly broad generality. These attempts offer good dictionary

definitions, but that generality prevents the use of such a definition within a theoretical framework. Brewer's (1999, 328) definition is a relevant example in this respect:

> Interdisciplinarity generally refers to the *appropriate combination of knowledge* from many *different specialties* – especially as a *means* to shed *new* light on an actual *problem*.[2]

The aspect of complexity, methodological effort and inquiry, motivation of the unsolved problems, and novelty of the results are reflected (through the italicized words above) in this dictionary definition, and other definitions within literature are usually more or less descriptive versions of this one.

What is clear – and all researchers agree – is that interdisciplinarity is an *epistemic* phenomenon. Since epistemology provides a kind of generality and primary foundation transcending typologies and boundaries specific to the concept of a particular discipline, this is where we should start in pursuing a theory of interdisciplinarity: the epistemic nature of interdisciplinarity.

In this paper, I argue that a third (neglected) aspect of interdisciplinarity is essential for clarifying its nature and for accounting further for the problems and challenges interdisciplinary does pose. This aspect is the epistemic outcome of the ongoing process of practicing interdisciplinarity, a theoretical entity of a structural nature developed through interdisciplinarity, which is grounded on the content of the disciplines as bodies of knowledge. This concept stems from the epistemic motivations of interdisciplinarity, among which the *theoretical* advancement of a discipline should also be included.

By analyzing these motivations in relation with the nature of the structures developed through interdisciplinarity, we can argue that the challenges interdisciplinarity is claimed to pose are of *the same* nature as the more or less special challenges scientists encounter within the evolution of their own disciplines. Hence, presenting interdisciplinarity as a special virtue of scientific practice is somehow exacerbated to the extent that that virtue is "as special as" the virtue of dealing successfully with the complexity of the disciplines themselves. My arguments run within a theoretical framework specific to the philosophy of science as well as the classical structural view on scientific theories.

The main aim of this paper is to clarify the nature of interdisciplinarity, by revealing its theoretical side in close relation to its motivations and dynamics. For this clarification to be possible, a primary conceptual clarification is needed and will be pursued.

[2] My emphasis on the words.

In the first section, I will try to clarify the criteria of specificity that we should follow in characterizing disciplines, because interdisciplinarity and its investigation have been seen as tightly dependent upon disciplinarity. Then I will argue for both weakening some distinctions and sharpening others for the purpose of this paper. In the second section, I will argue that interdisciplinarity does not mean or imply non-disciplinarity. I propose a structural scheme of how interdisciplinary processes go, focusing on the concepts of availability of the methods, concept linking, and theoretical modeling, and I conclude that the concepts of the disciplines in process of interdisciplinary interaction stand in certain *kinds* of relations, of a logical-epistemic-cognitive nature. These relations on one hand render interdisciplinarity amenable to a theoretical account, and on the other hand, the properties of these relations account for the dynamics of the phenomenon of interdisciplinarity. In the last section I draw conclusions.

1. Discipline – Defining, Delimiting, or Just Distinguishing?

In the study of interdisciplinarity, the focus has been on the essential concept of 'discipline'. With the attempts to provide a complete definition for 'discipline' and the adequate placement of the concept into the framework of investigations, various main aspects of the concept were stressed: social, historical, semantic, epistemic, and scientific. Within these main aspects, the typology is enlarged by every facet of such aspects: educational, expertise, authoritative (within the social aspect), distinction, reference (within the semantic), epistemic content, the truths, the normative, organization, regimentation and institutionalizing of knowledge (within the epistemic), resistance and evolution (within the historical), epistemic virtues and methodology (within the scientific aspect). The mutual dependence of these unfolded aspects makes difficult any attempt toward a descriptive definition by which to ensure a non-contextual stable use of the term, consistency with the disciplinary social practice, and eventually embedding in a conceptual framework within which to investigate interdisciplinarity theoretically.

1.1. Defining

The difficulty of stating an adequate definition comes first with the attempt to separate its use in ordinary discourse (including educational contexts) from that in a theoretical-scientific context. Regarding the former type of discourse, we have gotten used to the term 'discipline' to distinguish between and delimit what we were taught in primary through secondary schools and have associated the discipline with its teacher. Now, searching for a "meta-disciplinary" definition seems to be influenced by this habit, not to mention the Latin etymology of the word, based on

the root *discere* (to learn): *disciplinis* (disciplines) and the derived *disciplulus* (disciples); per this origin of the term, a *discipline* is a matter of knowledge or skills that can be taught by a teacher to his or her disciples (Alvagonzáles 2011). The latter type of discourse poses a particular kind of problem: Observe that sciences (or disciplines) dealing with (or at least referring to) the concept of discipline span a broad range, from philosophy of science and epistemology to social sciences (educational, policy, psychology, management, and so on), and thus it is difficult for a definition to fit all the conceptual and theoretical frameworks of these sciences or disciplines. For instance, the theoretical-philosophy disciplines are more concerned with the discipline's body of knowledge as a theoretical unit than with its social dimension, while social sciences are more inclined toward its aspects of education, interpersonal relationships, and social implications. The problem arises when trying to define interdisciplinarity through disciplinarity – since *all* the disciplines should be related to the concept of interdisciplinarity, *the same* definition should be employed in a metatheoretical/metascientific investigation of interdisciplinarity.

Definitions or descriptions for the concept of *discipline* that have been provided represent various ways of combining and weighting the aspects listed at the beginning of this section. Some researchers emphasize the educational-social dimension while others focus on the epistemological-theoretical one. This is not the place to review them or make a classification of these definitions.[3] I will just note that the previously mentioned double tendency reflects two distinguishable views on the concept of disciplinarity: one takes the discipline to be a body of knowledge with a certain epistemic autonomy and authority (the former delimits it and the latter makes it teachable), and thus the *content* of the discipline is taken to be more relevant for a definition than its relations to the society – call it the *internalist* view; the other sees the discipline as both its content and *social relations* (including academic, educational, and organizational-institutional) under the principle that content is in fact dependent on these social relations – call it the *externalist* view.[4]

The externalist view seems to be dominant, and this is explicable through the prevalence of the social sciences among the disciplines dealing with interdisciplinarity. The externalist views range from admitting the content as characterizing disciplines to radical ones, giving exclusivity to the social dimension; for instance, Apostel (1972), for whom sciences and disciplines do not exist, but only persons or groups practicing them. The structuralist approach of the content and dynamics of science, originating in the conceptual and methodological framework

[3] A well-organized review can be found in Chettiparamb 2007, 2-5.

[4] The terms for this typology and the typology itself are not original; they have been used in the literature, for instance in (Klein 2002).

initiated by J. D. Sneed and W. Stegmüller in the 1970s allowed the access of analytical philosophy into the scientific domains, as well as interdisciplinary communication between philosophers, scientists, and historians of science (Kuhn 1977, 289-291). In the view of T. Kuhn (1970, 182-187), the disciplines, which are seen as related to the concept of paradigm, are characterized by four elements – symbolic generalizations, models, values, and exemplars. For Kuhn, these structural elements shape the scientific communities and define problems and solutions, and not vice versa. In what follows I will adopt an internalist view.

1.2. Delimiting

Now it is time to adopt a simplifying position regarding definitions. The main reason we struggled to find an adequate definition for 'discipline' is that in order to investigate interdisciplinarity, we have to engage the concept of 'disciplinarity,' tightly related to *inter*disciplinarity. Disciplinarity is a *general* practice or principle of practice, regardless of the particular discipline to which it may refer in a particular context. Despite the various dimensions 'disciplinarity' may have, it means and should be used in the current context as "staying/remaining within the boundaries" of a discipline in the course of a scientific endeavor, having as a goal either problem solving or theoretical advance. The related concepts 'staying within' and 'boundaries' can make sense only as relating to the content, structure, and/or internal "affairs" of a discipline (including methodology and all its epistemic production and values) organized or structured so as to allow "the bounding" in some sense. This conceptual dependence imposes an internalist view for the concept of discipline in the context of interdisciplinarity (while not rejecting the externalist one). It is the *content* of a discipline that makes us able to 'stay within' it as investigators and not its social extension, even though that content is humanly produced, managed, and bounded.

The concept of boundary (of a discipline) may have a double sense: one related to the (bounded) content and the other semantic and related to the criteria of naming disciplines. Indeed, one can refer to boundary as the criterion of distinguishing between disciplines in order to have a consistent usage of their names in various discourses.

However, these criteria should not be merely conventional; naming a discipline should be as rigorous as the constitution of the discipline is. For the former sense, the concept of boundary does pose a serious problem: A discipline is not a static construct, but rather, has a certain dynamic concerning both its content and social implication. That dynamic is time dependent and human dependent. As such, how should we understand 'boundary,' so that 'staying within' it will make sense

despite these dependencies? How is this problem related to interdisciplinarity? We shall return to this point later.

1.3. Distinguishing

In the attempts to define 'discipline,' the goal of distinguishing objectively between disciplines has been followed, and this goal again has put forward the internalist view. The distinction aspect has led to structural definitions regarding the internal content of a discipline. Almost all such definitions have taken its objects of investigation and its methodology to be the two main constituents of the content of a discipline. For the former, their nature and ontology is expressed differently across the definitions; for instance, Boisot (1972, 90) takes "observable and/or formalized objects," but also "phenomena that are materialization of the interaction between these objects" to be the objects of study of a discipline. Heckhausen (1972, 83-84) distinguishes between the "material field" as "a set of objects" and the "subject matter" as "the point of view from which a discipline looks upon the material field." Squires (1992, 202) adopts a simple semantic stance by naming 'object' "what they are about." The latter element – methodology – expressed through "laws" (Boisot 1972, 90), "operations" (Squires 1992, 202) or "methods"/"analytic tools" (Heckhausen 1972 and many others) is what actually constitutes the structure of the content, as operating upon, linking, and making connections between the objects of study and the various concepts, theories, and truths of a discipline. I may add 'procedures,' 'techniques,' 'principles,' 'rules,' 'norms' and so on and note that methods have their own degree of rationality, generality, or epistemic authority. They may range from primary functions of human reason (like perception, association, or primary induction) to methods that are discipline-specific. Regarding this latter type, Bauer (1990, 106) transforms it into a second-order epistemic criterion of distinction adding to the content distinction through objects and methodology, by noting that

> Disciplines differ in epistemology, in what is viewed as knowledge, and in opinion over what sort of knowledge is possible. They differ over what is interesting and what is valuable.

Methodology and its mode of validation and evaluation is also what distinguishes scientific from non-scientific disciplines. This distinction based on methodology has been stressed by several authors who investigated interdisciplinarity; however, for our structural approach under the internalist view, the distinction is of little relevance, as we shall see further. The distinction is however of high relevance for the externalist views, especially in what concerns the historical and educational aspects of a discipline.

1.3.1. Distinguishing Through Concepts

We shall keep the two main constituents of the content of a discipline for the purpose of distinction – the objects of investigation and methods – in their broad sense; that is, the objects irrespective of a specific ontology or ontological commitment, and the methods with no epistemic distinction, hierarchy, or categorical taxonomy. At this point, let us note that distinction based on these two constituents does not work everywhere. There are distinct disciplines sharing both their objects of investigation and their methodology. To take an immediate example, consider astronomy, (physical) cosmology, and astrophysics. Their common objects of investigation are celestial bodies and phenomena, and they all use the same observational methods, as well as theoretical methods of physics (from various branches) and mathematics. What then essentially differentiates them? (Of course there are the historical criteria of delimitation, the amount or degree in which a specific method is used, and the placement of the methods in one branch or another of mathematics or physics.) The answer is: the stance towards their object of investigation, the primary concepts they developed about these objects, and the primary conceptual framework developed around them. Astronomy considers its objects of interest in their origin and evolution, being more focused on local celestial structures and systems, and interested in typologies, patterns of regularity, and the prediction of the dynamics of such systems. Astrophysics is more concerned with the physical nature of the celestial bodies and less with their positions and motion, while cosmology treats the objects within the largest-scale structures and dynamics of the Universe as a whole, being also concerned with fundamental questions about its origin, nature, structure, evolution, and fate. Other examples can be easily found in well delimited sub-disciplines of the same big discipline (for instance, atmospheric physics and meteorology within Earth science, or cognitive psychology and neuropsychology within psychology). Of course, the relevance of such examples is sensitive to various contextual factors, including historical and semantic; however, the conceptual aspect of the objects of investigation is related to methods not only in what concerns distinction between disciplines, but also in the effective way in which methods of a discipline do work. Indeed, the rational methods of investigation are not applied to any object in its material nature, nor as an abstract linguistic entity as referent, but to *concepts about* that object, even if this reverts in some trivial instances to the mere sense perception of the objects. Even an empirical method of research (say, collecting responses to a survey over a population sample, or participant observation in a community within some social sciences) is not applied directly to the material world, but to what our mind shapes through concepts as an image of it (through data in previous examples); this is also true for empirical,

experimental, or engineering disciplines. This view, in which we can only infer from what is thought traces back to Frege (1951/1892) and is based on the predicative nature of concept, which is fundamental for traditional logic and for some primary notions of mathematics (such as that of set, set membership, infinite sets, etc.). We should adopt it just because the content of a discipline (even non-scientific) as an epistemic unit has to submit to traditional logic.

Conceptualizing the objects of investigation before employing them in judgments, inferences, rules, and theories within a discipline means both adopting a view – however radical – on the inner nature of these objects and placing them within larger systems or structures of knowledge. The conceptual stance may differ dramatically for the same object of investigation across disciplines. For example, empirical concepts embeddable in a certain physical structure or having certain properties, describable in ordinary language of an empirical discipline, can be abstract mathematical concepts – identified through mathematical definitions – for applied mathematics or for mathematical physics. If they were not mathematical concepts, mathematical methods could not be applied to them. A population for a psychological discipline is a group of people with different psychological profiles, while for mathematical (statistical) psychology, it is a set of elementary events belonging to a probability field. If this were not so, statistical research or inference would not be possible within psychology. Therefore: 1. The nature and properties of the concepts are essential for a method to *be able to* be applied to them, just as a (mathematical) function can make sense only for its domain of definition. 2. Across disciplines, it is precisely the *difference* between those concepts about the same objects that makes interdisciplinarity possible, for if the concepts were similar, the methods of the same discipline would be operating on them (now or in the future). In the next section, we shall also see how concepts of different natures are linked across disciplines from a structural point of view.

Summing up as motivations the additional criterion of distinction between disciplines and the constitutive fundamental role conceptualizing does have for the methodology of a discipline, I will take *the concepts* to be the third essential element of the content of a discipline. These include the primary concepts about the objects of investigation, but also other concepts developed through theoretical advancement; along with the relations between them, they all form the conceptual framework of that discipline. Concepts have already been considered as part of an internalist definition of a discipline. In the previous examples, they are mentioned in the definition of Squires (1972, 202) as "their [disciplines'] stance toward that object [of investigation]"[5] and Heckhausen (1972, 84) as "the point of view from

[5] My insertions in brackets.

which a discipline looks upon the material field." What I want to emphasize is the epistemic relationships concepts do have with the objects of investigation, with each other, and with the methods operating on them: Concepts are about the objects of investigation – this is a semantic relation, but not merely conventional; the correspondence is made on the rational criteria of placing the concept into existing frameworks of the universe of knowledge about the respective topic, while not breaking any well-established systematic coherence with this correspondence.[6]

1.3.1.1. The Relationships between Concepts

Concepts stand in various types of relations to each other in the course of defining, making judgments, inferences, applications, and statements. As with the methods, concepts stand on one hand in an identity relation (a method would not be what it is without those concepts to or upon which it is applied), and on the other hand, it is a means of linking concepts epistemically. For instance, the method of basic induction links a data set associated with the evolution of a phenomenon to a predicted event. Two or more concepts linked through a rational method belong to the same epistemic structure. Note that the method linking two or more concepts is not only the "engine" creating that connection, but it also gives the *kind* of connection those concepts share, so that a relation can be defined (and not only conventionally denoted) as the class of the connections of a certain kind. In our previous example, basic induction put the concepts from the observational base and the prediction in a sort of inferential connection. A similar kind of connection is created between a set of hypotheses and the conclusion obtained by using the basic principles of logical inference or the usual deductive methods of science. Including a concept in a new definition creates a constitutive (or identity, if you prefer) connection between the newly defined concept and the one(s) included, while concepts linked through the application of a theory into another or in a specific context stand in an applicative kind of connection. The connection between the particulars and their corresponding generals is another kind, and so on. I will come back to the point of epistemic kinds of connection later.

1.4. The Nature and Epistemic Status of the Methods

Finally, for the methods, let us note that placing them within the content of a discipline does pose a sensible problem. Even though disciplines share methods as well as concepts, and even though there exist discipline-specific methods among

[6] Of course, such breaks occur sometimes, culminating with scientific revolutions. Within a chronology of the processes of constitution of a discipline, first correspondences always exist.

which are methods developed internally by a discipline, a method remains something that is *used*. This usage has linking concepts as its immediate goal, and it must be epistemically *justified* if we talk about a rational practice. However, that justification about usage cannot be attained within the discipline itself using the method, but beyond it and beyond any other, within a general epistemology of the rational methods. This is true not only for the primary methods available for *any* discipline (such as abstraction from experience, logical principles, induction, theoretical modeling, and so on), but also for discipline-specific ones. For instance, the method of statistical inference used in social sciences is not epistemically justified within those sciences, but at another level of reasoning. First, the mathematical nature of the concept brings a justification based on the epistemic value of the mathematical-logical necessity that we all accept regardless of our profession. Second, the empirical confirmation and the good track record of the inferred results are not concepts originally specific to the science using that method, but of all sciences using it, as well as to our common rationale in daily life. Even if we accept Bauer's (1990, 106) position on validating methods as being discipline-specific, I claim that that internal validation is just decisional – thus normative – and reduced to choose and select; the criteria of these two actions may be discipline-specific and may be justified for the objectives of a discipline, but the core epistemic validity of the methods lies in the *availability* of the method, which is beyond the discipline (say, transdisciplinary) and dictated directly from a level of discipline-free rationality. In other words, the distinct epistemology of each discipline in Bauer's terms (at least in what concerns methods) is subject to a second-order primary epistemology of reasoning.

From this perspective, it seems that the nature of *any* method is intrinsically transdisciplinary, in the current usage of the term.[7] On the other hand, taking the method as belonging to the epistemic content of a discipline is fairly justified: the method is the means of creating connections, which account for the internal systematic coherence of the content and the intelligibility of the system, but the method also accounts for the kinds of connections it makes between concepts, and this latter contributes to the *understanding* of that connection and also accounts for the *relevance* of a connection to be included in the epistemic content of a discipline.

Now, if we mentioned semantical and syntactical relations, definitions, statements, and theories as related to the content of a discipline, and – leaving for a moment the internalist view – taking into account that the content of a discipline

[7] This conclusion is in the vein of Lakatos's (1968) view on methodological unity of sciences, and opposed to the views of Cartwright (1999), Galison (1996) or Shapere (1984), who argue for a methodological differentiation depending on the nature of the disciplines.

should be communicated and taught, shouldn't we introduce *language* as the fourth element of content? Language contributes to the specificity of a discipline, especially in what concerns vocabulary. It is very likely that there are not disciplines sharing the same objects of investigation, concepts, and methods, and so perhaps language is not necessary as a new criterion of distinction. However, language certainly is present in the content. It suffices to mention the linguistic nature of conceptualizing, of definition, of predicative logic, and of scientific theories to see that language serves as both a tool and a method for creating and linking concepts (not to mention its role in communication). The options are either to include it in methodology as a primary method (function of the brain) or to take it as the fourth distinct element of content. The observation that methods are also expressible in language, just as concepts are, inclines the balance toward the former option. The observation that all the development and results of a discipline can be expressed in a body of statements formed in a language (what is to be communicated and taught in a curricular way) inclines it toward the latter option. Also bear in mind that our aim here is to investigate interdisciplinarity through disciplinarity and – for the sake of simplicity in such a complex context – we must employ only those concepts directly related to the two we mentioned, namely "boundary" and "staying within (the boundaries)." It seems that language does not submit to this attribute, except in the semantic criteria of naming disciplines, which we proposed to avoid. Indeed, language (either ordinary or scientific) – by its nature – is cross-bordering, flexible, and with a high degree of freedom in its semantics and domain of description. It suffices to mention mathematical language, which has exported predications and vocabulary into the ordinary language and into the languages of several disciplines. As such, the linguistic element is one that facilitates cross-bordering, and not bounding. This process goes in fact in a double sense, being so specific to interdisciplinarity: once language accesses the new discipline, it feeds back the language of the original discipline and so a new language is formed.[8] Given all these considerations and that we are talking here about the *epistemic* content of a discipline, not only the linguistic one, I will take language as an element of the content associated to all three previously included, but not independent or distinctive in its own right.

Summing up, in my internalist view, the content of a discipline would consist of these three components:

[8] The concept of interlanguage through interdisciplinarity and transdisciplinarity traces back to the works of Piaget (1972) and Lichnerowicz (1972), as well as to early structural approaches of unity of science.

- the objects/subjects that discipline investigates
- the conceptual framework that discipline uses in its investigation of the objects/subjects, development, or any of its objectives
- the methods which that discipline uses for advancement and accomplishing its tasks and objectives.

With this primary organization of the epistemic content of a discipline – not presuming it as exhaustive nor standing for a rigorous definition – we can take the next step in analyzing interdisciplinarity structurally and epistemically within a conceptual framework developed around the concepts of 'boundary' and 'staying within (the boundaries).'

2. Non-disciplinarity and Interdisciplinarity

In this section, I will argue that, at least under an internalist view, interdisciplinarity, transdisciplinarity, cross-disciplinarity, or multidisciplinarity does not mean or imply non-disciplinarity.

If disciplinarity is understood as relative to the concept of (a kind of) boundary for the epistemic content of the disciplines, then all these prefixed versions of '-disciplinarity' should mean crossing or passing through or breaking or even violating these boundaries. An immediate reflection on this supposed epistemic action would observe a. that those boundaries (if any) must be by their nature trespassable; b. that trespassing is justified and allowed at a certain level of meta-methodology; c. that trespassing (which is not quite justified as an adequate term given b.) occurs by means of one or more epistemic methods the disciplines usually use in their practice; d. the methods with which the boundary is crossed should belong[9] to the epistemic content of the discipline within which the epistemic action was initiated (where the problem was posed or the theoretical goal was designed, etc.).[10]

Now it is time to take a new action toward simplifying all that we described as the epistemic content of a discipline. As a necessary parenthesis to clarify the abstract concepts introduced when discussing methods in the previous section, I will describe in structural terms what the epistemic structure of a discipline would mean in order to represent boundary crossing.

[9] In the sense of "used" (see the view on methods in the previous section).

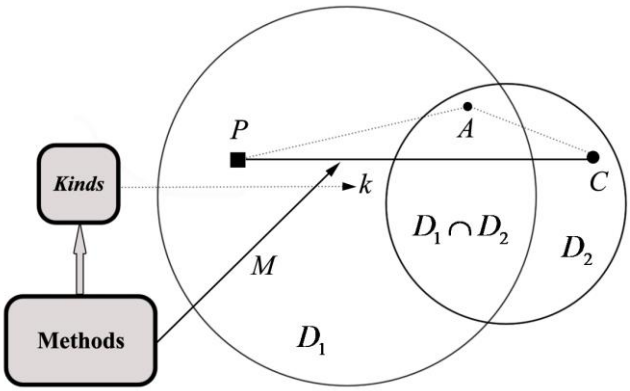
[10] The apparently distinct case in which a method is *imported* from another discipline for a particular endeavour can be cancelled in the view that once used, a method has linked concepts belonging now to the epistemic content of that discipline, and the method should belong as well, even if no longer used in the future (the argument has been detailed in the previous section).

2.1. Disciplinary Epistemic Structures, Dynamics and Bounding

In our framework, the crossing of the boundary works by the following simplified scheme:

Let D_1 be the original discipline within which a problem P (either practical or theoretical) is posed and addressed. P is formulated in the conceptual framework of D_1. We can see P as a concept or a group of concepts related to each other. For the sake of simplification, we take the former case, since a structure of concepts is itself a well identified concept. As an initial case, consider the aim to connect P to a concept C not belonging to D_1, but to a different discipline D_2 (the same simplification applies to C) in a way that will ultimately enrich our knowledge of P expressed in D_1. The general case is to connect more than two concepts, but for simplification we shall consider only binary relations, as they suffice to reflect our point on disciplinarity. Assume we found a method M that made that connection. The aim has been attained and the whole action is called interdisciplinary because a boundary of D_1 was crossed, at least in what concerns concepts.

1a. The inquiry stage



1b. The integration stage

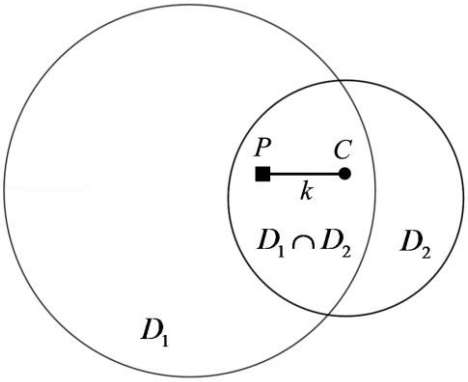


Figure 1. Linking concepts between different disciplines

The method M could be either one already used by D_1, developed within D_1 just for the current purpose, or "imported" from another discipline. Given my transdisciplinary view on the nature of methods expressed in the previous section, I will not in this context take the latter case as one of interdisciplinarity. Once used, a method is validated for that use (for the concepts it links), and it can be considered as belonging to D_1 in the sense that at any time it can be used again where it is adequate. Now let us take an insight into the nature of these actions and their outcome.

First, we should accept that concepts P and C were already somehow *connected* before the method M was applied. This connection may be made visible if we break the concepts down to constitutive concepts until reaching their primary concepts of constitution, among which one (A in the figure) or more is common to the two disciplines. Any concepts – regardless their nature, category, or regimentation – are connected to each other in the *network* of knowledge, as neither concept is developed *ex nihilo*, but from existing concepts; this is clear for the mathematical concepts and happens in any area of knowledge. As such, any two disciplines share common concepts (not necessarily about their objects of

investigation), so the intersection $D_1 \cap D_2$ is never empty.[11] However, having P connected with A and A with C does not cancel the endeavor of linking P to C, because those primary connections may not be relevant for our purpose – that is, we might not gain any new knowledge about P. For instance, if A is a concept constitutive to both P and C, we can say nothing about how P and C stand in a relation or in what kind of relation they do stand. Or, if A implies P and A implies C, we cannot say that C or P implies the other, since this does not reflect the transitivity of the implication. In other words, we need a connection between P and C that is epistemically *relevant* for the objective we propose. This is where the distinction between network of knowledge and structure of knowledge is essential. P and C are connected in the former; however, a connection between them in the latter is acquired only when it has the status of (belongs to) a *relation*. But what constitutes that relevance? In mathematical terms, a network is a graph connecting atomic nodes through paths, while a structure is a set of atomic nodes (objects/positions) together with a family of relations between them, that is, a set of sets of connections (n-uples of nodes of various arities). A relation represents a class of connections, and this class can be constituted either conventionally or by criteria of relevance of the representation that that structure does. As such, even though the nodes are atomic and lack any content or epistemology in the abstract structure, grouping the connections between them as relations grants the nodes a minimal epistemology given by the criteria of grouping. However, this epistemology is the one of that particular structure and particularity comes from that grouping; therefore, the conventional nature of the relations remains. In mathematics, any structure (in the sense of Bourbaki (1950)) is reducible to a set-theoretic structure, but any relation (as a class of connections) is *defined* mathematically. This does not happen for other non-mathematical representations of various systems such as classical structures, where the relational arrangement is conventional (that is, we just name or interpret the relations, but not define them). However, for a structure representing knowledge through linked concepts, we can acquire an epistemology of the structure from *outside* it, by assigning the connections what I call epistemic *kinds*. For instance, consequential (inferential), constitutive, applicative, predicative (characterizing or assigning properties), and negation are kinds of connections between concepts, constituting any rational system of knowledge. Their nature is both logical and cognitive, and they represent the rational processes of thought. If two or more concepts connected in the structure exhibit a connection of a certain

[11] I am using abusively the set-theoretic notation, just for illustration. If we consider the epistemic side of the content of a discipline (including methods), the notion of intersection is different from that applying to sets of objects.

epistemic kind, the relation holding that connection (and other connections of the same arity and type) acquires the same epistemic kind. This is what we require as relevant for the aimed connection between P and C in our context: an epistemic kind k, given by method M. Let us call a classical structure with connections having assigned epistemic kinds an *epistemic structure*.

Such a structure is usually a second-order structure because relations between relations may also hold. Think of *properties* of concepts, which are both second-order concepts and relations between first-order concepts. For example, in mathematics, the property of 'being divisible to' connecting integers is related to the property of primeness of a number, and the connection is of a constitutive kind, since primeness is defined through divisibility.

Not only do epistemic kinds ensure the relevance of the relations between concepts within the interdisciplinary action in context, and not only do they ensure understanding and logical coherence of the constructed systems of knowledge, but that relevance is included in the criteria for delimiting the content of a discipline by external authorities. As such, it is clear that the "boundable" content cannot consist only of concepts, but of the entire structure linking them through connections of certain epistemic kinds.

Now let us take an epistemologist-observer position for what we have just represented structurally as discipline boundary crossing. Once the (P, C) connection is realized with epistemic kind k, concepts P and C along with their connection come to belong to the same epistemic structure. This structural extending "moves" both concepts P and C into the common zone $D_1 \cap D_2$, since the connection is epistemically relevant. However, P was assumed to belong exclusively to the original discipline D_1, for if it also belonged to D_2 we wouldn't have any interdisciplinary action. One may argue that there are two stages of this process separated on a timeline – one of inquiry and application of method M, and one of integration – and only in the latter is the new connection assimilated into the existing structure; as such, in the first stage – the one of interdisciplinarity – the two concepts remain in separated domains (the two stages are pictured in Figure 1a and 1b along with all the denotations used above). My counterargument invokes the same concept of epistemic kind, which is not disciplinary. The relevance of the connection (P, C) is – at the moment of obtaining – acknowledged through k, which is not specific to either of the two disciplines or their boundaries or intersection. For the investigator, once the connection is obtained, the epistemic structure of D_1 is extended regardless of the conventional (and perhaps relevant) disciplinary bounding of that moment. Since it is the investigator alone (or team of investigators) that actually perform(s)

the interdisciplinary action and only their results are validated externally, we have an instance of interdisciplinarity (as crossing content boundaries) with no actual crossing (staying within the epistemic structure of the original discipline).

Indeed, the stage of integration does exist and is socially driven. The integration will apply not only to the two concepts linked, but will engage all the concepts with relevant connections to those two from both disciplines; therefore $D_1 \cap D_2$ extends itself with more than one connection and two concepts. The structural integration is selfaccelerated, and this represents the process of merging disciplines. It is the classic case of physics and biology; biology and chemistry; sociology and anthropology and many other pairs. Domains of disciplines are also in process of merging through structural concept linking. This is how sociobiology, quantum information science, cognitive neuroscience, and other relatively new disciplines were established.

Crossing discipline boundary (if a valid notion) is not done only through accessing concepts from another discipline. There is also the case when the aim is to connect P with a concept C also from D_1, through a method by which we access content from D_2. This case is illustrated in Figure 2 and described in structural terms as what we traditionally mean by theoretical modeling: We identify/observe a structure of concepts from D_2 that is homo- or isomorphic with a structure of concepts from D_1 that includes P and C (call S_2 and S_1 the two structures). This correspondence is not made arbitrarily or conventionally (although it may be made as such), but by following criteria of relevance as well as convenience. Among the former, the correspondence should preserve in both domains the systematic coherence of other structures involving concepts from those placed in correspondence. The validity of the model is tested against this coherence, and it might at any time be invalidated when advancing outside in the superstructure, including through empirical observation. Assume the corresponding concepts of P and C in D_2 are A and B [$f(P) = A$, $f(C) = D$, where f is the structural morphism]. By using the system of knowledge or theory from D_2 we infer or observe that A and B do stand in a certain relation in S_2 and – by means of the homo/ isomorphic feature of f – we infer that P and C do stand in a certain relation in S_1. Note that this inference is actually an *interpretation* back in S_1 of the known relation between A and B, that is it is described in the terms of D_1. Thus, we obtained the connection (P, C) through a double inference: one is within S_2, and the other between or across

S_2 and S_1, via *f* and its properties.[12] We may now fairly assume that if connection (*A*, *B*) is of the epistemic kind *k*, the inferred connection (*P*, *A*) is of the same kind; otherwise the systematic coherence would be broken in D_2.

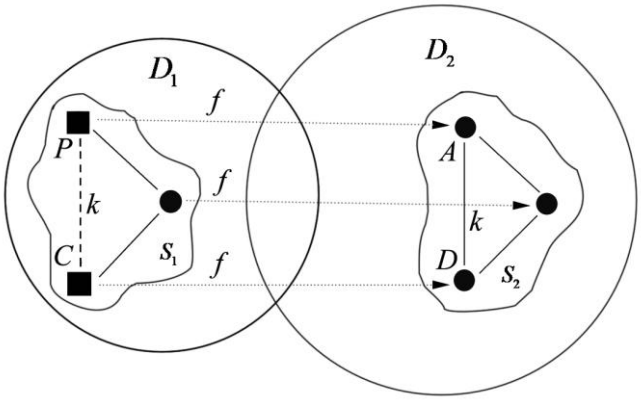


Figure 2. Linking concepts through theoretical modeling

Now, take again the observer position. Where or how was the boundary of D_1 crossed? It seems to have occurred nowhere. Concepts from D_2 were not accessed through *f*, as in the previous case, but put in a formal correspondence with those from D_1, then interpreted back in D_1 along with their connection. It is like a semantic relation – we can talk about something without seeing or touching it. No relation has been established between the concepts of the two disciplines. What we brought in D_1 is a truth of D_2, which we adapted to D_1 so as to have an epistemic kind for the connection (*P*, *C*). The method *M* through which we did this (theoretical modeling) is not discipline-specific, but universal. Is this import of truth through *M* a boundary crossing? The content of D_1 was not enriched with new concepts, nor with new relations, but with a new connection (as the imported truth interpreted) for an existing relation. Given that the inferred connection is relevant for D_1, the epistemic content of this discipline is actually the same as before the

[12] In the traditional terms of structural-theoretical modeling, the former inference is said to take place within the 'governing theory' of the model, while the latter is the interpretation step (when a new function is applied from one structure to another, not necessarily the inverse of *f*).

interdisciplinary action. Then, the only option remaining is to qualify the truth import as boundary crossing from D_2 to D_1. However, this import is actually the method we used in the action and does not belong to any discipline. Theoretical modeling, which is at its core a primary cognitive method, namely that of reasoning through analogy, just sheds light on a structural equivalence that we uses to advance knowledge in the target (original) discipline. But *observation* is in turn a primary cognitive method and cannot be granted any boundary crossing – we can observe from distance. Think also of metaphors as an example, which in language have the role of easing understanding or describing through syntactical analogy. Metaphors are based on the same structural morphism as theoretical models.[13] A metaphor with fictional subjects will never alter or cross into the non-fictional reality. If this is still seen as bringing something (new) to something from something separated, the entity that is brought seems not to cross any boundary of any sort. Overall, as in case 1, we have boundary crossing with no crossing for the epistemic content.

The universal method of theoretical modeling used in case number two is one of a special power. It can link structures from various domains and concepts of different natures and do transfer of truths from one domain to another. However, its results are subject to further confirmation. At some point in the evolution of the system of knowledge of the target domain, one result may be invalidated by observation, experience, or theoretical inconsistency with other substructures. If such a problem occurs, it is not the method which should be invalidated, but the model developed for its application – that is, the structures, their relations, and their interpretations. Regarding the latter, language of the target discipline plays again a major role here, being actually part of the model. This is why there is always a fine-grained competition of models in the advancement of rational disciplines.

Models are used not only between disciplines, but also between domains or theories within one discipline. The paradigmatic example is that of pure mathematics, for which the set-theoretic foundation allows various theories to be linked through structural morphisms and mathematical structures. Physics also uses its own models to advance. For instance, the model of colliding balls from Newtonian mechanics is used in the kinetic theory of gases (interpreting gas molecules as elastic balls); the gas model was used in applications in nanotechnology, energy research, and biology (Lizhang 2012); the gas model also found applications in recently developed theories on stars and galaxy formation (Binney & Tremaine

[13] Nonetheless, their validation against the external systematic coherence is more limited than in the case of theoretical models.

2008). Any discipline advances through theoretical modeling, by applying its established truths to new contexts if a structural similarity is found.

Theoretical modeling is what we usually call an *application* of one theory into a target domain – in fact, a non-trivial case of application, as an application mediated by a model. Thus, we can roughly say that D_2 was applied to D_1, and such application contributed content for D_1, but this is not content from D_2. The new content was a connection between concepts from D_1 "suggested" by the governing theory from D_2 and there isn't any integration stage as in case one in which the new content is assimilated in a common zone.

Traditional disciplines have evolved sufficiently through theoretical modeling to have "zones" called interdisciplinary, double-named and considered as a new academic discipline such as mathematical physics, mathematical biology, mathematical economics, mathematical neuroscience, and so on (having mathematics in the role of D_2); but also bioinformatics, bioeconomics, behavioral economics, socioeconomics, psychometrics, biopsychology, and so on. A discipline named in such prefixed forms 'D_2 (-"ized") D_1' refers to a "zone" of D_1 whose content (as statements, truths, and theories) is constituted by using models (and thus imported truths) from D_2. Whether teaching 'D_2 (-"ized") D_1' include also *how* those truths were reached in D_2 is an arbitrary choice; teachers must take seriously into account their expertise in D_2 in all its dimensions. In what concerns the epistemic content, I have argued above that the truths of the applied discipline are not imported as new content for the target discipline.

Mathematics is obviously the most frequently present discipline in such modeling processes, and this is not surprising at all if we think of its special status as a discipline. Specific features of mathematics and applied mathematics, like the nature of their structures, the mathematical language, the predicative logic and set-theoretic foundation account for its descriptive power and all the roles it plays in the constitution and advancement of the sciences, either natural or social. The descriptive power of mathematics is structural. The fact that the laws of physics are best formulated in mathematical language and mathematical descriptions of the empirical phenomena are possible (with a certain degree of idealization) within any scientific discipline, receives structural explanations. One widely accepted explanation is related to the richness in structures of pure mathematics compared with the structural needs of scientists (Maddy 2007, 341-343). This richness is self-generated through the structural apparatus of mathematics, where self-application

follows the same model based on structural morphisms. This structural richness and fertility is also externally influenced in a metabolic manner. Mathematics feeds from the problems of sciences, develops new theories just for addressing those problems, and puts their results at the disposal of the sciences, which access them through mathematical modeling. This perpetual process has now reached a stage in which the border between pure and applied mathematics is very diffuse. We have a mathematics of anything, however complex the thing may be, and as such we have a mathematics of any discipline dealing with that thing as its object of investigation. "Mathematics of" is simultaneously a model, a method, a description, a field of both pure and applied mathematics, and – sometimes – a new discipline, if named as such. We also have a physics of something, a biology of something, or sociology of something and it is just the evolution through models that makes us see an interdisciplinary zone around the investigated object, which we call "D_1 of D_2".

The method of mathematical modeling is not epistemically justified within mathematics and does not belong to mathematics; it is a particular case of theoretical (structural) modeling. In fact, the general applicability of mathematics has never received a satisfactory explanation at a metatheoretical level and thus raises philosophical questions. These questions developed around what in philosophy of science is called 'Wigner's puzzle:' *Why* is mathematics applicable in the sciences and physical reality, *how* do we rationally justify the use of mathematical models in the investigation of physical phenomena, and *how* do we explain their high rate of success, given the extreme difference in nature, ontology, epistemology, and logical category between the source (mathematical) and the target (physical-empirical) domains and also some special features of the mathematical practice?[14] For instance, some mathematical concepts and theories created and developed independently of any empirical contexts prove to apply to certain empirical contexts (Wigner 1960, 3, 7), perhaps after decades.

Among the criteria we follow in granting mathematics a special status among disciplines, one quite marginal might be that mathematics has never received a definition to reflect its nature and specificity, and this may account for our inquiry into defining or not defining disciplines. As a discipline, science, method, formal language, or whatever it is, mathematics has provided common conceptual frameworks for wide domains of most of the natural sciences, and methods of investigation and advancement of the disciplines; in this way, it has proved to be one of the main "engines" of what we call interdisciplinarity, by merging, unifying, and

[14] These questions developed around the influential paper of physicist Eugene Wigner (1960), the author of the syntagma "the unreasonable effectiveness of mathematics" and shaped a new domain within philosophy of science, namely philosophy of applicability of mathematics.

providing a common conceptual language. As a discipline, it has a content in which structures are auto-expandable, and the epistemic kinds of the connections are logical kinds. In addition, it seems that its applicative feature influences the development of this content. Given these, can we delimit such content – and how?

The epistemic-structural model is not quite a reductionist model for interdisciplinarity in the internalist view. Employing the concept of epistemic kinds in the constitution and qualification of the relations of the structure grants the set-theoretic structure a special epistemology which has both a logical and a cognitive dimension. Relating the methods to this structure as generators of the epistemic kinds increases the power of representation of the model, which is able to reflect the dynamics of the processes and is open to further refinement to reflect also the users' interventions in these processes. The nature of the main concepts employed (relations, connections, sets, kinds, structural analogy) is not discipline-specific,[15] but directly related to cognition, so our concern about epistemic circularity expressed in the introduction vanishes. There is no specific discipline dealing with this model, just our inner structural setup of reasoning.

Conclusions

Now is the time to take stock of what we have argued. We started from the common conception that we must study interdisciplinarity through disciplinarity, as being a kind of practice that combines knowledge from two separate disciplines, and this combination assumes crossing some boundaries of the combined disciplines; as such, disciplinarity is assumed to mean staying within those boundaries. Because interdisciplinary actions operate on the content of the disciplines and the concepts of 'boundaries' and 'staying within' (them) are relative to a (boundable) content (as a body of knowledge), it is adequate to have an internalist view on the discipline and consider the content in its epistemic-structural nature. In our structural framework, structures are seen as relations between concepts of certain epistemic kinds, and these kinds are assigned through the rational methods linking concepts which thus exhibit a transdisciplinary nature. Therefore, we have represented the interdisciplinary boundary crossing in two possible ways: linking concepts from different disciplines and linking concepts from the same discipline by using theoretical models sourced in another discipline. For both, I have argued that the

[15] The view that primary mathematical notions like sets and functions are not mathematical, but specific to our primary cognition, is still debatable within philosophy of mathematics. Relatively recent advancements in perceptual mathematics or protomathematics, among which is worth mentioning the works of Teissier (2005), Ye (2009), and Mujumdar & Singh (2016), may have important implications for this debate, favoring the view.

epistemic content of the original discipline is actually not passed, and as such we have interdisciplinarity with no boundary crossing. Interdisciplinarity (or any of the prefixed versions of '-disciplinarity') – in the common meaning – cannot be equated with and does not imply non-disciplinarity. The conclusion we ought to draw from this contradiction is that the two concepts we started from ('boundary' and 'staying within') were not adequately represented. What then are the options for changing that? How can we delimit a structurally represented content in a way other than *selecting* substructures? Choosing a type of closeness to certain operations instead of the mere set-theoretic delimitation does not work either, since the only possible operations are still over the concepts and represented as relations. If the methods themselves are the operators instead of assigning epistemic kinds – and thus make them discipline-specific in nature – the closeness through methods would assume that interdisciplinarity requires the development of brand *new* (crossing) methods different from discipline-specific ones, which is not the case in the real practice. Overall, it seems that any kind of set-theoretic- or topological-like bounding is inadequate for the epistemic content of a discipline. This is also in the vein of Nicolescu's (2014) view on boundaries. If this is the case, we have to draw a radical conclusion: From an epistemic-structural perspective, interdisciplinarity cannot coexist with disciplinarity, because delimiting is not possible.

The formalism of the model I have sketched can be integrated in both the framework of the general systems theory and Piaget's (1972) cognitive structuralist model of advancement of knowledge, where the key concepts are assimilation and adaptation. This is not the place to develop this unifying design, but it deserves further research. I will limit myself to mentioning that the cognitive origin of Piaget's concepts would benefit nowadays by the pioneering results of the mathematized neurosciences, so that the abstract nature of the structures of knowledge may find biological models. The epistemic structural model of (inter)disciplinarity supports this prediction if we consider the properties that epistemic kinds do exhibit at a first glance: all such relations are antisymmetrical, some of them are transitive, and structures are extended through the transitive ones. It would not be very surprising if these properties are found to be related to the neurophysiology of the brain, where the flow of electrical impulses over neural paths do have properties of sense and direction related to cognitive achievements.

Coming back to interdisciplinarity, the answer to the title question of section two is this: Finding a complete definition for the concept of 'discipline' is of no relevance for the investigation of interdisciplinarity in its common meaning, since disciplinarity and interdisciplinarity are not opposite each other, just as finding the complete definition of mathematics would never account for a metatheoretical

justification for its universal applicability nor change the way mathematical methods are applied in other disciplines. Delimiting a discipline in its content is possible set-theoretically in either its objects of investigation, its concepts, or its methods alone, but not all together; considering all three together with their natures and relationships requires a structural representation of the epistemic relevance and coherence of a cognitive system which cannot be bounded. Finally, distinguishing between various disciplines is possible, meaningful, and useful, as it ensures the semantic stability of the discourse on the topic, regiments the curricular side of the disciplines, and allows shortcuts for content access in discourse, education, and research. However, there is also a weak aspect of distinctiveness: Making sharp distinctions influences the conceptual stance we take for interdisciplinarity in many directions, for instance when we address the issue of problems and challenges of interdisciplinary practice.

In our epistemic structural framework, the two-case structural scheme of crossing boundaries is not only a counterexample for an inadequate conceptualization, but also a minimal structural representation of the actual interdisciplinary practice and advancement, through either direct concept linking or theoretical modeling. In this representation, there is no difference between how advancements are made within a certain structure named as a specific discipline for the purpose of distinction and reference, and across different such structures. The same kind of structural advancement is made in both situations, and the immediate goal of each action is to make relevant connections between concepts, including concepts of different natures, by assigning them the same[16] epistemic kinds everywhere. Viewing the methods of advancement as universal in the sense of epistemic availability and discipline-independent justification supports this equating.

Observe also that under this model, there is not much distinction between interdisciplinarity, cross-disciplinarity, multidisciplinarity, transdisciplinarity and other recently prefixed versions. This is not a weakness of any reductionist stance, but an effect of the combination of logical-epistemic-structural-cognitive components for the nature of the model.

In conclusion, the problems and challenges associated with interdisciplinary practice and so greatly stressed within the research of the phenomenon are the same *in nature* as those that investigators and experts encounter within their own discipline. These problems are not associated with crossing any boundary of content, but with the *complexity,* and the management of this complexity; it amounts to the potential breadth of knowledge in every domain paired with the anthropocentric

[16] In the sense of the finitude of those available or validated.

features of the investigators**:** possessing a driving curiosity, reaching proposed aims, and making new discoveries, all against the limited resources of reason and brain. Interdisciplinarity was *ab initio* associated with the concept of complexity, at least in what concerns education (Klein 2001; 2004), but a kind of complexity of combining knowledge rather than the complexity of the knowledge itself.

These challenges with respect to content complexity depend on historical contexts and evolution for every discipline. Take as two opposite examples Christian theology and mathematics. The former has not expanded much in content for the last century, as no influential new concepts have been developed or accessed (Pelikan 1989, vii-ix). The latter, mathematics, is the most fertile discipline, developing at an incredible rate new concepts and theories stemming from auto-application and difficult problems, including the problems of the sciences. There were brand new theories developed just for solving an "annoying" conjecture, which ultimately found other applications in other fields.[17]

Considering further the example of mathematics, imagine a mathematician graduating in this discipline 20-30 years ago and working now on a mathematical project requiring application of a mathematical theory developed recently. The problem or challenge this mathematician encounters is that of recreating the entire structural linkage between the new theory and the traditional concepts and theories taught in university. This is not an easy task and may be impossible in some circumstances. An option would be to collaborate with another more recently degreed expert in that theory, who is able to prepare relevant by-passes in the epistemic structure to ensure the understanding of the relations, saving time and other resources for the older colleague. Now imagine the same mathematician working on an applied-mathematics project in, say, medical imaging of the brain. In order to develop mathematical models and correspondences with the concepts of the target domain, the applied mathematician must become familiar (in some degree of reduction and convenience) with the anatomy of the brain and other medical concepts regarding medical images. This familiarity is required for the applied mathematician to make the relevant links with the mathematical concepts and theories. Is this challenge different from the former? Both assume effort, both have the option of collaborating, and in both, the mathematician struggles to connect concepts previously unconnected – of the same nature in the former case, of a

[17] Just to provide one illustrative example, think of Poincaré's conjecture, whose solution was provided 98 years after its statement in 1904. The solution belongs to the early history of algebraic topology. Generalizations of the conjecture to higher dimensions links to the concept of deformation in Riemannian geometry, with implications and applications for gravitation and cosmology.

different nature in the latter. The struggle is first with the complexity and not (or second) with interdisciplinarity.

The management of the complexity of knowledge has a strong social aspect despite the freedom of the investigators. Even if interdisciplinarity does not exist structurally, the phenomenon of interdisciplinarity does exist in this social aspect. It starts with the curricula conditioned by academic and governmental authorities and assignation of official expertise, and it continues with all interpersonal relationships developed around the interdisciplinary projects, including team management. In this social realm, boundaries are of a different nature and can be kept or crossed through different criteria. As such, we have interdisciplinarity through boundary crossing and a multitude of challenges that may be addressed by appropriate social sciences. Leaving the internalist view but retaining the conclusions from above, still accepting interdisciplinarity as a valid notion, maybe the right characterization for the concept of discipline would be the normative-authoritative "to get disciplined" (in the sense of Turner (2000, 47) or Parker (2002, 374), cited in Chettiparamb 2007). Another promising concept reflecting social boundaries is that of interdisciplinarity as a culture (Bauer, 1990).

The epistemic-structural model of disciplinary advancement, by reducing the processes to a limited number of kinds may suggest first reducing or collapsing the excessive typologies developed around this topic, and second, that a theory of interdisciplinarity can only be a *social*-science theory; other structural theories on this phenomenon may qualify only as general theories of knowledge. A third suggestion would be that within such theory, we must carefully discern between the challenges and problems of interdiciplinarity as formulated by the researchers, as some of them may be just semantic in nature and thus ingenuine.

References

Apostel, L. 1972. "Conceptual Tools for Interdisciplinarity: An Operational Approach." In *Interdisciplinarity: Problems of Teaching and Research in Universities,* 141-180. Paris: OECD.

Alvagonzáles, D. 2011. "Multidisciplinarity, Interdisciplinarity, Transdisciplinarity, and the Sciences." *International Studies in the Philosophy of Science* 25(4): 387-403.

Bangu, S. 2012. *The Applicability Of Mathematics In Science: Indispensability And Ontology*. Houndmills, UK: Palgrave Macmillan.

Bauer, H. H. 1990. "Barriers against interdisciplinarity: implications for studies of science, technology, and society." *Science, Technology, & Human Values* 15(1): 105-119.

Binney, J. and S. Tremaine. 2008. *Galactic Dynamics*, second edition. Princeton: Princeton University Press.

Boisot, M. 1972. "Discipline et interdisciplinarité." In *Interdisciplinarity: Problems of Teaching and Research in Universities*, 89-97. Paris: OECD.

Bourbaki, N. 1950. "The Architecture of Mathematics." *The American Mathematical Monthly* 57(4): 221-232.

Brewer, G. D.1999. "The challenges of interdisciplinarity." *Policy Sciences* 32: 327-337.

Cartwright, N. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Chettiparamb, A. 2007. "Interdisciplinarity: a literature review." *HEA Interdisciplinary Teaching and Learning Group, Centre for Languages, Linguistics and Area Studies, University of Southampton*.

Frege, G., P.T. Geach, and M. Black. 1951. "On concept and object." *Mind* 60(238): 168-180. Originally published as "Ueber Begriff und Gegenstand" in *Vierteljahresschrift für wissenschaftliche Philosophie* 16 (1892): 192-205.

Galison, P. 1996. "The context of disunity." In *The Disunity of Science: Boundaries, Contexts and Power,* edited by P. Galison and D. J. Stump, 1-33. Stanford, CA: Stanford University Press.

Heckhausen, H. 1972. "Discipline and Interdisciplinarity." In *Interdisciplinarity: Problems of Teaching and Research in Universities*, 83-90. Paris: OECD.

Klein, J. T. 1996. *Crossing Boundaries: Knowledge, Disciplinarities and Interdisciplinarities*. London: University of Virginia Press.

———. 2001. "Interdisciplinarity and the prospect of complexity: The tests of theory." *Issues in Integrative Studies* 19: 43-57.

———. 2002. "Unity of knowledge and transdisciplinarity; contexts of definition, theory and the new discourse of problem solving." In *Unity of knowledge in transdisciplinary research for sustainability*, Vol. 1, Encyclopedia of Life Support Systems, 35-69. Paris: Eolss Publishers Co.

———. 2004. "Interdisciplinarity and Complexity: An Evolving Relationship." *E:CO* 6(1-2): 2-10.

Kuhn, T.S. 1970/1962. *The Structure of Scientific Revolutions*. Chicago: Chicago University Press, second edition.

———.1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Lakatos, I. 1968. "Criticism and the methodology of scientific research programmes." In *Proceedings of the Aristotelian society*, Vol. 69, 149-186. Aristotelian Society, Wiley.

Leydesdorff, L. and I. Rafols. 2011. "Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations." *Journal of Informetrics* 5(1): 87-100.

Lichnerowicz, A. 1972. "Mathematic and transdisciplinarity." In: *Interdisciplinarity: Problems of teaching and research in universities*, edited by L. Apostel, G. Berger, A. Briggs, and G. Michaud, 121-127. Washington, DC: OECD.

Lizhang, W. ed. 2012. *Molecular Dynamics - Theoretical Developments and Applications in Nanotechnology and Energy*. Intech (electronic resource). Retrieved from https://www.intechopen.com/books/molecular-dynamics-theoretical-developments-and-applications-in-nanotechnology-and-energy.

Maddy, P. 2007. *Second Philosophy*. New York: Oxford University Press.

Mujumdar, A.G. and T. Singh. 2016. "Cognitive science and the connection between physics and mathematics." In *Trick or Truth? The Mysterious Connection Between Physics and Mathematics*, 201-217. Cham (Switzerland): Springer International Publishing.

Nicolescu, B. 2014. "Methodology of transdisciplinarity." *World Futures* 70(3-4): 186-199.

Parker, J. 2002. "A New Disciplinarity: Communities of Knowledge, Learning and Practice." *Teaching in Higher Education* 7(4): 373-386.

Pelikan, J. 1989. *The Christian Tradition: A History of the Development of Doctrine, Volume 5: Christian Doctrine and Modern Culture (since 1700)*. Chicago: University of Chicago Press.

Piaget, J. 1972. "The epistemology of interdisciplinary relationships." In *Interdisciplinarity: Problems of teaching and research in universities*, edited by L. Apostel, G. Berger, A. Briggs, and G. Michaud, 127-139. Washington, DC: OECD.

Shapere, D. 1984. *Reason and the search for knowledge*. Dordrecht: D. Reidel.

Squires, G. 1992. "Interdisciplinarity in Higher Education in the United Kingdom." *European Journal of Education* 27(3): 201-210.

Teissier, B. 2005. "Protomathematics, perception and the meaning of mathematical objects." In *Images and Reasoning*, edited by P. Grialou, G. Longo, and M. Okada. Tokyo: Keio University.

Turner, B. S. 2000. "What are Disciplines? And How is Interdisciplinarity Different?" In *Practising Interdisciplinarity*, edited by P. Weingart and N. Stehr, 46-65. London: University of Toronto Press.

Wigner, E. P. 1960. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." *Communications on Pure and Applied Mathematics* 13(1): 1-14.

Ye, F. 2009. "The Applicability of Mathematics as a Scientific and a Logical Problem." *Philosophia Mathematica*, doi:10.1093/philmat/nkp014, 1-22.

A KNOWLEDGE-FIRST ACCOUNT OF GROUP KNOWLEDGE

Domingos FARIA

ABSTRACT: The aim of this paper is to relate two trending topics in contemporary epistemology: the discussion of group knowledge and the discussion of knowledge-first approach. In social epistemology no one has seriously applied and developed Williamson's theory of knowledge-first approach to the case of group knowledge yet. For example, scholars of group knowledge typically assume that knowledge is analyzed in terms of more basic concepts, such as group belief or acceptance, group justification, and so on. However, if Williamson's theory of knowledge is correct, these are not good analyzes for understanding group knowledge. For, in such framework, knowledge is not analyzed in terms of belief and justification, and the same should apply to group knowledge. Thus, we propose to analyze which consequences Williamson's theory has for social epistemology, namely for an understanding of group knowledge. The questions that will guide this article are the following: What is a knowledge-first approach to group knowledge? And what does a knowledge-first approach teach us with regard to one of the most pressing issues of social epistemology, namely the dispute between summativists and non-summativists accounts of groups? We claim that a knowledge-first account of group knowledge can be offered and that it favors non-summativism.

KEYWORDS: group knowledge, group epistemology, knowledge-first epistemology, extended mind thesis

1. Introduction

The aim of this paper is to relate two trending topics in contemporary epistemology: the discussion of *group knowledge* and the discussion of *knowledge-first* approach. In social epistemology no one has seriously applied and developed Williamson's theory of knowledge-first approach to the case of group knowledge yet.[1] For example, scholars of group knowledge typically assume that knowledge is analyzed in terms of more basic concepts, such as group belief or acceptance, group justification, and so on (Tuomela 2004; Corlett 2007; Hakli 2007; Gilbert 2014; Klausen 2014; Habgood-Coote 2019, and Lackey 2020). However, if Williamson's

[1] A recent exception is the work done by Simion, Carter, and Kelp (2021); but the focus is different from what we present here. It is worth stressing that some group knowledge theories, although not a direct application of Williamson's epistemology, are compatible with knowledge-first epistemology, such as Kallestrup (2019) and Bird (2020).

theory of knowledge is correct, these are not good analyzes for understanding group knowledge. For, in such framework, knowledge is not analyzed in terms of belief and justification, and the same should apply to group knowledge. Thus, we propose to analyze which consequences Williamson's theory has for social epistemology, namely for an understanding of group knowledge.[2] The questions that will guide this article are the following: What is a knowledge-first approach to group knowledge? And what does a knowledge-first approach teach us with regard to one of the most pressing issues of social epistemology, namely the dispute between summativists and non-summativists accounts of groups? We claim that a knowledge-first account of group knowledge can be offered and that it favors non-summativism.

This article is divided in 4 sections. In section 2 we will start with a basic presentation of the main ideas of knowledge-first epistemology. In this part we emphasize that, according to Williamson (2000), individual knowledge is the most general factive mental state. In section 3, we extend Williamson (2000)'s account to group knowledge, hence offering a knowledge-first account for group epistemology. Within a knowledge-first framework, we will defend a non-summativist account of group knowledge, according to which *group knowledge* is the most general collective factive mental state. And, in section 4, we will argue that one can understand *collective mental states* based on the extended mind thesis; thus, there is no resistance to a knowledge-first account of group knowledge.

2. Knowledge-First Epistemology

Let us start with the *knowledge-first* approach.[3] Williamson (2000, 33) argues that "the concept *knows* cannot be analyzed into more basic concepts," such as belief, truth, and justification. Instead, knowledge is prior to other epistemic kinds, the concept of *knowing* being a theoretical primitive. In other words, knowledge is a starting point for explaining other notions: we can analyse and explain belief, justification, evidence, assertion, etc., by reference to knowledge. For instance, "believing *p* is, roughly, treating *p* as if one knew *p*" (Williamson 2000, 47); "a belief is fully justified if and only if it constitutes knowledge" (Williamson 2014, 5); "all and only knowledge is evidence" (Williamson 2000, 193); and so on.

[2] We use the term "groups" to refer only to organized or structured groups, such as committees, research teams, and so on.

[3] It is important to note that Williamson (2000) is the originator of knowledge-first epistemology and its most important proponent. And there are other knowledge-firsters besides Williamson (for example, Carter (2017), Simion (2019), Antognazza (2020)). However, in this paper we want to focus our attention only on Williamson's version.

The main reasons for supporting this approach, in which knowledge is the central concept, are the following: Not all concepts are susceptible to analysis in terms of more basic necessary and sufficient conditions, at least in a non-trivial and non-circular manner. This seems to happen with several common concepts like red, heap, bald, tall, old, etc., but also with the concept of knowledge. So, against the orthodoxy, it seems that there aren't non-circular necessary and sufficient conditions for knowledge, just as there aren't such non-circular conditions for analysing, for example, the color red. This is confirmed inductively by the continued failure to solve the Gettier problem; for "a succession of increasingly complex analyses have been overturned by increasingly complex counterexamples," as Williamson (2000, 30) underlies.

Although knowledge cannot be analyzed in and reduced to more basic components, a sufficiently informative characterization can be presented. Williamson (2000, 34)'s proposal is to give a positive and externalist understanding of knowledge as a state of mind. More precisely, he holds that "knowing is the most general factive stative attitude, that which one has to a proposition if one has any factive stative attitude to it at all." In other words, knowledge is the most general factive mental state, that is, the most general attitude that one can have to true propositions only.

This characterization needs some clarification. First, "knowing is *merely* a state of mind;"[4] namely, there is a mental state such that being in that state is a necessary and sufficient condition for knowing p. Consequently, contrary to the orthodox approach, knowledge is not a hybrid of mental and non-mental elements.[5] Thus, knowledge is not a conjunction of belief (the mental component), truth (the non-mental component), and something else. Instead, knowledge purely is a mental state, given its resemblance to other mental states. This understanding of knowledge as a mere state of mind is a "presumption" that we must accept, unless we have a powerful reason to reject it.[6]

Secondly, unlike mental states like imagining, desiring, hoping, etc., the mental state of knowing is *factive*. A factive attitude is one that a subject S can have only to truths; for, from 'S knows that p' we can validly deduce that 'p is true.' The fact that knowledge is a factive attitude does not prevent it from being a mental state. As Williamson (2000, 22) points out, "factive attitudes have so many similarities to the non-factive atitudes that we should expect them to constitute mental states too."

[4] See Williamson 2000, 21.

[5] Note, however, that *believing truly* is not a mental state since, in this case, there is a combination of a mental state with a non-mental state. See Williamson 2000, 28.

[6] See Williamson 2000, 22.

Namely, this factive dimension of mental states must be understood in the light of an externalist approach according to which mental states can depend on the external world.[7] For this reason, the essence of such factive mental states "includes a matching between mind and world."[8] Other factive mental states include perceiving, remembering, regretting, etc., and it would be strange to postulate, for example, a mental state of imagining but no mental state of remembering, and so on.

Thirdly, and finally, knowing is not just a factive mental state, but the *most general* of that kind. For, the other factive mental states, like perceiving that p or remembering that p, imply knowing that p. In other words, the other factive mental states imply the truth of their content because they entail knowledge.[9] For example, if a subject S really sees that it is raining, then S knows that it is raining. There are thus several specific ways in which one can know, such as seeing, remembering, etc. Williamson (2000, 39) seeks to elucidate this idea more formally, showing that such factive attitudes are expressed in natural language through a *factive mental state operator* (an FMSO) which is semantically unanalysable. The central points of FMSO can be summarized as follows: 'Know' is an FMSO; and if ϕ is an FMSO, then from 'S ϕs that p' one may infer both 'p' and 'S knows that p.' This allows us to capture the idea of knowing as the most general factive mental state. For the sake of my argument, let us take this approach for granted. Does it apply to group knowledge and, if so, how?

3. Towards Knowledge-First Group Epistemology

We commonly attribute knowledge to groups. For example, we say that "the United Nations knows that the coronavirus is widening global inequality;" "the World Health Organization knows that vaccines against coronavirus have a high probability of success;" "the jury knows that the accused is innocent." How should we understand these group knowledge attributions? Is group knowledge always reducible to knowledge of the members of the group or not? In response to these problems, there are two approaches: summativism and non-summativism. According to the summativist view, ascribing knowledge to a group is an indirect way of ascribing such knowledge to its members. Its first formulation is attributed

[7] According to Williamson (2000, 58), "if taking the externalist attitude of rational belief to a given content can contribute to one's mental state, why cannot taking the externalist attitude of knowledge to that content also contribute to one's mental state? (…) [Once content-externalism is admitted] the denial that knowing is a mental state [is] ill-motivated."

[8] See Williamson 2000, 40.

[9] It is relevant to note that 'believe truly' does not count as factive mental state, given that it has a non-mental component, that is, the truth. See Williamson 2000, 39.

to Quinton (1976), but there are several versions of summativism, some being more plausible than others. Lackey (2020) developed a minimal version of summativism which provides a *necessary* but not sufficient condition for group knowledge. This minimal version holds that a group g has knowledge K *only if* some members of g have that knowledge K. So, the minimal summativist view implies the following characterization of group knowledge:

> Summativism $=_{df}$ Necessarily, a group g has knowledge K only if at least one individual i is both a member of g and has K.

$$\Box(Kg \rightarrow \exists i(i \in g \wedge Ki))$$

By contrast, the non-summativist view of group knowledge rejects this thesis, holding that group knowledge might diverge from individual members' knowledge. So, a group g can have knowledge K even when no member of g has K. There are also several versions of non-summativism, such as Gilbert (1989; 2014), Bird (2010; 2014), Carter (2015), and Kallestrup (2019). But all of these approaches accept this characterization of group knowledge:

> Non-summativism $=_{df}$ Possibly, a group g has knowledge K even when no individual member of g has K.

$$\Diamond(Kg \wedge \neg\exists i(i \in g \wedge Ki))$$

If Williamson's approach to knowledge-first is correct, how should group knowledge be understood? Should it be understood in a summative way? Or, instead, in a non-summative way? When Williamson discusses the nature of the evidence, he appears to be inclined to accept non-summativism. In this regard Williamson (2000, 185) argues that

> The communal case is needed: science depends on public evidence, which is neither the union nor the intersection of the evidence of each scientist. We can ascribe such knowledge by saying that p is known in community S, or that we know p, which is not equivalent to saying that some, many, most, or all of us know p.

Although Williamson (2000) does not develop his argument beyond this quote, he seems to accept non-summativism and reject summativism. Therefore, according to Williamson, group knowledge isn't reducible to individual knowledge and to individual factive mental states. Instead, one has to resort to *collective factive mental states* in order to accommodate for group knowledge. In this framework, while individual knowledge is an individual factive mental state, group knowledge is a collective factive mental state. To develop Williamson's argument, it is worth stressing that the general structure of Williamson (2000)'s argument is as follows:

1. Evidence equals knowledge.

2. Communal or group evidence does not equal or reduce to individual evidence.

3. Therefore, communal or group knowledge does not equal or reduce to individual knowledge.

The first premise is dependent on his externalist theory of evidence. In such theory, for every individual or community S, Williamson (2000, 185) equates S's evidence with S's knowledge. So, S's total evidence is simply S's total knowledge. This thesis is called "$E = K$". According to this theory, a hypothesis is supported by S's evidence if only if that hypothesis is supported by S's knowledge.

Now, according to the second premise, communal or group evidence is not (reducible to) individual evidence. Indeed, it is possible that E is part of a community or group's evidence, while none of its members has E. We can support this idea with the help of *divergence arguments*, that is, situations in which the epistemic states of some group differ from the epistemic states of individual members of that group. Williamson (2000, 185) agrees that such a situation occurs in science since public scientific evidence is not the same as the evidence available to each scientist. Although Williamson does not offer a concrete example, we will use an example inspired by Bird (2014).

It is possible that a scientific research team deals with complex evidence such that the evidence cannot be obtained by any individual alone. As Bird (2014, 55) holds, "no modern science depends for its conclusions just on the intellect and the evidence of the senses of a single scientist nor even of a local team of scientists." Instead, several individuals are given roles in gathering different pieces of evidence. This happens, for example, in the scientific works developed at CERN: results are published in papers authored by hundreds of authors, but the evidence and content of these papers is not fully known by each one of them.[10] In such a case there is a *distributed cognition*, i.e. a division of cognitive labor within a group or research team.[11] Here the task of gathering evidence is divided into subtasks. Each subtask is assigned to different individual members, depending on the area of expertise. However, no member of the group is able to grasp each other's tasks. In such scientific work it is possible that the communal or group evidence is different from the individual evidence. This can be illustrated with this toy example inspired by Bird (2010, 34–35; 2014, 57–58):

> RESEARCH TEAM: Dr. X is a physicist and Dr. Y is a mathematician. Both are collaborating on a project to gather evidence E_β to support a conjecture β. This

[10] See Cetina 1999, Ridder 2019, and Palermos 2020.

[11] The classic example of *distributed cognition* is presented by Hutchins (1995).

> project has three parts. The first part is about gathering physical evidence E_α which is a task that Dr. X will do alone. The second part is about gathering mathematical evidence that E_α implies E_β which is a task that Dr. Y will do alone. And the last part is an application of *modus ponens* to the results of parts one and two, for which Dr. X and Dr. Y provided a prewritten text like this: if we have evidence E_α and if we show that evidence E_α implies evidence E_β, we also have evidence E_β. They arrange for an assistant to publish a scientific paper with these three parts if and only if the assistant receives from Dr. X the evidence E_α and from Dr. Y the evidence that E_α implies E_β. Suppose further that Dr. X and Dr. Y have no other communication with each other; but each one reached the desired evidence without the other knowing this result and, then, a paper with the evidence E_β is published.

In such a case the research team has the evidence E_β; but none of its individual members has it.[12] Because Dr. X only has the evidence E_α; and Dr. Y only has the evidence that E_α implies E_β; but neither individually has or is aware of the other scholar's evidence. Nor does the assistant have such evidence E_β, given that he limited himself to assembling the three parts and publishing them in a single document. Since $E = K$, we can conclude that in this case the research team has knowledge while none of its individual members has that knowledge. Thus, non-summativism is true. There are many other examples to help establish that conclusion. For instance,

> LAW-COURT: Suppose that a jury, as a group, following the best epistemic standards has evidence E that supports the belief that a particular refugee defendant is innocent. However, imagine that each individual member is *biased* and individually lacks the evidence E; thus, each individual member has no evidence to support the belief that the refugee defendant is innocent (instead, each one privately believes that the refugee defendant is guilty). In this case, a group has some evidence E, but each individual member does not have E. Since $E = K$, we can hold that, in such case, the group, the jury, has knowledge while none of its individual members has that knowledge.[13]

> PILGRIMS: Imagine that Joseph and Mary want to go from location α to location γ. Each individual only partially remembers the path (for instance, Joseph has evidence from α to β and Mary has evidence from β to γ). Therefore, none of these individuals has evidence of the full path from α to location γ. However, suppose that Joseph and Mary have been working together as a group; in such a case, the

[12] As an objection, based on the work of Lackey (2014), it can be said that in RESEARCH TEAM we only have evidence *production* and we have no evidence *possession*. But we can answer that whoever is reading this research team's paper naturally attributes such evidence to that team. Hence the team also possesses evidence. See Ridder 2013.

[13] This case is adapted from Kallestrup 2016.

> retrieval of all the information about the path is done by Joseph and Mary working together. So, as a group, they remember the full path from α to location γ and hereby have evidence of that path. And we can say that Joseph and Mary, as a group, have dispositional evidence of the full path even before consulting each other. Since $E = K$, and even before they consult each other, this group has dispositional knowledge of the full path, while none of its individual members has that knowledge.[14]

If these cases (or at least one of them) are plausible, premise 2 of Williamson (2000)'s argument is established. Thus, based on the thesis that $E = K$, as supported in premise 1, it can be concluded that communal or group knowledge does not equate or reduce to individual knowledge. Therefore, non-summativism is true.[15] But if that is so, and given that individual knowledge is the most general factive mental state, then it is necessary to appeal to *collective factive mental states* in order to account for group knowledge. Following Williamson (2000), it seems that a consequence of knowledge-first approach to group knowledge is to acknowledge factive mental states at the *collective level* (in ways that are irreducible to the individual level). It is thus necessary to provide and develop a plausible account of collective minds and collective mental states in non-summativist terms. But are there collective factive mental states?

4. Group Knowledge as a Collective Mental State

One may be tempted to answer this question negatively. The main resistance to accepting collective mental states is based on the following reasoning: Mental states are located in minds, that is, in heads or brains; groups do not have heads or brains; therefore, groups cannot have minds or mental states.[16] However, this is not a sound

[14] This idea is supported by studies of cognitive psychology; see Wegner, Giuliano, and Hertel 1985.
[15] It is possible to criticize such a conclusion holding, based on Wray (2001), Meijers (2002), Hakli (2007), that if groups have knowledge, then groups have beliefs; but groups cannot have beliefs (given that group doxastic states are voluntary); therefore, groups cannot have knowledge. As reply it can be said that the group doxastic states in RESEARCH TEAM and in PILGRIMS seem to be involuntary. Furthermore, it is very doubtful that involuntariness is a necessary condition for something to be a belief, because there are beliefs based in deliberation. Another possible reply is to defend that knowledge does not imply belief, even in a knowledge-first framework, as supported by Antognazza (2020). Also as an objection, following Lackey (2020), one can try to analyze cases like RESEARCH TEAM, LAW-COURT, and PILGRIMS as groups that are *in a position to know* rather than groups that know. However, we can reply by saying that the groups in such cases actually know the target propositions, since they have at least a dispositional belief and knowledge that can be readily retrieved for active deployment in reasoning; and in LAW-COURT case the group has not only a disposition to believe, but has an occurrent belief.
[16] See Tollefsen 2006.

argument, because we can plausibly deny the premise that says "mental states are located in minds and minds are in heads." More precisely, we can deny such premise and support our conclusion that there are collective factive mental states with the help of the Extended Mind Thesis (EMT).[17] According to EMT, the realization of mental states sometimes extends beyond skin and skull's boundaries to include non-biological material components and/or other biological individuals. We can formulate this thesis with the slogan *"mental states ain't all in the head."*[18]

In support of EMT, Clark and Chalmers (1998, 8) propose a *Parity Principle* (PP), supported by functionalist intuitions, and offer examples of the mind's partial realization in the environment.[19] The main idea of PP can be summarized in this way: when devices and other things in the world function sufficiently like things we normally regard as cognitive, they too should be regard as cognitive and as parts of minds. In order to illustrate the mind's partial realization in the environment, Clark and Chalmers (1998, 12–13) propose an analogy between Inga's and Otto's cases:

> CASE 1 (INGA): Inga hears about an exhibition at the Museum of Modern Art (MOMA) in New York. She starts to think, recalls it's on 53rd Street and sets off.
>
> CASE 2 (OTTO): Otto suffers from Alzheimer's, and as a result he always carries a notebook. When Otto learns useful new information, he always writes it in the notebook. He hears about the exhibition at MOMA, retrieves the address from his notebook, and sets off.

By PP, we can hold that Otto's mind includes his notebook, because the latter is *functionally equivalent* with a part of Inga's brain. In other words, following PP, there is no relevant functional difference between the role of the internal memory storage for Inga and the role of the notebook for Otto. Then, just like Inga had a belief about the museum's location before consulting her biological memory, so did Otto have a belief about the museum's location before consulting his notebook. Both Inga and Otto have a *dispositional belief* and *dispositional knowledge* concerning the museum's location. So, according to EMT, not only Inga but also Otto knew,

---

[17] This thesis is also known as "active externalism" or "vehicle externalism." See Carter & Pritchard 2018a and Carter & Pritchard 2018b. A different but related thesis is known as the "extended cognition thesis" and is about cognitive processes and not mental states. See Pritchard 2010 and Kelp 2011.

[18] See Clark & Chalmers 1998, 8.

[19] The original formulation of PP is as follows: "If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the mental process, then that part of the world *is* (so we claim) part of the mental process" (Clark & Chalmers 1998, 8).

even before consulting his notebook, where the museum was. Thus, some of Otto's mental states are determined, in part, by the contents of his notebook.

However, EMT is ambiguous and there are at least two versions of this view, when applied to Otto's case, namely:[20]

> EMT-1: It is Otto who believes that the museum is on 53rd Street.
>
> EMT-2: It is the coupled Otto-notebook system (for short, 'Otto+'), rather than Otto, that believes that the museum is on 53rd Street.

According to EMT-1, the realization of individuals' mental states sometimes extends beyond the skull boundaries to include non-biological material components and/or other biological individuals. In this interpretation, Otto himself is the subject of extended mental states.[21] By contrast, following EMT-2, the subjects of extended mental states are hybrid or coupled systems constituted by the conjunction of individuals and artifacts or other individuals. In this latter interpretation, Otto+ coupled system involves Otto's brain and nervous systems plus other features of his environment, such as his notebook.[22]

Besides that, following EMT-2, Otto+ is a distinct agent of which both Otto and the notebook are parts. In other words, it is worth observing that the *coupled system* Otto+ is not numerically identical to Otto. Indeed, if Otto were numerically identical to Otto+, then Otto would no longer suffer from any deficiency in memory, given that the notebook functions as Otto's memory. But Otto suffers from a deficiency in memory (it is only the coupled system Otto+ that lack such deficiency).[23] Therefore, Otto is not numerically identical to Otto+. On this basis, we can claim that the coupled system Otto+ has more beliefs than the biological individual Otto; and something similar happens with groups.

But is there a criterion to prevent the arbitrary formation of coupled systems? What would the criterion to form a *coupled system* be? Clark and Chalmers (1998, 17) propose the following criteria: constancy, availability, automatic endorsement,

---

[20] This ambiguity of interpretations is underlined by Adams & Aizawa (2008), Nakayama (2013), Miyazono (2017), Milojevic (2018), and Harris (2019).

[21] This is typically the standard interpretation of EMT.

[22] It is relevant to note that Clark and Chalmers (1998) sometimes seem to be inclined to accept EMT-2 instead of EMT-1. They hold, for example, that "the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right" (Clark & Chalmers 1998, 8). Such *coupled system* need not be paired with non-biological artifacts only, given that "it is entirely possible that one partner's beliefs will play the same sort of role for the other as the notebook plays for Otto" (Clark & Chalmers 1998, 17).

[23] See Preston 2010.

and prior endorsement.[24] Regarding EMT-2, we can say that Otto+ is a coupled system (constituted by Otto and his notebook) because Otto has constant access to his notebook, the system is easily available to him, with an automatic and prior endorsement. However, Clark and Chalmers (1998, 17) do not regard these criteria as necessary and sufficient conditions.[25]

There are several reasons (like the one presented in the penultimate paragraph) to prefer EMT-2 interpretation over EMT-1. First of all, following Miyazono (2017), the *systems reply* (SR) to Searle (1980)'s Chinese Room thought experiment supports EMT-2.[26] According to SR, while the person inside of Searle's Chinese Room fails to understand Chinese, the *broader system* of which she is a part understands Chinese. Analogously, it is the hybrid or coupled system realized by Otto and his notebook that has a dispositional belief in the museum's location. Consequently, Otto, as an individual, does not have certain mental states that are possessed by the broader system (Otto+) of which he is part. Thus, if functionalists have a reason to accept SR, then functionalists also have a reason to accept EMT-2.

Another advantage has to do with the fact that EMT-2 handles some objections better than EMT-1. An important objection to EMT, known as "Otto 2-step objection," maintains that there are two stages to Otto's behavior: Otto believes that the museum's location is recorded in his notebook and, based on that belief, he consults his notebook to find out the museum's location. Thus, Otto does not believe in the museum's location before consulting his notebook.[27] This objection can be stated in this way, where *t1* stands for the time interval just before Otto consulted the notebook and *t2* for the time interval just after he consulted the notebook:

1. At *t1* Otto believes that the address of the museum is written in his notebook.
2. At *t1* Otto does not believe that the museum is on 53rd street.
3. At *t2* Otto believes that the museum is on 53rd street.

In support of 2, we can say that Otto cannot simply think in order to know the museum's location and that he needs to do something more. But if so, and if it is intuitive to accept 2, then EMT-1 is false. However, this objection affects only EMT-1 and not EMT-2. According to EMT-2, we can accept 1, 2, and 3 because what this version adds is the following:

[24] Such conditions are also known as "trust and glue" conditions. See Clark 2008, 46.
[25] See Clark 2008, 80.
[26] The Chinese room argument aims to refute the possibility of strong artificial intelligence.
[27] See Rupert 2009, Wikforss 2014, and Harris 2019.

4. At *t1* Otto+ believes that the museum is on 53rd street.

5. At *t2* Otto+ believes that the museum is on 53rd street.

In the same way, it can be argued, following Harris (2019), that other standard objections to the extended mind thesis, such as the cognitive bloat objection, affect only EMT-1 and not EMT-2. The main idea of the cognitive bloat objection is that EMT implies that a subject has an implausible range of extended beliefs and knowledge.[28] For example, if a student has the same relationship with a cheat sheet as Otto has with his notebook, then we can say that the student who cheats during an exam has the same knowledge and credit as a student who is academically honest and does not cheat. But this is an implausible consequence. However, again, this consequence applies only to EMT-1 and not to EMT-2. For, according to EMT-2, the cheating student, as an individual, does not have the relevant knowledge and does not deserve credit on his exam result. Still, this student is part of a broader cognitive system that has that knowledge. So, EMT-2 seems more resistant to objections than EMT-1.

Finally, and most importantly for our purposes here, EMT-2 offers support to the idea of collective mental states.[29] For, according to EMT-2, the relevant interactions between individuals and artifacts or other individuals give rise to a *coupled system* that functions as a cognitive system in its own right.[30] So, we can hold that group minds or collective mental states are realized in *coupled systems* that are constituted primarily by interactions between humans. In other words, similarly to what happens in the Otto case, in which there is a coupled system (i.e. Otto+) with mental states formed by interactions between Otto and his notebook, it is also possible that a group and certain relevant interactions between individuals form a coupled system with mental states.

Thus, group mentality is a form of EMT-2; that is, group minds are instances of coupled systems. This way of understanding collective mental states, as being a mere instance of EMT-2, is hereby not ontologically mysterious. In addition, EMT-2 offers a way of understanding group knowledge as a mental state at the collective level, along the lines of the knowledge-first approach. In other words, EMT-2 offers a plausible account of collective mental states in a non-summativist way. For instance, in the PILGRIMS case (§3), we can see that Joseph and Mary have been working together as a group so that we can say that they form a *coupled system* that

[28] The cognitive bloat objection is developed, for example, by Allen-Hermanson (2012), Wikforss (2014), and Aizawa (2018).

[29] See Tollefsen 2006, Theiner, Allen, & Goldstone 2010, and Harris 2020.

[30] "Relevant interactions" are those underlined by "trust and glue" conditions, such as constancy, availability, automatic endorsement, and prior endorsement.

has dispositional knowledge of the full path, although none of its individual members has such knowledge. Likewise, in the LAW-COURT and RESEARCH TEAM cases, it appears that similar coupled systems are formed, that is, systems that have more mental states than their individual parts.

In short, as the resistance to collective mental states is misplaced (given that EMT-2 is plausible), then we can hold that group knowledge is the most general collective factive mental state, within a knowledge-first epistemology and along non-summativist lines.

5. Conclusion

This paper examined the consequences of *knowledge-first epistemology*, namely the thesis that knowledge is the most general factive mental state, with regard to group knowledge. The knowledge-first approach, when applied to *group knowledge*, requires to account for factive mental states at the collective level. This is a first consequence. We argue that we can make sense of *collective mental states*, in a non-summativist way, with the help of the Extended Mind Thesis as long as we interpret it as EMT-2. This knowledge-first account of group knowledge has several advantages, since it opens the path to explaining group belief, group justification, group evidence, group assertion, etc., with reference to group knowledge, reversing the traditional order of explanation. For instance, consider the following claims concerning group epistemology: a group believes that p insofar as that group treats p as if it knew p; a group belief is fully justified if and only if it constitutes group knowledge; the evidence available to some group is the group's knowledge; a group should assert p only if this group knows p; and so on. Thus, in principle, we have an elegant and unified framework for understanding the phenomena of group epistemology. Of course, defending the tenants of this new approach in detail should wait for another occasion.[31]

[31] Acknowledgements: I am grateful to Ricardo Santos, Federico Lauria, Bruno Jacinto, Michel Croce, Bogdan Dicher, David Yates, Célia Teixeira, Diogo Santos, Diogo Fernandes for helpful comments and discussion on an earlier version of this paper. Earlier versions of this paper have been presented at the *6th Congress of the Brazilian Society for Analytic Philosophy*, the *4th International Congress of the Portuguese Philosophical Society*, and the *LanCog Seminar Series in Analytic Philosophy*. I am grateful to the audience for helpful discussion. Any errors or omissions are my responsibility. Work for this paper was supported by the post-doctoral project CEECIND/01066/2017 of the Portuguese Foundation for Science and Technology. Affiliation: LanCog, Centre of Philosophy, University of Lisbon. E-mail: domingosfaria@campus.ul.pt.

References

Adams, Frederick and Kenneth Aizawa. 2008. *The Bounds of Cognition*. Wiley. https://www.ebook.de/de/product/6741551/frederick_adams_kenneth_aizawa_the_bounds_of_cognition.html.

Aizawa, Kenneth. 2018. "Extended Cognition, Trust and Glue, and Knowledge." In *Extended Epistemology*. Oxford University Press. https://doi.org/10.1093/oso/9780198769811.003.0004.

Allen-Hermanson, Sean. 2012. "Superdupersizing the Mind: Extended Cognition and the Persistence of Cognitive Bloat." *Philosophical Studies* 164 (3): 791–806. https://doi.org/10.1007/s11098-012-9914-7.

Antognazza, Maria Rosa. 2020. "The Distinction in Kind Between Knowledge and Belief." *Proceedings of the Aristotelian Society*, October. https://doi.org/10.1093/arisoc/aoaa013.

Bird, Alexander. 2010. "Social Knowing: The Social Sense of 'Scientific Knowledge'." *Philosophical Perspectives* 24 (1): 23–56. https://doi.org/10.1111/j.1520-8583.2010.00184.x.

———. 2014. "When Is There a Group That Knows?" In *Essays in Collective Epistemology*, edited by Jennifer Lackey, 42–63. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199665792.003.0003.

———. 2020. "Group Belief and Knowledge." In *The Routledge Handbook of Social Epistemology*, 274–83. Routledge. https://doi.org/10.4324/9781315717937-27.

Carter, J. Adam. 2015. "Group Knowledge and Epistemic Defeat." *Ergo, an Open Access Journal of Philosophy* 2 (20190926). https://doi.org/10.3998/ergo.12405314.0002.028.

———, ed. 2017. *Knowledge First: Approaches in Epistemology and Mind*. Oxford University Press. https://www.ebook.de/de/product/34876757/knowledge_first_approaches_in_epistemology_and_mind.html.

Carter, Adam, Jesper Kallestrup, Andy Clark, and Duncan Pritchard, eds. 2018a. *Extended Epistemology*. Oxford University Press. https://www.ebook.de/de/product/30119534/extended_epistemology.html.

———, eds. 2018b. *Socially Extended Epistemology*. Oxford University Press. https://www.ebook.de/de/product/33313185/socially_extended_epistemology.html.

Cetina, Karin. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, Mass: Harvard University Press.

Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford New York: Oxford University Press.

Clark, Andy and David Chalmers. 1998. "The Extended Mind." *Analysis* 58 (1): 7–19. https://doi.org/10.1093/analys/58.1.7.

Corlett, J. Angelo. 2007. "Analyzing Social Knowledge." *Social Epistemology* 21 (3): 231–47. https://doi.org/10.1080/02691720701674049.

Gilbert, Margaret. 1989. *On Social Facts*. Routledge.

———. 2014. *Joint Commitment: How We Make the Social World*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199970148.001.0001.

Habgood-Coote, Joshua. 2019. "Group Knowledge, Questions, and the Division of Epistemic Labour." *Ergo, an Open Access Journal of Philosophy* 6 (20200523). https://doi.org/10.3998/ergo.12405314.0006.033.

Hakli, Raul. 2007. "On the Possibility of Group Knowledge Without Belief." *Social Epistemology* 21 (3): 249–66. https://doi.org/10.1080/02691720701685581.

Harris, Keith. 2019. "Whose (Extended) Mind Is It, Anyway?" *Erkenntnis*, September. https://doi.org/10.1007/s10670-019-00172-9.

———. 2020. "Group Minds as Extended Minds." *Philosophical Explorations* 23 (3): 234–50. https://doi.org/10.1080/13869795.2020.1799663.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, Mass: MIT Press.

Kallestrup, Jesper. 2016. "Group Virtue Epistemology." *Synthese* 197 (12): 5233–51. https://doi.org/10.1007/s11229-016-1225-7.

———. 2019. "Groups, Trust, and Testimony." In *Trust in Epistemology*, 136–58. Routledge. https://doi.org/10.4324/9781351264884-6.

Kelp, Christoph. 2011. "Extended Cognition and Robust Virtue Epistemology." *Erkenntnis* 78 (2): 245–52. https://doi.org/10.1007/s10670-011-9301-3.

Klausen, Søren Harnow. 2014. "Group Knowledge: A Real-World Approach." *Synthese* 192 (3): 813–39. https://doi.org/10.1007/s11229-014-0589-9.

Lackey, Jennifer. 2014. "Socially Extended Knowledge." *Philosophical Issues* 24 (1): 282–98. https://doi.org/10.1111/phis.12034.

———. 2020. *The Epistemology of Groups*. Oxford University Press.

Meijers, Anthonie. 2002. "Collective Agents and Cognitive Attitudes." *ProtoSociology* 16: 70–85. https://doi.org/10.5840/protosociology20021621.

Milojevic, Miljana. 2018. "Extended Mind, Functionalism and Personal Identity." *Synthese* 197 (5): 2143–70. https://doi.org/10.1007/s11229-018-1797-5.

Miyazono, Kengo. 2017. "Does Functionalism Entail Extended Mind?" *Synthese* 194 (9): 3523–41. https://doi.org/10.1007/s11229-015-0971-2.

Nakayama, Yasuo. 2013. "The Extended Mind and the Extended Agent." *Procedia - Social and Behavioral Sciences* 97 (November): 503–10. https://doi.org/10.1016/j.sbspro.2013.10.265.

Palermos, Spyridon Orestis. 2020. "Epistemic Collaborations: Distributed Cognition and Virtue Reliabilism." *Erkenntnis*, April. https://doi.org/10.1007/s10670-020-00258-9.

Preston, John. 2010. "The Extended Mind, the Concept of Belief, and Epistemic Credit." In *The Extended Mind*, 354–69. The MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0015.

Pritchard, Duncan. 2010. "Cognitive Ability and the Extended Cognition Thesis." *Synthese* 175 (S1): 133–51. https://doi.org/10.1007/s11229-010-9738-y.

Quinton, Anthony. 1976. "Social Objects." *Proceedings of the Aristotelian Society* 76 (1): 1–28. https://doi.org/10.1093/aristotelian/76.1.1.

Ridder, Jeroen de. 2013. "Epistemic Dependence and Collective Scientific Knowledge." *Synthese* 191 (1): 37–53. https://doi.org/10.1007/s11229-013-0283-3.

———. 2019. "How Many Scientists Does It Take to Have Knowledge?" In *What Is Scientific Knowledge?*, 3–17. Routledge. https://doi.org/10.4324/9780203703809-1.

Rupert, Robert. 2009. *Cognitive Systems and the Extended Mind*. Oxford University Press. https://www.ebook.de/de/product/9235551/robert_d_rupert_cognitive_systems_and_the_extended_mind.html.

Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–24. https://doi.org/10.1017/s0140525x00005756.

Simion, Mona. 2019. "Knowledge-First Functionalism." *Philosophical Issues* 29 (1): 254–67. https://doi.org/10.1111/phis.12152.

Simion, Mona, J. Adam Carter, and Christoph Kelp. 2021. "On Behalf of Knowledge-First Collective Epistemology." In, *Doxastic and Propositional Warrant*, edited by P. Silva and L. Oliveira. Routledge.

Theiner, Georg, Colin Allen, and Robert Goldstone. 2010. "Recognizing Group Cognition." *Cognitive Systems Research* 11 (4): 378–95. https://doi.org/10.1016/j.cogsys.2010.07.002.

Tollefsen, Deborah Perron. 2006. "From Extended Mind to Collective Mind." *Cognitive Systems Research* 7 (2-3): 140–50. https://doi.org/10.1016/j.cogsys.2006.01.001.

Tuomela, Raimo. 2004. "Group Knowledge Analyzed." *Episteme* 1 (2): 109–27. https://doi.org/10.3366/epi.2004.1.2.109.

Wegner, Daniel M., Toni Giuliano, and Paula T. Hertel. 1985. "Cognitive Interdependence in Close Relationships." In *Compatible and Incompatible Relationships*, 253–76. Springer New York. https://doi.org/10.1007/978-1-4612-5044-9_12.

Wikforss, Åsa. 2014. "Extended Belief and Extended Knowledge." *Philosophical Issues* 24 (1): 460–81. https://doi.org/10.1111/phis.12043.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press. https://doi.org/10.1093/019925656x.001.0001.

———. 2014. "Knowledge First." In *Contemporary Debates in Epistemology*, edited by J. Turri and E. Sosa, Second Edition. Wiley-Blackwell.

Wray, K. Brad. 2001. "Collective Belief and Acceptance." *Synthese* 129 (3): 319–33. https://doi.org/10.1023/a:1013148515033.

A CONNEXIVE CONDITIONAL

Mario GÜNTHER

ABSTRACT: We propose a semantics for a connexive conditional based on the Lewis-Stalnaker conditional. It is a connexive semantics that is both classical and intuitive.

KEYWORDS: connexivity, Lewis-Stalnaker conditional, semantics

1. Introduction

Connexive logic and conditionals are intertwined. The implications of connexive systems have long been thought to be candidates for analysing indicative and subjunctive conditionals (Angell 1962; McCall 1966; 1975; Cantwell 2008; Kapsner and Omori 2017). McCall (2014, 964) says that the logic of causal and subjunctive conditionals is connexive "since 'If X is dropped, it will hit the floor' contradicts 'If X is dropped, it will not hit the floor'." Weiss (2019) puts forth sound and complete connexive logics for conditionals, using an algebraic semantics in the tradition of Nute (1980). However, what is badly missing is an intuitive semantics for a connexive conditional. As Kapsner (2019, 516) recently pointed out, no strongly connexive logic has "an intelligible and well-motivated semantics."

Why is an intuitive semantics for a strongly connexive conditional so hard to come by? One of the reasons is that strongly connexive logics are usually many-valued and it is hard to give some of those values an intuitive meaning. Another reason is that many connexive logics do not validate the principle of classical logic that anything follows from an arbitrary contradiction. The consequence relation of those logics is paraconsistent: the relation is non-trivial even if the premises are inconsistent (Priest et al. 2018). This non-trivial reasoning from inconsistent premises poses at least a challenge for intelligibility (Rudnicki 2021).

Here we propose an intuitive semantics for a connexive conditional. We do so by defining the connexive conditional in terms of a Lewis-Stalnaker conditional. Notably, we remain within the realm of consistent classical semantics for conditionals (in the tradition of Stalnaker (1968), Lewis (1973a), and Chellas (1975)). The connexive semantics we propose is a modal extension of classical logic (Kripke 1963). Unlike many connexive logics, our conditional semantics validates, in particular, the principle of bivalence.

Section 2 outlines what a connexive conditional is and explains why the current contenders for an intuitive semantics for conditionals are not connexive.

Section 3 introduces the idea to restrict the guiding ideas of connexivity to consistent antecedents. Section 4 proposes our semantics for a non-restrictedly connexive conditional. Section 5 concludes the paper.

2. What is a Connexive Conditional?

There are two guiding ideas behind connexive logics. One of them is that no proposition can entail its own negation (Anderson & Belnap 1975; Wansing 2020). This idea can be understood as follows: for no proposition A, $A > \neg A$ is satisfiable, where $>$ expresses a conditional or an implication. The other guiding idea of connexive logic is that if A implies C, then A does not imply $\neg C$. In symbols, if $A > C$, then $A \not> \neg C$.[1]

Nowadays, connexive logics are taken to express the two guiding ideas by validating the following principles.

$$\neg(A > \neg A) \qquad \text{(AT')}$$

$$(A > C) > \neg(A > \neg C) \qquad \text{(BT)}$$

Indeed, a logical system is currently said to be connexive only if AT' and BT are theorems and the conditional is non-symmetric (Wansing 2020).[2] It is worth mentioning that many non-classical connexive logics satisfy AT' and BT, but still violate our guiding ideas. In the logic of Cantwell (2008), for instance, $A > \neg A$ is satisfiable, and so is $A > C$ and $A > \neg C$ at the same time. As Kapsner (2012, 143) puts it, it "seems clear to me that a logic like Cantwell's [...] clearly violates the original intuitions that first prompted the adoption of [AT'] and [BT]." The 'original intuitions' or guiding ideas behind connexive logic are, of course, validated by other non-classical connexive implications, e.g. Angell (1962) and McCall (1966).

The two guiding ideas are highly plausible for an intuitive semantics of conditionals. And yet the mainstream contenders for the semantics of our everyday conditional violate the guiding ideas. Jackson defends that the material implication $\rightarrow$ of classical logic gives the truth conditions of *if* (Jackson 1979; 1980). However, $A \rightarrow \neg A$ is true whenever A is false. On the variably strict semantics for conditionals, $A > \neg A$ is true whenever A is impossible (Stalnaker 1968; Lewis 1973b). And likewise on the strict analysis: where $\Box$ is a necessity operator, $\Box(A \rightarrow \neg A)$ is true whenever A is impossible (Gillies 2007; Kratzer 1989; 2012). Similarly, $A \rightarrow C$ and $A \rightarrow \neg C$ are both true whenever A is false. The Stalnaker-Lewis conditionals $A > C$ and $A > \neg C$

[1] The two guiding ideas roughly correspond to the two intuitions in Kapsner (2012, 142).

[2] Non-symmetry requires that $(A > C) > (C > A)$ is not valid. A logic is said to be connexive iff in addition to AT', BT, and non-symmetry, AT and BT' are theorems of the logic. AT says $\neg(\neg A > A)$, and BT' says $(A > \neg C) > \neg(A > C)$.

are both true whenever A is impossible, and so are the corresponding strict implications.

As we have just seen, the mainstream contenders for the everyday conditional violate the plausible guiding ideas of connexive logic. The first guiding idea speaks decisively against the material implication. Whenever A is false, $A \rightarrow \neg A$ is true. And whenever A is true, $\neg A \rightarrow \neg\neg A$ is true. Whether or not A is true, the first guiding idea is violated.[3] The first guiding idea, taken as a desideratum, challenges the material implication as giving the truth conditions for *if*. But perhaps the material implication is for other reasons the weakest of the current contenders anyways (Starr 2019; Edgington 2020).

3. Restriction to Consistent Antecedents

Why do the variably strict and strict conditionals violate the guiding ideas of connexive logic? The reason seems to be that, from a classical perspective, an inconsistent or contradictory proposition entails everything—including its negation. We make the natural assumption according to which a proposition is inconsistent or contradictory iff it is logically impossible. To be explicit, we say that a proposition A is logically possible iff there is a possible world where A is true. A proposition A is thus impossible iff there is no world at which A is true. Our assumption then says: for each world at which a proposition is evaluated, there is a possible world where the proposition is true iff the proposition is classically consistent. Furthermore, we say that a world which satisfies a proposition A is an A-world. Hence, there is no A-world *only* if A is inconsistent.

Under our assumption, there is an obvious way to save the guiding ideas for the variably strict and strict conditionals. One may restrict the guiding ideas to consistent—or equivalently logically possible—antecedents. The first idea would become: for no logically possible proposition A, $A > \neg A$. The second idea would become: for any logically possible proposition A, if $A > C$, then $A \not> \neg C$. And indeed, the strict and variably strict conditionals validate those restricted principles. $\Box (A \rightarrow \neg A)$ is true at a world iff $\neg A$ is true at *all* worlds. By the restriction, we only look at antecedents for which there is a possible world where the antecedent is true. So there is a world where A is true and thus not all worlds make $\neg A$ true. We have that $\Box (A \rightarrow \neg A)$ is not satisfiable if A is logically possible. As to the second restricted guiding idea: suppose there is an A-world and all worlds are $\neg A \vee C$-worlds. Well then there is an A-world which is a C-world, and so it is not the case that all worlds are $\neg A \vee \neg C$-worlds.

[3] Moreover, one of AT' and AT is violated whether or not A is true.

A variably strict conditional $A \boxdot\!\!\rightarrow C$ is true at a world w just in case C is true at all A-worlds most similar to w. The restricted guiding ideas are validated. Whenever there is an A-world, $A \boxdot\!\!\rightarrow \neg A$ is false at any world. Moreover, suppose there is an A-world and C is true at all A-worlds most similar to w. Well then there is a most similar A-world at which C is true, and so it is not the case that $\neg C$ is true at all A-worlds most similar to w. So the restricted guiding ideas are validated by variably strict and strict conditionals.[4]

The idea to restrict the guiding ideas of connexive logic is not new. Lenzen (2020), for example, believes that Aristotle and Boethius—the name givers for AT' and BT respectively—intended the theses to apply only to conditionals with consistent antecedents. Lenzen (2019) states the following restricted version of AT':

$$\Diamond A > \neg(A > \neg A) \qquad \text{(LEIB1)}$$

Most variants of variably strict and strict conditionals validate LEIB1, as Lenzen and Kapsner (2019) already pointed out.[5] The latter introduces the restriction to non-contradictory antecedents under the label of *humble connexivity*. But not everyone thinks the connexive principles should be restricted. Restricting AT' and BT to consistent (or satisfiable) antecedents, so claims for example Wansing (2020, 62), makes the "whole enterprise of connexive logic pointless from the standpoint of classical logic."

Are the connexive theses restricted to satisfiable antecedents more intuitive than the original versions? Does a contradictory antecedent entail its own negation? Perhaps. Iacona (2019, 6), for instance, defends that the truth of the conditional "If it is snowing and it is not snowing, then it is snowing" is "not implausible." But, even if so, the truth of $A \wedge \neg A > B$ for any B is hardly intelligible. A contradiction does not help to understand anything. To see this, notice that the conditional "if it is snowing and it is not snowing, then it is snowing and not snowing" may well be true according to Iacona's defense. But how is the world like if it snows and it does not at the same time?

It remains to be seen whether the restricted versions of connexivity will replace the established concept of connexivity. In the meantime, we propose a semantics that puts the restriction into the meaning of the conditional rather than restricting the connexive theses themselves. The result is an intuitive semantics for an unrestrictedly connexive conditional.

[4] The variably strict and strict conditionals likewise validate AT' and BT.

[5] Lenzen correctly remarks that LEIB1, named after Leibniz, is a theorem of almost all systems of normal modal logic.

4. A Connexive Conditional

We have seen that a variably strict conditional $A \boxdot\!\!\rightarrow C$ is true at a world w just in case C is true at all A-worlds most similar to w. We modify this semantic clause to obtain a connexive semantics. The main idea is to require for the truth of a conditional that there is an antecedent world. A conditional $A > C$ is true at w iff there is an A-world and C is true at all A-worlds most similar to w. In brief, we define $A > C$ as $\Diamond A \wedge (A \boxdot\!\!\rightarrow C)$.[6]

The variably strict conditional $A \boxdot\!\!\rightarrow C$ is automatically true (at any world) when the antecedent A is impossible. By contrast, the conditional $A > C$ is automatically false (at any world) when the antecedent is impossible. If A is (logically) impossible, then there is no A-world, and so no most similar A-world w_A. Likewise, if there is an A-world, there is a most similar A-world. We have baked the restriction to logically possible (or consistent) antecedents into the truth conditions of >. To reiterate, $A > C$ is false at any world if A is logically impossible.

Here is a handy way to state the truth conditions for our conditional. Let $[C]$ abbreviate the set of all C-worlds, and $[W_A]$ the set of all A-worlds most similar to w. We can thus simplify our semantic clause:

> $A > C$ is true at w iff there is w_A and $[W_A] \subseteq [C]$.

Why is this semantics connexive? A first reason is that > satisfies our guiding ideas. For no proposition A, $A > \neg A$ is satisfiable. Indeed, $A > \neg A$ is true at no world. To see this, consider two cases. First, suppose A is possible. Then there is a most similar A-world w_A, but all w_A satisfy A and thus do not satisfy $\neg A$. In symbols, $[W_A] \nsubseteq [\neg A]$. Second, suppose A is impossible. Then there is no A-world, and a fortiori no most similar A-world w_A. Hence, $A > \neg A$ is false at every world. So AT' is a tautology for >.[7]

The conditional satisfies the other guiding idea as well. Suppose $A > C$ is true at an arbitrary world w. Then there is w_A and $[W_A] \subseteq [C]$. So there is at least one most similar A-world that satisfies C. This implies that not all most similar A-worlds are $\neg C$-worlds. Hence, $A > \neg C$ is false at the arbitrarily chosen world w. A weak form of BT is thus a tautology for >:

---

[6] Lewis (1973a, 25) and Lewis (1973b, 438) strengthen his more subtle semantics for conditionals in similar ways. Priest (1999, 145-6) strengthens the strict conditional in a similar way and arrives at a semantics for a "simple" connexivist logic. However, he does so in a more complicated way using a "symmetrised account."

[7] By similar reasoning, $\neg A > A$ is false at every world. So AT is also a tautology for >.

$$(A > C) \rightarrow \neg(A > \neg C) \qquad \text{(BTw)}$$

This establishes that our conditional is *connexive* according to Weiss (2019).[8] Moreover, our conditional satisfies the rule form of BT:

$$(A > C) \models \neg(A > \neg C) \qquad (\text{BT}^r)$$

This makes our conditional also *weakly connexive* according to Wansing and Unterhuber (2019).[9]

As is clear by now, our conditional validates the principle of conditional non-contradiction, also known as Abelard's First Principle:

$$\neg((A > C) \wedge (A > \neg C)) \qquad \text{(ABEL1)}$$

We have illustrated ABEL1 in the introduction by means of quoting McCall.

It should be mentioned that our conditional does not validate BT. The reason is that $A > C$ may be impossible. Then there is no $w_{A>C}$. And so $(A > C) > \neg(A > \neg C)$ is false (at any world).[10] But since our conditional validates the guiding ideas and the other connexive principles, we think that this result rather casts doubt on BT than our conditional. BT is too demanding for expressing the second guiding idea.

Finally, and to support the point of the last paragraph, our conditional satisfies BT when the main connective is replaced by a variably strict conditional which may be vacuously true:

$$(A > C) \mathrel{\square\!\!\rightarrow} \neg(A > \neg C) \qquad \text{(VBT)}$$

Suppose $A > C$ is impossible. Then there is no $w_{A>C}$. Hence, $(A > C) \mathrel{\square\!\!\rightarrow} \neg(A > \neg C)$ is vacuously true.[11]

5. Conclusion

We have proposed a semantics for a connexive conditional. The conditional $A > C$ is true at a world w just in case there is an A-world *and* C is true at all A-worlds most similar to w. This strengthening of the Lewis-Stalnaker conditional validates the two guiding ideas behind connexive logics, as well as the connexive theses AT, AT', BTw, BTw', BT^r, BT^p, and ABEL1. Moreover, our conditional validates VBT, a version of BT where the main connective is a variably strict conditional. Indeed, the variably

[8] Again, by similar reasoning, $A > C$ is false at any world if $A > \neg C$ is true there. So BTw', $(A > \neg C) \rightarrow \neg(A > C)$, is also a tautology for >.

[9] Our conditional satisfies the rule form of BT' as well.

[10] BT' is also no tautology for our conditional.

[11] VBT' is valid for our conditional as well.

strict conditional $\Box\!\!\rightarrow$ and our conditional $>$ are interdefinable using only Boolean connectives. Where $\bot$ stands for an arbitrary contradiction, we have

$A > C \stackrel{\text{df}}{=} \neg(A \Box\!\!\rightarrow \bot) \wedge (A \Box\!\!\rightarrow C)$, and

$A \Box\!\!\rightarrow C \stackrel{\text{df}}{=} \neg(A > A) \vee (A > C)$.

Unlike others, we have not restricted the connexive theses to consistent or possible antecedents. Rather we have baked this restriction into the semantics of our conditional. The result is a semantics for (causal) conditionals which is as intuitive as the Lewis-Stalnaker semantics when the antecedent is (logically) possible. When the antecedent is (logically) impossible, our conditional is automatically false. Intuitively, the conditional "If the vase is dropped and not dropped at the same time, it will hit the floor" is at least not true. And the conditional is not exactly intelligible. For all practical purposes, we may consider it false.[12]

References

Anderson, A. R. and N. D. Belnap. 1975. *Entailment: The Logic of Relevance and Neccessity, Vol. I*. Princeton University Press.

Angell, R. B. 1962. "A Propositional Logic with Subjunctive Conditionals." *Journal of Symbolic Logic* 27(3): 327–343. doi:10.2307/2964651.

Cantwell, J. 2008. "The Logic of Conditional Negation." *Notre Dame J. Formal Logic* 49(3): 245–260.

Chellas, B. F. 1975. "Basic Conditional Logic." *Journal of Philosophical Logic* 4(2): 133–153.

Edgington, D. 2020. "Indicative Conditionals." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University. Fall 2020 edn.

Gillies, A. S. 2007. "Counterfactual Scorekeeping." *Linguistics and Philosophy* 30(3): 329–360.

Iacona, A. 2019. "Strictness and Connexivity." *Inquiry* 64(10): 1024-1037.

Jackson, F. 1979. "On Assertion and Indicative Conditionals." *The Philosophical Review* 88(4): 565–589.

———. 1980. "Conditionals and Possibilia." *Proceedings of the Aristotelian Society* 81: 125–137.

[12] Is the negation of "If the vase is dropped and not dropped at the same time, it will hit the floor" true? On our semantics, 'it is not the case that if the vase is dropped and not, it will hit the floor' is true. But we will never say such conditionals in an ordinary context.

Kapsner, A. 2012. “Strong Connexivity.” *Thought: A Journal of Philosophy* 1(2): 141–145. doi:10.1002/tht3.19.

———. 2019. “Humble Connexivity.” *Logic and Logical Philosophy* 28(3): 513–536.

Kapsner, A. and H. Omori. 2017. “Counterfactuals in Nelson Logic.” In *Logic, Rationality, and Interaction. LORI 2017*, edited by A. Baltag, J. Seligman, and T. Yamada, vol. 10455, 497–511. Berlin, Heidelberg: Springer.

Kratzer, A. 1989. “An Investigation of the Lumps of Thought.” *Linguistics and Philosophy* 12(5): 607–653.

———. 2012. *Modals and Conditionals: New and Revised Perspectives*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.

Kripke, S. A. 1963. “Semantical Analysis of Modal Logic I Normal Modal Propositional Calculi.” *Mathematical Logic Quarterly* 9: 67–96.

Lenzen, W. 2019. “Leibniz’s Laws of Consistency and the Philosophical Foundations of Connexive Logic.” *Logic and Logical Philosophy* 28(3): 537–551.

———. 2020. “A Critical Examination of the Historical Origins of Connexive Logic.” *History and Philosophy of Logic* 41(1): 16–35.

Lewis, D. 1973a. *Counterfactuals*. Oxford: Blackwell.

———. 1973b. “Counterfactuals and Comparative Possibility.” *Journal of Philosophical Logic* 2(4): 418–446.

McCall, S. 1966. “Connexive Implication.” *The Journal of Symbolic Logic* 31(3): 415–433.

———. 1975. “Connexive Implication.” In *Entailment. The Logic of Relevance and Necessity*, edited by A. Anderson and N. Belnap, vol. 1, 434–46. Princeton: Princeton University Press.

———. 2014). Connexive Gentzen. *Logic Journal of the IGPL* 22(6): 964– 981.

Nute, D. 1980. *Topics in Conditional Logic*. Dordrecht: Reidel.

Priest, G. 1999. “Negation as Cancellation, and Connexive Logic.” *Topoi* 18(2): 141–148.

Priest, G., K. Tanaka, and Z. Weber. 2018. “Paraconsistent Logic.” In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University.

Rudnicki, K. 2021. “The Psychological Aspects of Paraconsistency. ” *Synthese* 199: 4393–4414. https://doi.org/10.1007/s11229-020-02983-8.

Stalnaker, R. 1968. “A Theory of Conditionals.” In *Studies in Logical Theory (American Philosophical Quarterly Monograph Series)*, edited by N. Rescher, no. 2, 98–112. Oxford: Blackwell.

Starr, W. 2019. “Counterfactuals.” In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University.

Wansing, H. 2020. “Connexive Logic.” In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Metaphysics Research Lab, Stanford University.

Wansing, H. and M. Unterhuber. 2019. “Connexive Conditional Logic.” Part I. *Logic and Logical Philosophy* 28(3): 567–610.

Weiss, Y. 2019. “Connexive Extensions of Regular Conditional Logic.” *Logic and Logical Philosophy* 28(3): 611–627.

ONTOLOGICAL SOLUTIONS TO THE PROBLEM OF INDUCTION

Mohammad Mahdi HATEF

ABSTRACT: The idea of the uniformity of nature, as a solution to the problem of induction, has at least two contemporary versions: natural kinds and natural necessity. Then there are at least three alternative ontological ideas addressing the problem of induction. In this paper, I articulate how these ideas are used to justify the practice of inductive inference, and compare them, in terms of their applicability, to see whether each of them is preferred in addressing the problem of induction. Given the variety of contexts in which inductive inferences are made, from natural science to social science and to everyday thinking, I suggest that no singular idea is absolutely preferred, and a proper strategy is probably to welcome the plurality of ideas helpful to induction, and to take pragmatic considerations into account, in order to judge in every single case.

KEYWORDS: induction, uniformity of nature, natural kinds, natural necessity

In his famous critique, Hume challenged the legitimacy of the inductive inference, namely the inference from the premise "observed Fs have been Gs" to the conclusion "all Fs are Gs."[1] He was also the first one who responded to the challenge, by introducing an ontological principle, i.e., the *uniformity of nature*, joining which to the inductive premise was supposed to legitimate the inference. He was again the first one who challenged the response, as the principle cannot be established without circularity (Hume 1739). Since then, and despite the initial failure, not only was this idea not forgotten, but also further similar attempts were subsequently made by others, to solve the problem, in more or less the same way as the uniformity of nature does. Among those suggestions, two ideas were distinguished: *natural kinds* and *natural necessity*. While both were employed in the literature to address the problem of induction, there is a subtle difference between them which might be historically interesting. While the term natural kinds was initially invoked to support the inductive inference, the term natural necessity was introduced initially in metaphysics and only then be employed to legitimate the induction. Therefore, three ontological alternatives are now available, joining each to the inductive premise have been argued to legitimate the inference. In this paper, I will compare

[1] There is also an alternative schema for the inductive inference, with the premise "Observed F1… Fn have been Gs" to the conclusion "Fn+1 is G." To the extent of this paper, there is no significant difference between them, and because of its prevalence, we work with the first one.

them, in terms of their applicability and usefulness, to see whether each of them is preferred in addressing the problem of induction. I begin with the more favored idea, i.e. natural kinds, then go to the natural necessity, and finally come back to the idea of uniformity.

Natural Kinds and the Inductive Inference

The tradition of natural kinds, as Hacking (1991) noted, has been mostly concerned with the problem of induction. Mill (1843) invoked *Kinds* when he spoke of "subsidiary operations to induction." John Venn, by whom the term *natural kinds* was coined, also proposed this idea while searching for a basis to apply probability rules in the domain of natural objects and events (Venn 1876). According to him, these rules owe their applicability on artificial objects to the overall resemblance embedded, in advance, in the objects involved in trials, a resemblance which is not evident in natural objects. The idea of natural kinds, then, is supposed to provide this resemblance in nature, required for probability rules to be applied therein.[2] In Chakravartti's words, "The primary motivation for thinking that there are such things as natural kinds is the idea that carving nature according to its own divisions yields groups of objects that are capable of supporting successful inductive generalizations and predictions" (Chakravartti 2007, 152).

Nonetheless, what has been rarely articulated is details of this support, and the way in terms of which the idea of natural kinds fulfills its legitimating role in the inductive inference. Here is where my work begins. In order to articulate the details, I refer to Sankey' s account of natural kinds, as it is developed properly and addresses the problem of induction clearly. As Sankey (1997) stated:

> We are rational to employ induction when we form our beliefs about the future because nature is, in fact, uniform. It is uniform in the sense that the fundamental kinds of things which exist are natural kinds of things, which possess essential sets of properties. Because all members of a kind possess the same essential properties, unobserved members of a kind will possess the same properties as members of the kind which have already been observed. This is why, when we infer that an unobserved object will have a property which observed objects of the same kind have, we turn out to be right. For having such a property is just part of what it is to be an object of the same kind as the other objects.[3]

[2] Attempts of the same kind can be seen in Russell (1948), Quine (1969), Boyd (1999), Kornblith (1993), Sankey (1997), Ellis (2001), Chakravartti (2007), Bird (2018), etc., noting that Russell's case is somehow different, as he ultimately preferred to refer to Keynes's principle of limited variety rather than natural kinds.

[3] As clear, Sankey's account is one among many accounts of natural kinds. But its essentialist

According to Sankey, were all instances so far observed of a substance with say atomic number 26 have the melting point of 1536° C, we can conclude that all substances with atomic number 26 have the melting point of 1536° C. The reason is that having atomic number 26 is sufficient for a substance to be recognized as iron. Moreover, having the melting point of 1536° C is among the essential properties of the iron. Supposing invariability of kind essences over time and individuals, unobserved instances of iron can be claimed to possess both properties of having "atomic number 26" and "melting point 1536° C." Then, if a substance with the first property is observed, the second property can be legitimately ascribed. Here is my articulation of this inductive inference, based on a classical essentialist conception of natural kinds presupposed in Sankey's account:

Inference (A)

(1) All observed Fs have been Gs.

(2) All Fs belong to kind K.

(3) G is among the essential properties of kind K.

(4) If something is F, it necessarily is G.

Therefore,

(5) All Fs are Gs.

In the schema above, the conjunction of lines (2) to (4) and the premise (1) seems to legitimate the inference. But let us see how these lines are themselves justified. Line (4) is presumably entailed by the conjunction of (2) and (3), as necessity is generally presumed to be involved in essentialism. Line (5) is directly entailed by (4), and (1) is an observed matter of fact. What we must worry about are premises (2) and (3), which are not well-grounded. The idea which philosophers of natural kinds had in mind when they invoked these categories seems to be that (2) and (3) are justified because they provide the best explanation for an unexpected fact stated in (1). On this account, natural kinds solve the problem of induction through an IBE for the inductive premise.[4]

---

character, inherited from Ellis' account, makes it a typical instance of the classical conception of natural kinds.

[4] There is also another view among philosophers of natural kinds (Kornblith 1993; Boyd 1999) in which these categories are appreciated because they provide the best explanation for the *reliability of induction* (or *inductive success*), not the inductive premise. In this view, natural kinds are not supposed to solve the justificatory problem of induction. For an inference externalist, there seems to be no significant difference between the justification of an inductive inference and the explanation of its reliability. For an inference internalist, however, there is a significant difference

Natural Necessity as an Alternative

Considering that (4) is entailed by premises (2) and (3), and they are inferred thanks to their explanatory power as regards (1), one may ask why not initially postulate (4), rather than begin with (2) and (3), in order to legitimate the inference? In other words, instead of appealing to natural kinds and essential properties to explain (1), why not invoke necessary relations directly? Since (4) entails (1), necessary relations seem to be capable of affording a rival explanation for (1), with also an advantage, if the simplicity of explanation is taken into account and simpler is taken as metaphysically weaker. Regarding the premises, two metaphysical postulates are involved in (2) and (3): the ideas of natural kinds and essentialism, while only one metaphysical postulate is involved in (4): the idea of necessity. Since the idea of natural kinds, as it is postulated above, involves itself necessary relations between properties,[5] premise (4) turns out to be weaker. It urges us to recast the inference in a simpler form:

Inference (B)

(1) All observed Fs have been Gs.

(4) If something is F, it necessarily is G.

Therefore,

(5) All Fs are Gs.

This parsimonious version, surprisingly, is followed by an undesired consequence: the idea of natural kinds is removed from the inference, as the inference is vindicated without appealing to natural kinds. It recalls, surprisingly, an alternative solution to the problem of induction, offered by philosophers of natural necessity (Shoemaker 1980; Foster 1983; Armstrong 1983; BonJour 1998; Tooley 2011). Before articulating this suggestion, let us pause here to see whether no other advantage is supplied by natural kinds, compared to the natural necessity, which invites us to keep kinds. Among others, a virtue traditionally associated with natural kinds has been their capability to systematize the generalizations, namely to include a multitude of interconnected properties in a kind, and allow us to implicitly make these generalizations, while working with kinds. Kinds' epistemic virtue is therefore

between them. For him, then, the above-mentioned second view would not be a solution to the problem of induction, since this problem challenges the justification of inductive inference, and needs something more to be answered, something like what the tortoise asked from Achilles discussing the deduction, in Carroll (1895).

[5] Not to deny other accounts of natural kinds, Boyd (1999) for example, treats essentialism and necessitation differently.

a pragmatic one. It is, in other words, a matter of economy of thought, which contributes significantly to the abbreviation of the laws of nature postulated in the scientific theories. However, not in every case of scientific laws, natural kinds are legitimate to be presupposed. In the law of momentum conversation, i.e., $m_1v_1 = m_2v_2$, for example, only properties are illustrated, and no kind, implicitly or explicitly, is involved. Then, one may claim that there are at least some generalizations in science to which natural kinds do not contribute.

The fact that natural kinds are only pragmatically useful, and even in some cases not pragmatically useful, corroborates the idea that, ontologically speaking, kinds are not as substantial as properties. It could support views of natural kinds in which these categories were taken to be interim tools of practicing science. Recall, for example, Russel's claim that the doctrine of natural kinds "is only an approximate and transitional assumption on the road towards fundamental laws of a different kind" (Russell 1948, 391), or Quine's claim that "we can take it as a very special mark of the maturity of a branch of science that it no longer needs an irreducible notion of similarity and kind" (Quine 1969). It also works as an argument against the ontologies which regard kinds as substantial as properties,[6] as far as the epistemic utility of natural kinds was regarded to contribute to their substantiality.

Hence, as argued above, there are at least some inductive inferences which can be legitimated by appealing to the idea of necessity. Now, one may wonder which kind of necessity should be involved in this reconstruction. Considering that the *explanandum*, "all observed Fs have been Gs," is confined to the actual world, the metaphysical necessity, which ranges over all possible worlds, would make *the explanatory hypothesis* redundantly strong. Then, as far as natural necessity is considered to be weaker than metaphysical necessity, as it is restricted to the worlds governed by natural or physical laws rather than metaphysical laws, the necessity in line (4) is preferred to be regarded as natural necessity.[7]

This line of argument is taken further by Beebee (2011) who noticed that natural necessity is yet too strong. In her words, "the fact that what calls for explanation is only that the observed *F*s have been *G*s is important, since alternative explanations come into play, aside from the one that postulates timeless necessary connections" (Beebee 2011, 509). Natural necessity, according to her, was the best explanation, if the explanandum had been "all Fs have been Gs." But when only "*so far* observed Fs have been Gs," the best explanation seems not to be a natural

[6] Lowe's *The Four-Category Ontology: A Metaphysical Foundation for Natural Science* (2006) is an instance.

[7] Philosophers who take natural necessity as a sort of metaphysical necessity (Shoemaker 1980; Swoyer 1982; Ellis 1999; and Fine 2002) naturally skip this moderation.

necessity, unless it conjoins with another statement, i.e., "unobserved Fs have been Gs, too." The rationale in Beebee's argument is that a hypothesis may be the best explanation for A&B, without being the best explanation for A or B. Hence, she introduced a moderate version of necessity, namely a "time-limited necessity" (N_t), as a preferred explanation for the inductive premise (1). This notion of necessity, however, would not be a solution to the problem of induction anymore, simply because (5) is not entailed by N_t.

Time-limited necessity is a relation with a time index, according to which the conjunction between F-ness and G-ness necessarily holds until *t*, without holding necessarily thereafter. But how can a relation be genuinely necessary and expires at a particular time? Beebee (2011) and Psillos (2017) attempted to answer this question. But I want to address the question in a different way, since what is involved in N_t (F,G) can also be recast in a more familiar way, by employing a Goodmanian indexed-predicate. Then t (F,G) can be replaced by N (F,G_t), where Gt means "G until t."[8] While the inductive premise is explained by both relations, none of them supports inductive inference, in a way that rules out inductive skepticism. Now, the question is that among the predicates G and G_t, which one is preferred? And as clear, it is the very question raised by the new riddle of induction. Then, the debate on the timeless necessity and time-limited necessity is connected to the debate on ordinary predicates and indexed-predicates. As time-limited necessity leaves Hume's problem unsolved, answering the old problem of induction turns to depend on an answer given to the new problem of induction. Then so far as we have any convincing solution to set grue-like predicates aside, we have likely a solution to prefer N(F,G) over N(F, G_t).

Cases in Favor of the Uniformity

Now, let us go back to the inference (B), to see whether the rationale ascribed to set natural kinds aside is applicable to the necessity itself. In other words, isn't there any weaker metaphysical alternative to support the inference, compared to necessity? As far as the explanation for (1) is concerned, and the explanation is treated as an entailment, one may ask why not remove the line (4), and employ (5) directly as an explanation for (1)? Is not it possible to use (5) as the best explanation for (1) and take it to be true, based on an IBE? A third inference comes with the positive answer to this question:

Inference (c)

[8] Such predicates were introduced by Armstrong (1983) as quasi-universals, but just to be rejected as opposed to genuine universals.

(1) Observed Fs have been Gs.

Therefore

(5) All Fs are Gs.

Obviously, this inference may look premise-circular, as the conclusion is indeed the conjunction of premise (1) and a hypothetical statement: "non-observed Fs are Gs," and as the latter sentence has nothing to do with the explanation of (1), it seems that the burden of explanation is ultimately on the first conjunct, which is at the same time the *explanandum*. Without denying the circularity, it seems to me that it goes indeed back to the IBE structure, not to this particular case, or to the inference itself. Note that an explanation referred to here is not a causal explanation, which offers, for example, a mechanism responsible for Fs' being G. What is meant by the explanation in this inference is a formal explanation, which involves an entailment to the *explanandum,* that obtains in the above inference. Then, this apparent circularity can be ruled out, as far as IBE is recognized as a legitimate rule of inference in cognitive activities.

Adequacy of uniformity as a solution to the problem of induction, however, has been challenged extensively by the necessitists. As Foster (1983) argued, necessity is required when it comes to distinguishing between uniformities by accident and uniformities by necessity. Necessity, in Armstrong's terms (1983), is also an ontological basis which warrants that uniformities observed between the instantiations of Fs and the instantiations of Gs hold in every point in time and space. Furthermore, necessity is fruitful to explain the uniformity observed, as Tooley (1977) illustrated. Nonetheless, there are two cases which show that the justification of induction cannot exclusively depend on the necessity:

First, successful inductions are not confined to natural science. In social science, and also in everyday thinking, we commonly make successful inductive inferences, while necessity is very hard to be supposed in the social realm. As in natural science, there are some law-like statements in social science, which support inductive inferences, while these statements are often claimed to be not necessary. Take, for example, the relation between money supply and inflation, which goes as follows: 'Inflation will happen, if the money supply grows faster than the economic output.' To suppose any kind of necessity in this relation seems not conceivable. Some philosophers who understand laws as something necessary even preferred to call regularities in social science as 'trends', underlining their difference from laws, and saving the latter for the regularities in nature (Little 1986). Theories of social reality, Searle's (1995) for example, also leave no room for any kind of necessity,

metaphysical, natural or conceptual, in the social realm.[9] According to Searle, the institutional reality is constructed through the acts of collective intentionality and is grounded on social conventions, which keeps arbitrariness an essential part of the social phenomena.

Second, the solutions for the problem of induction are not confined to ontological solutions. As it is known, Bayesian theories have been very successful in addressing this problem. Thanks to the conditional probabilities, it can be shown that likeliness of a belief regarding unobserved cases can be updated through the next observations, following such an equation: $P(E'/E) = (E'/H)\ P\ (H/E)$, where E' stands for an unobserved instance, and E for all observed instances. As clear, a Bayesian approach is indifferent to the nature of theories, pieces of evidence, properties, the relation between them, and even the agent who grasps the evidence. No matter whether he is Russell or its turkey, Pavlov or its dog, and whether the context is natural or social. The logic of belief update is in all cases uniform. While connecting necessity to the justification of induction leads to discrimination between inductions in the natural and social sciences, Bayesian approaches treat them uniformly. In other words, any kind of discrimination between inferences based on their ontological basis seems to be untenable in a Bayesian approach. Putting it in terms of natural kinds, one can say that no matter whether a kind is natural or social, or even artificial, what supports an inductive inference is its reliance on the kind-hood, not the naturalness of a kind. Then, social and artificial kinds can be as justificatory to the induction as natural kinds are. In other words, making inductive inferences over screws and iPhones, or over money and police, are as justified as making inductions over horses and woods. The fact that dogs are ancestrally wolves which were evolutionarily manipulated by human beings makes no difference in tenability of inductions over each species. Therefore, to make the justification of induction depends on the necessity will be accompanied by two unpleasant consequences: it is against our observations regarding practicing social science, and it underlines discrimination which is not found in other successful solutions to the problem of induction.

Conclusion

Inductive inferences are made in different contexts, based on different evidence, and by different subjects. In some cases, in natural science for instance, natural kinds are fruitful means to support these inferences. In other cases, some natural laws for

[9] While necessity is ruled out from the social realm by most philosophers, there were also some exceptions, Hegel for example, who found necessary interconnections in the social realm.

example, the natural necessity is a proper ontological idea to support an inductive inference. In some cases, in social science or everyday thinking, nonetheless, necessity is neither an economic nor a tenable idea, to be supposed to support the inductions. There, uniformity of nature is perhaps all we need to support an inductive inference. Then, when it comes to an ontological solution to the problem of induction, a proper strategy is probably to welcome the plurality of ideas helpful to induction, and to take pragmatic considerations into account in order to judge in every single case.[10]

References

Armstrong, D.M. 1983. *What is a Law of Nature?* Cambridge University Press.

Beebee, H. 2011. "Necessary Connections and the Problem of Induction." *Nous* 45: 504–527.

Bird, A. 2018. "The Metaphysics of Natural Kinds." *Synthese* 195 (4): 1397-1426.

BonJour, L. 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Boyd, R. 1999. "Homeostasis, Species and Higher Taxa." In *Species*: *New Interdisciplinary Essays*, edited by Robert A. Wilson, 141–185. Cambridge, MA: Massachusetts Institute of Technology Press.

Carroll, L. 1895. "*What the Tortoise Said to Achilles.'' Mind* 104 (416): 691–693.

Castro, E. 2014. "On Induction: Time-limited Necessity vs. Timeless Necessity." *Teorema* 33: 67–82.

Chakravartty, A. 2007. A *Metaphysics for Scientific Realism*. Cambridge University Press.

Ellis, B. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.

Fine, E. 2002. "The Varieties of Necessity." In *Conceivability and Possibility*, edited by Tamar Szabo Gendler and John Hawthorne, 253-283. Clarendon Press.

Foster, J. 1983. "Induction, Explanation and Natural Necessity." *Proceedings of the Aristotelian Society* 83: 87–101.

Goodman, N.1977. *The Structure of Appearance*. Boston Studies in Philosophy and History of Science. Springer.

Hacking, I. 1991. "A Tradition of Natural Kinds." *Philosophical Studies* 61: 109–26.

Hume, D. 1739/1978. *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, revised by P. H. Nidditch. Oxford: Clarendon Press.

[10] Acknowledgment: I am grateful to Hossein Sheykh Rezaee, José Antonio Díez, and Carl Hoefer for many helpful comments on previous drafts of this paper. This research was funded by the Iranian Institute of Philosophy.

Kornblith, H. 1993. *Inductive Inference and Its Natural Ground*. Cambridge, Mass: MIT Press.

Little, D. 1986. *The Scientific Marx*. University of Minnesota Press.

Lowe, E.J. 2006. *The Four-Category Ontology*: *A Metaphysical Foundation for Natural Science*. Oxford University Press.

Mill, J. S. 1843. *A System of Logic, Ratiocinative and Inductive, Part II*, University of Toronto Press.

Psillos, S. 2017. "Induction and Natural Necessities." *Journal for General Philosophy of Science* 48 (3): 327-340.

Quine, W.V. 1969. "Natural Kinds." In *Ontological Relativity and Other Essays*, New York: Colombia University Press.

Russell, B. 1948. *Human Knowledge: Its Scope and Limits*. London: Routledge.

Sankey, H. 1997. "Induction and Natural Kinds." *Principia* 1: 239–254.

Searle, J. 1995. *The Construction of Social Reality*. New York: The Free Press.

Shoemaker, S. 1980. "Properties, Causation, and Projectability." In *Applications of Inductive Logic*, edited by L. Jonathan Cohen and Mary Hesse, 291–312. Oxford: Oxford University Press.

Swoyer, C. 1982. "The Nature of Causal Laws." *Australasian Journal of Philosophy* 60: 203-23.

Tooley, M. 1977. "The Nature of Laws." *Canadian Journal of Philosophy* 7(4): 667–698.

———. 2011. "The Skeptical Challenges of Hume and Berkeley: Can They Be Answered?" *Proceedings of the American Philosophical Association* 85(2):27–46.

Venn, J. 1876. *The Logic of Chance*. London: Macmillan.

ALSTON, ARISTOTLE, AND EPISTEMIC NORMATIVITY

Benjamin W. McCRAW

ABSTRACT: Alston (2005) argues that there is no such thing as a single concept of epistemic justification. Instead, there is an irreducible plurality of epistemically valuable features of beliefs: 'epistemic desiderata.' I argue that this approach is problematic for meta-epistemological reasons. How, for instance, do we characterize epistemic evaluation and do we do we go about it if there's no theoretical unity to epistemology? Alston's response is to ground all epistemic desiderata, thereby unifying epistemology, in truth and truth-conduciveness. I argue that this move over-unifies epistemology, in effect, giving us a single criterion for epistemology on par with the epistemology-by-justification approach he rejects. Perhaps surprisingly, we find a similar theoretical worry in Aristotle's argument about the science of metaphysics. Aristotle's resolution in this problem by the 'analogy of being' provides a parallel framework to resolve the worries with Alston's approach. In particular, I argue that we can focus epistemic evaluation on the person of epistemic virtue: this category will be focal, unifying the disparate desiderata, without reducing to one thing all epistemic values or relations that desiderate must bear to the central value. A virtue-centric account of epistemic normativity follows: one that can remain genuinely pluralistic and yet unified as well.

KEYWORDS: Alston, Aristotle, justification, epistemic desiderata, virtue epistemology

William P. Alston's *Beyond "Justification": Dimensions of Epistemic Evaluation* attempts to reorient epistemology away from focus on the chimera of a univocal concept of epistemic justification towards a new pluralistic approach to epistemic normativity. In place of justification, Alston offers us his 'epistemic desiderata' approach wherein we should view the positive epistemic status of a belief deriving from any number of irreducible but epistemically valuable features. So, if we wish to analyze the positive epistemic status of some belief, we must look to a plurality of doxastic practices rather than analyze some singular, univocal notion of 'justification' that grounds that belief's status.

In this paper, I sketch an approach to epistemic normativity arising from Alston's theory of epistemic desiderata and problems with it. Namely, I shall suggest a virtue-theoretic account of epistemic normativity which solves problems with Alston's approach: an account inspired by Aristotle. Section One briefly details

Alston's theory of epistemic desiderata as an approach to normativity. From his pluralistic approach to epistemically desirable features, Section Two examines a meta-epistemological problem or worry arising from his pluralism. In particular, I suggest that an indefinite plurality of values undermines the very identity of epistemological inquiry and a lack of unity in terms of epistemic evaluation generates theoretical problems for such evaluation. Section Three turns to Alston's solution in giving truth-conducivity as the central role epistemic evaluation. Section Four argues that Alston's solution will not work. Alston insists that only truth-conducivity grounds epistemic desirability; making his theory ultimately as single-valued and non-pluralistic as the 'justification theory' he attacks. A framework to solve the meta-epistemological problem comes in Section Five. Inspiration comes from Aristotle's problem of the science of metaphysics and the solution in the analogy of being. I argue that a parallel in epistemology can preserve a pluralist approach to epistemic desiderate which avoids the worries Alston's view faces. Finally, in Section Six, I argue that the paradigmatic person of intellectual virtue can explain all of the ways that a belief may be epistemically desirable on Alston's own list. Thus, we can unify these disparate relations into one norm (=the person of virtue) without collapsing them all into each other. And, in the process, we can sketch and defend a virtue-theoretic approach to normativity.

1. Alston's Epistemic Desiderata

Drawing on his decades of work in epistemology, *Beyond "Justification"* crowns Alston's efforts in examining justification, reliabilism, internalism/externalism, and his later doxastic practice approach to epistemology. However, this work aims to remove 'justification' from pride of place in epistemic analysis. He begins with the state of justification theory: citing many examples proposed for the nature of justification over the past several decades of epistemological investigation. Listing the putative analyses of 'justification' gives one a feel for the depth and breadth of the philosophical quarrels over the concept. Alston gives the following on his list: to be justified in believing *p* means:

- One cannot be reproached for being confident that *p*
- Believing that *p* violates no epistemic/doxastic/noetic obligations and/or duties
- Believing/Accepting that *p* when you have good reason to think it true
- Having a right to believe that *p*
- Believing that *p* fulfills one's epistemic responsibility in seeking the truth (and/or avoiding error)
- Believing that *p* is permitted by adequate/correct norms, rules, and/or procedures
- Believing that *p* is evidentially probable
- Believing that *p* is based on something reliable with respect to truth

- Believing that *p* is based on adequate grounds
- Believing that *p* is produced by reliable faculties, properly functioning faculties, and/or epistemic virtues
- Believing that *p* fits one's evidence (2005, 12-15)

The point of Alston's catalogue is obvious: there is no single thread upon which one can tug to unite all of the various theories of evidentialism, reliabilism, coherentism, foundationalism, or any other 'ism' from the massive literature.

When looking at this list reflective of the history of 20th C. analyses of justification, we see no unity or agreement or overall at all. Alston thinks of this plurality as a datum to be explained; offering two sorts of explanations (2005, 21). First, one can accept that there *really is* some univocal concept of justification. In order to make good on this claim, one would need to show that all of the extant theories really converge or reduce to whatever univocal concept or that any theory diverging from this concept is false (regardless as to whether the correct account is extant or to be given). These other false theories are just symptoms of the depth of difficulty in epistemological evaluation. That so many gifted philosophers fail simply shows how hard the concept is to analyze—not that there is no such concept to be analyzed in the first place.

The second explanation denies the univocal nature of 'justification.' Adopting this explanation implies that we see epistemology differently. We shouldn't tilt after epistemological windmills: there just is no such 'thing' as justification that a theory can adequately analyze. Instead, there is an irreducible plurality of things that serve as justifiers—i.e. epistemic values—that no single analysis can capture. This to accept an epistemic desiderata approach to epistemic evaluation. Instead of one single theory of what makes a belief epistemically good, we have a legion of such value-conferrers. Thus:

> [a]ll we have is the plurality of features of belief that are of positive value for the cognitive enterprise. They need no validation from a connection with a supposed master epistemic desideratum picked out by 'justified' (2005, 22).

Given a plurality of epistemic desiderata, a catalogue of failed or incomplete attempts to analyze 'justification' merely serves to show that the attempt is wrong-headed from the very beginning.

But, why should we go with Alston's explanation? That is, why think that the plurality of theories on the list shows that the list is wrong-headed rather than just wrong? How do we select which explanation to adopt? Alston's answer is to begin, if possible, from a theoretically neutral point. Justification-theory is a matter of epistemic normativity: which epistemic norm(s) give rise to justified belief. From normativity of belief, we can move to epistemic value. So, the place to begin in

deciding how to approach epistemology is that "[i[f I am justified in believing that p, my doxastic state is *one that is desirable from an epistemic point of view*" (2005, 23). It is this role played by epistemic value/desirability that gives us our theoretically neutral point. And that's because we can (generally or by-and-large) agree on a major epistemic value: namely, truth. For Alston, *the* epistemic value that we desire is to believe the truth; to have our cognition track reality. How does this incline us towards the epistemic desiderata approach? Each theory catalogued and each analysis on the list belongs there because it picks out something epistemically valuable. Reliability, possession of adequate evidence, based in virtues, produced by proper functioning, doing nothing epistemically irresponsible or irrational, and so on seem to be ways that put one in a better position to believe the truth (or disbelieve what's false). So, by valuing the truth, we can explain why there is an irreducible list in the first place: because there are many ways that something can relate to the truth. And, thus, there are many ways for some doxastic practice or belief forming mechanism to be epistemically valuable. Therefore, the epistemic desiderata approach—grounded in the primary value of truth—better explains the 'irreducible list of justification accounts' datum.

2. The Meta-Epistemological Problem

Alston's argument and epistemic desiderata approach aims to turn epistemology on its ear. And, by attacking the central concept of justification, his position revises the aim of traditional epistemology. But there is a meta-epistemological worry here about just what implications follow from an Alstonian desiderata approach. What are we doing when we do epistemology? With a pluralistic list of irreducible epistemic desiderata, it's unclear how we might go about analyzing knowledge or warrant or your preferred kind of positive epistemic status. In fact, Alston has to use that intentionally and necessarily vague phrase "positive epistemic status" throughout the book. And, further, how do we think about wisdom, understanding, rationality, etc. in ways consistent with Alston's deep and abiding pluralism? We need not assume that there is a set of clear necessary and sufficient conditions for these concepts or anything so clear-cut as that but, rather, my question is higher order. Just what is epistemology or epistemological inquiry? It seems we can get no better than 'investigation into epistemically valuable states/processes/etc.' on Alston's view.

My meta-epistemological worry concerns the twin notions of identity and unity. If we don't investigate something or somethings, then what is epistemology? If all we can manage is an indeterminate list of objects of inquiry ending in only an ellipsis, how can we say what it is to do epistemology? This concern here focuses on

identity—what is epistemic evaluation—and I fear that Alston's desiderata approach gives no possible answer or set of answers. In short, we can construe epistemology as devoted to epistemic evaluation but merely saying *that* tells us nothing substantive about just what we're doing when we do epistemology. This facet of the meta-epistemological worry, then, is that without some determine concept or set of concepts or, better, field of study, it's hard to see any disciplinary unity to the field and, thus, any identity when it comes to epistemologically distinctive philosophical inquiry. This threatens to dissolve epistemology from a theoretical standpoint.

Further, it seems difficult to see how one could reject any feature of a belief as epistemically desirable if we can't manage some standard(s) arising out of our list of desiderata. Why reject wishful thinking, for example, in a principled way from our list of epistemically valuable doxastic features? Intuitively, such thinking lacks positive epistemic status—it just isn't epistemically valuable. But why? If there is no thread (or threads) running through the pluralism of desiderata, how can we explain why we leave this sort of thinking off the list.

3. Alston's Solution

Luckily enough, Alston isn't so de(con)structive himself: he sees the force of his position and tries to clarify 'epistemic evaluation' in terms that will preserve epistemology as a genuine field of inquiry and/or examination. For him, we need to be clearer about the 'epistemic' modifier when we think about epistemology as 'epistemic evaluation.' After his negative, 'there's no such thing as justification' arguments, we find Alston's beginning of his positive epistemological program.

> We evaluate something epistemically…when we judge it to be more or less good or bad from the epistemic point of view, that is, for the attainment of epistemic purposes. And what purposes are those?...I suggest that the primary function of cognition in human life is to acquire true rather than false beliefs about matters that are of interest or importance to us (2005, 30).

Epistemic evaluation (as opposed to moral, political, aesthetic, et al.) centers on the primary value of truth. On this approach, we have a path to give some kind of unity to epistemology as a field of study or philosophical inquiry. Instead of just 'studying epistemically valuable states' or 'philosophical investigation into thinking,' we have truth occupying the prominent role in reorienting epistemology. Though we still have an irreducibly pluralistic list of epistemic values, these desiderata are nevertheless *epistemic* and, therefore, necessarily bound up with evaluation eyed towards truth. Hence, we can avoid the meta-epistemological worries because of the unifying work done by truth: where we lose the sense of 'justification' in epistemology, truth takes up the central axiological role in grounding the set of

desiderata. By looking at Alston's list of desiderata more closely, we can see just how crucial the value of truth becomes from his desiderata approach to epistemic normativity.

Alston's catalogue of epistemic values has five key sections with some subsections. Let's give the list straightaway and see how Alston discusses it. All of these figure as key features that can confer epistemic value on a belief (i.e. are epistemically value with respect to truth).

I. Truth
II. Truth-Conducive Desiderata
    1. Having adequate evidence, reasons, grounds, etc. for the belief in question
    2. Being based on adequate evidence, reasons, grounds, etc.
    3. Being produced by a reliable process
    4. Formed by properly functioning faculties
    5. Formed by epistemic virtue(s)
III. Desiderata dealing with the formation and discrimination of true beliefs
    6. Having higher order access to the evidence (et al.) for the belief in question
    7. Having higher order knowledge or well-grounded belief counting for the belief in question
    8. Being able to defend the (probable) truth of the belief in question
IV. Deontological desiderata
    9. The belief in question is held permissibly
    10. The belief is formed/held responsibly
    11. The formation of the belief includes no violations of intellectual obligations/duties
V. Epistemically valuable features of belief systems
    12. Explanation
    13. Understanding
    14. Coherence
    15. Systematicity (2005, 39-57)

Given the catalogue and the role of truth for epistemic evaluation, we can see Alston's assessment of the list. Clearly, truth (I), must play a central role. But it confers a pride of place to (II) as well. If truth is the primary epistemic value, then those features of beliefs that require truth-conducivity are equally crucial for epistemic evaluation. But that's not the only way that (II) fits in the picture. And that's because both (III) and (V) play their roles as epistemic desiderata by connection to (II). For (III), Alston argues that "they earn the title of [epistemic desiderata] in an indirect way by contributing to S's being in a position to arrange things in a way that is favorable to acquiring truth beliefs rather than false beliefs" (2005, 50). In short, (III) occurs on the list because they, in some fashion, make one more likely to obtain the truth; that is, they are indirectly truth-conducive. Alston has a similar assessment for (V):

> there is a clear connection of these desiderata to truth. It consists of their cognitive desirability depending on their being associated with a favorable balance of truth over falsity in the body of beliefs to which they apply… (2005, 51)

That is, we can see a role for (V) insofar as they, also indirectly, contribute to one's beliefs being generally truth-conducive.

The list of desiderata, when combined with a focus on the value of truth, therefore allows us to link the major sections of the list together.

> What we have seen is that of the desiderata short of truth itself, the directly [truth-conducive] desiderata in Group II are clearly the most basic sense they are most clearly related to true belief itself, by virtue of being always rendering beliefs true or likely to be true. The items in Groups III and V have a more derivative status as desiderata through more indirect connections with true belief or the likelihood of such (2005, 51).

So, (I) is the core or foundational value that (II) directly picks up and, by their indirect relation to (II), Groups (III) and (V) find their place in the catalogue of epistemic desiderata. Alston eschews any *reductive* analysis here but, by linking to truth through (II), he can show how the catalogue displays a kind of unity through truth-conducivity. And that "kind-of" unity is the sort of answer we need to solve the meta-epistemological worry above. We have unity (in a way) without reducing epistemological evaluation to a single, univocal concept like 'justification.' We can have a pluralist, desiderata approach and a coherent epistemological methodology and aim in one theory.

Unfortunately, Alston's solution won't give us the answer we need, and that's because it moves us too close towards the non-pluralist approach against which he sets his desiderata theory. In the next section, I'll digress briefly from Alston to see how a similar problem and solution drives an important theory from Aristotle's metaphysics and turn in the section thereafter to my critique of Alston. In understanding the Aristotelian problem of the science of metaphysics, we have a parallel problem forces Aristotle to a similar solution as we see in Alston. And that will drive both my criticism and my own solution to the meta-epistemological problem that Alton's theory of desiderata threatens.

4. Why Alston's Solution Won't Work

Alston attempts to unify epistemic evaluation via a central focus on truth and truth-conducivity. The unity that grounding all desiderata in (II) provides allows us to make sense of what we do when we do epistemology. Alston praises his desiderata approach for its strikingly pluralist view of epistemic value. And it seems that way: we have a diverse list of epistemically desirable features that can't be reduced to

others. Recall that Section Two gives a list of *five* kinds of epistemic desiderata. And we've seen that (II) prompts (I) via reliability with (III) and (V) indirectly relating to (II). These interrelations with (II) at the center provides the unity we are seeking to solve the meta-epistemological worry from Section Two. But (IV)—deontological desiderata—has been conspicuously absent from our discussion so far. Where do they fit into Alston's theory? That's a bit of a trick question: they don't. The deontological elements of (IV), Alston argues, requires concepts like holding beliefs responsibly, permissibly, or in way that violate no epistemic obligations or duties. This seems to require, via the Kantian 'ought implies can' dictum, that we have control over our beliefs. If I *ought* to believe in certain ways (responsibly, permissibly, …), then believing must be under my control. And yet our beliefs aren't really under our control; they arise in us or occur in us without our volition.

> Thus none of the deontological candidates makes the grade. [Versions of 9] fail through the failure of the version of voluntary control of belief presupposed by each to be a real possibility for human beings. And [11] fails through not being connected with truth in the right way. Thus they will receive no further attention in the development of the [epistemic desiderata] approach to the epistemology of belief (2005, 80).

I skipped (IV) in my earlier discussion because Alston ultimately rejects this group as a genuine set of desiderata. Foley-ian considerations face a similar assessment. For Foley, epistemic rationality (=justification, more or less) connects with what can survive reflective, critical self-scrutiny. Thus, we have no requirement that rationality entail reliability (or truth-conducivity). And it's that final point to which Alston objects:

> [s]ince I take a status of a belief to be an *epistemic* desideratum only if it is desirable from the point of view of the aim of having true rather than false beliefs on matters of importance and/or interest, I do not recognize being justified or epistemically rational in these senses as distinctively *epistemic* desiderata (2005, 93).

Since Foley-ian rationality implies no truth-conducivity, it fails to be a genuine *epistemic* desiderata on Alston's view.

I cite the rejection of (IV) and Foley-ian rationality/justification to make the following point: truth-conducivity is *the* necessary condition for genuine 'epistemic desiderata' status for Alston. If some epistemic feature of a belief (process, web of beliefs …) doesn't directly or indirectly contribute to truth-conducivity or reliability, it has no place in Alston's theory. Thus, Alston's theory isn't all that pluralist or inclusive. This theory works on the notion that it's truth-*conducivity* that really does the epistemic heavy lifting when it comes to epistemic normativity and evaluation. This realization leads towards two points of criticism.

First, (II) seems to doing all of the really significant work in the theory. Alston's theory clearly makes (II) central but his view goes beyond making it central towards making it *the* criterion for status as an epistemic desiderata. Now, clearly the features in (III) and (V) do not *reduce* to truth-conducive features in (II) but those features are *epistemically* valuable (for Alston) only by virtue of their indirect truth-conducivity. Thus, they would seem to belong as subsections of (II). Indirect truth-conducivity is still truth-conducivity, after all.

Because of his insistence on truth-conducivity, it seems as though all genuine epistemic desiderata really fall under one group—Group (II). The result is that his epistemic desiderata theory isn't really pluralist. Given his arguments about non-truth-conducive features, we see that only *one* type of desiderata really counts; namely, those that reliably lead to the truth.

Alston only rejects 'justification' as the single conferrer of positive epistemic status while it provides another single category that plays the same role. Truth-conducivity functions the same in conferring such status as the 'justification' Alston attacks. A genuinely pluralistic approach cannot admit of higher-order unification. In short, Alston falls victim to the very same thing he criticizes of 'justification' theory. Though the properties are pluralist, the *relations* they must bear to one epistemic value (truth) or *what makes them epistemically valuable* admits of no real diversity. All the real work is done by truth and truth-conducivity.

Second, Alston seems to restrict genuine epistemic status to truth-conducivity without really arguing for it. He begins with the plausible claim that having true beliefs and avoiding false ones provides the primary epistemic value upon which to construe epistemic normativity. But the immediate response is not truth *simpliciter* but the reliable attainment of truth. That is, Alston moves from the claim that truth is epistemically valuable to the claim that *only features that successfully attain truth are epistemically desirable*. And that, it seems to me, doesn't follow without more argument. One can accept the value of X without insisting that the only thing desirable is the actual attainment of X. For instance, it seems possible to think that being properly motivated to obtain X can be desirably even if 'being motivated to obtain X' doesn't entail the actual obtainment. An instance of this sort of approach would be James Montmarquet (2000). He defends an 'internalist' approach to epistemic virtue. Like Alston, he thinks that truth is valuable and that epistemology should focus on it as the proper end or aim of our cognition. But what he denies is that we should construe this epistemic teleology in terms of success or reliability. Specifically, he says that 'trying' to get to the truth, in in the right way(s) presumably, is epistemically desirable—it's an *epistemically good* thing to try to get to believe truths (2000, 136).

And this is a crucial point: we can maintain some end as a value conferrer without requiring the *attainment* of that end. Let's return to Group (IV) and Foley-ian rationality. If we can make sense of believing permissibly or responsibly (solving Alston's doxastic involuntarism worries), then would we think of that as aimed at the truth? That seems quite plausible. To believe irresponsibly would usually lead one to believe in ways that would not promote true beliefs. For instance, I take it that wishful thinking/believing is irresponsible. And that would not promote believing the truth. But does responsible believing entail reliability? Probably not. As a matter of contingent fact, responsible believing is probably reliable or truth-conducive, but it would be overly strong to say that responsible believing *entails* that your beliefs are likely to be true. Similarly, would Foley-ian, surviving-self-scrutiny reasoning typically lead to the truth? As with responsible/permissible believing; probably. But does survival of self-scrutiny entail problem truth or reliability? Clearly not. Many scientific beliefs in the distant past, for example, survived self-scrutiny millennia ago and such beliefs turned out to be unreliable. But, nevertheless, self-scrutiny does *aim* at the truth and, to my lights, would be an epistemically good thing even without requiring that such scrutiny be truth-conducive, reliable, or successful.

Take the other side of the epistemological coin. We can't rule out, *a priori*, that intuitively non-valuable doxastic practices, such as wishful thinking, cannot find their way onto the list. The old epistemological standby of the evil demon world suffices here, but we might also appeal to considerations which suggest that certain cognitive heuristics can be both reliable and yet intuitively irrational.[1] Prizing truth-conducivity as *the only* epistemic value-conferrer won't only exclude plausibly valuable doxastic practices like those above but also possibly include irrational yet reliable features, too.

So, we have two sets of objections here. First, we see that Alston doesn't really give the epistemic pluralism he claims. Instead, he effectively replaces 'justification' with truth-conducivity rather than offering a revisionary or revolutionary pluralism about epistemic normativity. At the end of the analysis, there's only one kind of epistemic normativity or value: that of truth-conducivity. And that single category of desiderata is in deep conflict or tension with the avowed pluralism motivating the desiderata approach. Second, Alston unduly focuses on the *attainment* of truth as epistemically valuable. Granting the value of truth, we can maintain other relations besides success, reliability, or truth-conducivity as epistemically desirable: notably aiming at or trying to get to the truth.

[1] See, e.g. Elgin (1988) and Bishop (2000).

5. A Better Solution

What Alston needs is a way to unify a genuine plurality without making all features' value collapse on the single value of truth. What is needed is a plurality, not only of epistemic desiderata, but of the relations that a doxastic practices may bear to the central, focal epistemic value to count as legitimate desiderata. I suggest looking to a source that may seem completely irrelevant to the topic at hand for help: namely, Aristotle's problem of the genus of being and the science of metaphysics.

Here the problem: in the *Metaphysics*, each science—i.e. theoretical field or discipline—must have a distinct species or category of thing that it investigates. Yet, metaphysics—the study of being *qua* being—lacks this: Aristotle argues in *Metaphysics* (B.3) that there is no genus of being. Famously, Aristotle maintains that "being is said in many ways" (Γ.2). But without the unity (univocity) of being and, thus, some discrete category as the subject for metaphysics, it's hard to see how metaphysics can be a genuine science. (This is parallel to my meta-epistemological worry about Alston's desiderata approach in Section 2.) What Aristotle needs is some way to unify 'being' without thinking of reality as one single *thing*.

The famed 'analogy of being' comes into play here. Aristotle accepts that there really is no category of 'being' or 'everything that exists.' But accepting the non-univocity of being doesn't imply that it's equivocal either. He carves out a space between purely synonymous and ambiguous meaning. The meanings of 'being,' he argues, are linked together but not in a way that reduces their meaning to one single thing. His example is health. We call a person, urine, and a diet—for instance—healthy. But is the health of a diet and the health of urine the same? Not at all: what it is for urine to be healthy is not at all what it means for a diet to be healthy. But thought their meanings are distinct, they are related. When urine is healthy it is a symptom of a healthy person and when a diet is healthy it is what promotes health in a person. So, while the health of a diet and urine aren't identical, they are connected by relation to the health of a person. Neither a diet nor urine could be healthy were it not for their relation to the health of a person. 'Healthy,' then is not univocal (because it is 'said in many ways') but it isn't entirely equivocal either (because the meanings are connected). We can call this *via media* between univocal and equivocal usage *analogical* or, following Owen (1960), *focal* meaning. The use is analogical (in the Greek sense of *analogia* as a relation or proportion) because the use of the analogical senses—while distinct—always converge in some sense that is primary.[2] For health, saying that urine and a diet are healthy necessarily focuses or converges on the primary use of 'healthy' as predicated of a person.

[2] Hence the term "focal:" the analogical senses *focus* on some primary meaning.

The category of 'substance' or 'thing' plays the primary or focal role in metaphysics. Everything that is exists by virtue of being a substance or a property of/in a substance. So, while the existence of a property (redness, say) is different from that of a substance (an apple, say), the property exists only because of the substance of which it is a property. The redness of the apple exists only because of the apple—not vice versa. So, while there are many senses of 'being,' they are all connected to and focused upon the being of substances.

We have unity—in the focal category of substance—with legitimate plurality—many different *kinds* of things exist as the subject of metaphysics and these different kinds of things bear different *relations* to that one central category (substance). Aristotle, then, defends a fully genuine (ontological) pluralism that's also sufficiently unified. The unity, via the category of substance, gives metaphysics the theoretical underpinning needed to solve the 'no genus of being' problem.

As I see it, we have the framework for a solution for an epistemic desiderata approach paralleling Aristotle's solution we sketched in the previous paragraph. First, we can agree with both Alston and Aristotle that some category, group, or what have you is primary. Focusing on one epistemic value in this way with make possible that group's unifying role. And that's what we need to solve the meta-epistemological problem. Second, we can avoid the non-pluralism with which Alston ends by accepting several ways that a belief can relate to the focal value as epistemically desirable. That is, we can accept the unifying potential of some central epistemic value but reject the narrow focus of reliable success that Alston insists by recognizing *only* conducivity.

However, this provides only a skeleton or a framework for a solution—not the solution itself. Which epistemic value/desideratum should be parallel to Aristotle's substance—which is the focal value that can appropriately unify epistemology? And how can we go about specifying the pluralist relations to this value in a theoretically appealing way? That is, even given one value (=truth) with many different types of relations won't go far enough in saying just how we think those different relations fit together. An indefinite list of relations ending in an ellipsis does us no better than the same for a list of desiderata.

6. Virtue Theory and Epistemic Desiderata

For the purposes of evaluating Alston's theory, we have seen that a truth-conducive or reliabilist criterion of epistemic normativity is lacking. We find Foley-ian, deontological, responsibility, etc. based accounts of beliefs to be epistemically valuable and, therefore, a strictly reliabilist theory of normativity can't explain the desirability of these features.

An evidentialist approach to epistemic normativity fails in precisely parallel ways. For instance, an evidentialist can't explain the value of Group II (3)-(5): reliable production, formed by properly functioning faculties, and grounded in epistemic virtues, respectively. If it's evidence that confers value, it's hard to see how reliability or proper functioning plays any substantial role. Of course, one may obtain evidence by one's faculties, but the quality of the evidence has to split from the reliability or proper functioning of those faculties lest that evidentialism slide into reliabilism or proper function theory itself.

I want to draw a general conclusion here. Neither reliabilism nor evidentialism can explain intuitive candidates for epistemic desiderata. And that's because they focus on relatively narrow properties of *beliefs* (i.e. being produced by reliable faculties and being based on evidence). However, if we consider Alston's Group (IV)—deontological desiderata—as genuinely epistemically valuable, we can draw more general implications. Group (IV) includes the following:

9. The belief in question is held permissibly
10. The belief is formed/held responsibly
11. The formation of the belief includes no violations of intellectual obligations/ duties

I want to center on (9) and (11). It's hard to really think about these desiderata without considering the crucial role played by the believer in holding these beliefs. (9) says that a belief must be *held* permissibly. Presumably, a belief is held in a certain way only by consequence of the believer believing in a certain way. Regardless as to whether the believing is volitional, non-volitional, indirectly volitional, or what have you, the subject doing the believing in a particular way is absolutely essential to (9). Similarly, (11) makes free use of intellectual obligations or duties in specifying one kind of epistemic value a belief may have. But, as with (9), the duties don't apply to the beliefs themselves but rather to the person doing the believing. In particular, (11) states that the believer must believe in certain way for that belief to have epistemic value. Thus, for (9) and (11) *at least*, there must be some way to account for epistemic normativity capturing the believer's role in coming to her beliefs. A theory of epistemic desiderata must include grounds for epistemic normativity *not* based solely in the properties of beliefs. Instead, there must be some agent-central or belief-central values at work to really explain these desiderata.

Reliabilism and evidentialism, as well as any belief-based theory of epistemic normativity, cannot account for the full, pluralistic range of desiderata we want to consider. At the very minimum, we must include some *agent-centered* elements to the theory. This point is the first step in my own answer: to take the person or agent-

based approach as the basis. Instead of asking whether we can account for these agent-based elements in a belief-based theory, I suggest we should follow the virtue epistemologist's maneuver to ask the opposite question: can we account for the epistemically desirable belief-based features in an agent-based theory? I think the answer is 'yes' and it does so in a way that fits the framework for a solution to the meta-epistemological problem discussed in Section Six.

Virtue epistemology directs epistemology primarily towards the agential features of believing and defines/grounds key epistemological concepts in terms of the properties of the agent. Epistemic properties of beliefs—based on evidence, reliably produced, etc.—are then accounted in terms of the properties of agents. Turri, Alfano, and Greco (2019) call this the "direction of analysis" thesis; virtue epistemology moves from agent-based concepts to belief-based concepts (rather than the non-virtue, traditional movement from beliefs to agents). So, the question becomes: can we explain epistemic normativity—i.e. epistemic desiderata—in terms of properties of agents and, from there, move to the normativity of beliefs? The key to the 'yes' question is to avoid fixing primarily on the virtues of an agent themselves but rather on some paradigmatically intellectually virtuous person and what's true of *them*. Thus, my view takes a (paradigmatically virtuous) *person* as the primary norm and explains other desiderata by relation to that person's thinking, motives, properties, or what have you.[3]

I doubt one could take some set of intellectual virtues, defined either in terms of dispositions, abilities, traits, or whatever, and use them to account for the value of reliably produced beliefs. Evidence is straightforward, I presume, since standard accounts of epistemic virtues will include those that direct the proper obtaining, weighting, maintaining, and using of evidence. Thus, there's a clear role for evidence *qua* desideratum on a virtue theory. But reliabilism is harder to see. Here's where the focus on the person of virtue becomes crucial. And that's because the paradigmatically rational person will most likely have reliable and properly functioning faculties (leaving demon world scenarios aside). Now, we may not *define* virtues in ways that make them reliable or truth conducive,[4] but I take it that the intellectually ideal person of virtue will have such faculties even if those virtues aren't defined by reliabilist criteria.

[3] Compare this to Aristotle's emphasis on the person of virtue (*phronimos*) as a standard for ethics. His definition of 'virtue' in Book II of the *Nichomachean Ethics* explicitly builds the judgment of the *phrominos* into the nature of moral virtue. Both his view and mine, therefore, emphasis the ideally virtuous person as crucial in understand moral and epistemic value, respectively.

[4] See, e.g. Montmarquet (2000), Wedgwood (2020), and Wright (2009; 2010).

Let's go through Alston's sets of desiderata and see what we can say about the paradigmatic person of virtue. Group (II) is Alston's hinge set of desiderata as the truth-conducive set. It includes:

1. Having adequate evidence, reasons, grounds, etc. for the belief in question
2. Being based on adequate evidence, reasons, grounds, etc.
3. Being produced by a reliable process
4. Formed by properly functioning faculties
5. Formed by epistemic virtue(s)

As mentioned above, we find a very plausible role that evidential concerns like (1) and (2) play in a virtue theory. Given the plausible assumption that evidence-based virtues will be among the person of virtue's store, these desiderata are easy to explain. And our previous discussion sets up (3) and (4): the person of virtue will typically have faculties that function reliably/properly. And (5)'s fit on a virtue approach is obvious.

Group (III) desiderata include:

9. Having higher order access to the evidence (et al.) for the belief in question
10. Having higher order knowledge or well-grounded belief counting for the belief in question
11. Being able to defend the (probable) truth of the belief in question

Will the person of (intellectual) virtue be in a position to defend her beliefs and have higher order access/knowledge/justification (when possible)? That strikes me as obvious if we are indeed talking about the *paradigmatically* virtuous agent. Certainly virtues of intellectual curiosity and determination will spur the person of virtue to put herself in the best epistemic position possible with respect to her beliefs. Groups (IV) and (V) provide no serious problem either. They pick out:

9. The belief in question is held permissibly
10. The belief is formed/held responsibly
11. The formation of the belief includes no violations of intellectual obligations/ duties

And

12. Explanation
13. Understanding
14. Coherence
15. Systematicity

Now, granted the approach is aretaic rather than deontological,[5] will the person of virtue violate any epistemic obligations or do anything irresponsible or

[5] *A fortiori*, Linda Zagzebski (1996, 241-242) has defined epistemic duties in terms of epistemic

impermissible? Presumably not. Even if there is an occasion where the virtuous agent doesn't follow some rule, some epistemic analogy of the 'do you steal a loaf of bread to feed a starving family' sort of scenario, this will be atypical. By and large, the person of virtue will act in perfectly permissible and responsible ways and, accordingly, the person of *intellectual* virtue will *believe* in perfectly permissible and responsible ways. Thus, (9)-(11) fit a virtue approach to epistemic normativity nicely.

What of (12)-(15)? Will they characterize the belief systems of the person of intellectual virtue? Again, I think the answer is obviously affirmative. If one has the ideal set of virtues, will one's beliefs explain and be explained; will that person understand a great deal; will that person's beliefs cohere with one another; and will that person's beliefs effect a systematic approach to a range of subjects? If we are considering, again, the paradigmatic person of virtue, I can't imagine answering 'no' to these questions. Of course, typical and fallible agents fail in all sorts of ways but we are considering the ideal virtuous agent.

What of the Foley-ian critical self-scrutiny rejected by Alston? Again, we must ask: does the ideal person of virtue reflect critically on her own beliefs, doxastic practices, commitments, and so forth? Yet again, the intuitive answer seems to be to be 'yes.' I doubt we could or would consider an unreflective, uncritical person as ideally intellectually virtuous. So, we can explain the value of Foley-ian features by their connection to the person of intellectual virtue.

Only one group remains: truth. For Alston, Group (I) hinges the set of desiderata because of its relation to (II). Truth-conducivity has value because it promotes the primary epistemic value of the truth itself. Where does a virtue approach to epistemic normativity place truth? Like Alston, truth plausibly has a central role in epistemic evaluation. We can see this by asking what drives the person of virtue. Following Linda Zagzebski, we can think of the overarching goal of an intellectual virtue as 'cognitive contact with reality'—i.e. truth (1996, 167). Now, I've argued that a truth-conductive approach restricts our desiderata inappropriately, but recall from Section Four that truth can be valuable in non-conducive or success-based ways. Appropriately trying to get the truth or aiming at the truth can be valuable as well. And it's this 'aiming for the truth' that makes an epistemic virtue a virtue and, thus, is the most fundamental characteristic of the person of virtue. Clearly, then, a virtue approach appreciates the value of Group (I) since it encodes the truth in its axiology of virtues (at least, from a Zagzebskian perspective).

---

virtue, so the connection between aretaic and deontological concepts, in epistemology, may be even closer than on the account offered here.

How precisely does this help us solve the meta-epistemology problem from the outset? All of the groups Alston affirms and denies status as desiderata have a place in virtue theory. By using the paradigmatic person of virtue as some sort of standard or criterion, we can explain the epistemically desirable features of both agents and beliefs incorporated in the various desiderata mentioned. Thus, we can accept a pluralism about epistemic normativity deeper than Alston's and yet retain the primary value of truth he also emphasizes. There are many epistemically valuable relations to the truth on this view: reliable success in attaining it, responsible seeking of it, defending one's beliefs in light of it, reflecting on one's cognition with an eye towards it, cultivating traits aimed at it, obtaining evidence for it, and so on. All of these different relations mark out different features holding of the person of intellectual virtue; her function, character, etc. unifies these epistemically desirable relations in various ways. Where Alston slips is in thinking that only one of these relations *really* makes something epistemically valuable. My view accepts all of these relations as genuinely desirable and, at the same time, shows that they all hold for the person of virtue without requiring *one single way* they must relate to this focal value.

My solution, therefore, keeps with Aristotle's use of substance in the *Metaphysics*. I emphasize the crucial role of the person of virtue as a way to explain the value of a pluralistic set of relations that are mutually irreducible and not inter-definable. The person of virtue, thus, unifies the diverse epistemic desiderata but she does not do so in a way that makes all of these desiderata ultimately collapse into one category: we have a genuine pluralism that's also robustly unified. And that sort of theory fits exactly on the framework developed in Section Six.

7. Conclusion

Given my arguments and position in this paper, we can find an illuminating role for virtue theory on epistemic normativity. Leaving aside particular worries about Alston or uses of Aristotle, I that that point as the major upswing of my theory. The hopelessly vague 'aiming that the truth' works through the more perspicuous person of intellectual virtue as a way to explain the seemingly disparate and unrelated sorts of epistemic values one might consider. However, these values do have a relation to each other insofar as they each relate to the person of virtue. Ultimately, then, the value of truth filters through the person of virtue as some sort of norm to ground epistemic value and theories of normativity. But the normativity of the ideally virtuous person doesn't collapse the other values into each other: truth-conducivity really is epistemically good just as evidence, critical self-reflection, responsible believing, and so on. Accordingly we have a view relating epistemic values to

epistemic norms to epistemic desiderata in a way that allows for unity with a robust plurality. Alston had the right sort of view about where we should be heading even if his particular route won't get us to that destination.

References

Alston, W. P. 2005. *Beyond "Justification": Dimensions of Epistemic Evaluation*. Ithaca: Cornell University Press.

Owen, G. E. L. 1960. "Logic and Metaphysics in Some Early Works of Aristotle." In *Aristotle and Plato in the Mid-Fourth Century*, edited by Ingemar Düring, 163-190. Göteborg: Humanities Press.

Montmarquet, J. 2000. "An 'Internalist' Conception of Epistemic Virtue." In *Knowledge, Belief, and Character: Readings in Virtue Epistemology*. edited by G. Axtell, 135-148. Lanham: Rowman & Littlefield.

Turri, J., M. Alfano, and J. Greco. 2019. "Virtue Epistemology." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, URL = https://plato.stanford.edu/archives/fall2019/entries/epistemology-virtue/.

Wedgwood, R. 2020. "The Internalist Virtue Theory of Knowledge." *Synthese* 197(12): 5357-5378.

Wright, S. 2009. "The Proper Structure of the Intellectual Virtues." *Southern Journal of Philosophy* 47(1): 91-112.

Wright. S. 2010. "Internalist Virtues and Knowledge." *Acta Analytica* 25(2): 119-132.

Zagzebski, L. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

DEONTIC LOGIC, WEAKENING AND DECISIONS CONCERNING DISJUNCTIVE OBLIGATIONS

Michael J. SHAFFER

ABSTRACT: This paper introduces two new paradoxes for standard deontic logic (SDL). They are importantly related to, but distinct from Ross' paradox. These two new paradoxes for SDL are *the simple weakening paradox* and *the complex weakening paradox*. Both of these paradoxes arise in virtue of the underlying logic of SDL and are consequences of the fact that SDL incorporates the principle known as weakening. These two paradoxes then show that SDL has counter-intuitive implications related to disjunctive obligations that arise in virtue of deontic weakening and in virtue of decisions concerning how to discharge such disjunctive obligations. The main result here is then that theorem T1 is a problematic component of SDL that needs to be addressed.

KEYWORDS: deontic logic, weakening, disjunction, decision theory, obligation

1. Introduction

SDL is the standard system of deontic logic that was developed importantly by von Wright and others in the 20th century.[1] The typical axiomatization of SDL is as follows:

A1. All well-formed tautologies.

A2. $\mathbf{O}(p \rightarrow q) \rightarrow (\mathbf{O}p \rightarrow \mathbf{O}q)$.

A3. $\mathbf{O}p \rightarrow \neg\mathbf{O}\neg p$.

R1. If p and $p \rightarrow q$, then q.

R2. If p, then $\mathbf{O}p$.[2]

So, SDL is just the modal logic KD with the operator $\mathbf{O}$ interpreted as "obligation" and $\mathbf{P}$ as "permission" such that $\mathbf{O}p$ is equivalent to $\neg\mathbf{P}\neg p$. From these axioms the following theorem importantly follows:

T1. $\mathbf{O}p \rightarrow \mathbf{O}(p \vee q)$.

[1] See von Wright 1951; McNamara 2019; and Hilpinen 2001.

[2] See Hilpinen 2001 and McNamara 2019.

Notice that T1 is a form of weakening that is embedded in SDL. Let us then begin to consider some important and problematic implications of SDL that arises in virtue of T1.

Suppose it is true that $\mathbf{O}p$. By T1 it follows that $\mathbf{O}(p \vee q)$. But note that this holds for *any and every* q_i whatever and it is reasonable to suppose that there are an infinite number of such alternatives or at least vastly more such alternatives than any agent could ever consider or even be aware of.[3] This may not seem to be immediately or obviously problematic, but it indeed is.[4] As we shall soon see, this is manifestly the case with respect to our intuitive and quite commonplace notion of disjunctive obligation and in the context of *decisions* with respect to the satisfaction of disjunctive obligations.

Let us suppose that a particular agent has the obligation to p and by T1 this implies that the agent also then has the obligation to do p or one of the large or infinite number of alternatives to p. For example, suppose Jane has an obligation to meet her friend Bonny for lunch on Tuesday. From this it follows that Jane has a disjunctive obligation to meet her friend Bonny for lunch on Tuesday or to meet her friend Bonny for lunch on Wednesday or to meet her friend Marge on Tuesday, etc., etc. In such a case it is reasonable to assume that such an agent can satisfy this sort of derived disjunctive obligation by doing p or any one of the alternatives to p and that the agent is *not* obligated to do each of the alternatives. But, this is at odds with SDL. Moreover, as we will see satisfying disjunctive obligations ultimately requires *choosing* to do p or one of the alternatives to p. However, given T1, it turns out that it is impossible to do this rationally in the context of orthodox decision theory and the rational obligation it imposes on us concerning choices. So from $\mathbf{O}p$, T1 and the framework of orthodox decision theory it will be shown that we can derive the perplexingly absurd conclusion that it is impossible to have the obligation to p by a sort of reductio (at least provided we maintain orthodox decision theory). Let us examine these lines of argumentation in more detail in what follows. First, let us begin by introducing Ross' paradox, as it importantly presages the new (but related) criticisms of SDL introduced here.

---

[3] This is simply a consequence of the fact that natural languages and even reasonably realistic artificial languages are constituted by a potentially infinite set of sentences. This is simply a result of the compositionality of such languages.

[4] The problems alluded to here are related to but importantly *different* from Ross' paradox, as originally presented in Ross 1941 and Ross' free choice permission paradox from the same paper. Ross' paradox has nothing directly to do with iterated applications of T1 and the additional problems this raises.

2. Ross' Paradox

In 1941 Alf Ross introduced a seminal problem for SDL. This has come to be known as Ross' Paradox.[5] The paradox arises in virtue of SDL and it crucially involves T1. The problem arises from pairs of claims like these:[6]

(M1) It is obligatory that the letter is mailed.

(M2) It is obligatory that the letter is mailed or the letter is burned.

The formal SDL analogs of these claims are, respectively:

(M3) **O**p

(M4) **O**(p ∨ q)

By T1, M4 follows from M3 and it would seem to be the case that the obligation in M4 can be satisfied *by burning the letter*. But this is clearly unacceptable. From one's having an obligation to mail the letter it should not follow that such an obligation can be fulfilled by burning the letter, especially given the reasonable supposition that burning the letter is (by assumption) forbidden and the fact that burning the letter would make it impossible to mail it. Something is clearly then wrong with SDL in virtue of its incorporation of T1. Ross' paradox crucially involves T1 and thus it is clearly the case that this problem results from the general idea of logical weakening (which is a familiar and seemingly unproblematic principle of propositional logic). As we shall see, however, the sort of *deontic weakening* that gives rise to Ross' paradox also gives rise to additional problems that are importantly related to, though distinct from, Ross' paradox.

3. The Simple Weakening Paradox

Suppose that an agent has an obligation to do p and that SDL is correct. By T1 this entails that the agent in question has a derived obligation to do p or q. However, in such a case it does not appear to be true that one must do all of the disjunctive actions and it does not appear to be true that one must do any one of the disjunctive options in particular. Crucially, this is supported by what Alan Donagan tells us about disjunctive obligations when he explains that: "…from the fact that I have a duty to save either a or b, it does not follow that I have a duty to save a and a duty to save b" (Donagan 2014). But it does seem to be the case that one must do something in such cases of disjunctive obligation. The intuitive idea is then that one can discharge such an obligation by doing either action. However, an agent facing such an

[5] See Ross 1941.

[6] See Ross 1941, 62 and McNamara 2019.

obligation is neither obligated to do both of the options nor is she specifically obligated to do the one and specifically obligated to do the other. This is all seemingly correct and, on this intuitive basis, the following principle concerning disjunctive obligation appears *prima facie* to be correct:

P1. $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p \,\&\, \neg\mathbf{O}q \,\&\, \neg\mathbf{O}(p \,\&\, q)$.[7]

Again, this is just the intuitively grounded claim that the obligation to do p or q entails that it is not obligatory to do p, that it is not obligatory to do q, and that it is not obligatory to do both p and q. So, it would seem to be true that in the case of a disjunctive obligation an agent can discharge such an obligation by doing any one of the disjunctive actions and one need do no more than one of the two options in order to do so. But, this generates an immediate contradiction. We have derived $\neg\mathbf{O}p$ from $\mathbf{O}p$ as $\neg\mathbf{O}p$ is a conjunct of the consequent of P1 that follows from $\mathbf{O}p$ and T1. Let us call this the simple weakening paradox.

3.1 The Complex Weakening Paradox

So, something must go here and one obvious and natural suggestion is that P1 is the culprit. So, despite its intuitive plausibility, one might be tempted to deny P1 and replace it with the following principle:

P2. $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}(p \,\&\, q)$.

This alternative–and rather less intuitive–stance concerning disjunctive obligation defuses the simple weakening paradox with respect to SDL by simply omitting the implications of the formulas $\neg\mathbf{O}p$ and $\neg\mathbf{O}q$ that appear in P1.[8] P2 is just the claim

[7] One might initially think that this principle is overly strong, but the argument given in this section will work using any related principle such that $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p$. In ordinary usage, it seems clearly to be the case that disjunctive obligations very often, or perhaps even always, do not entail specific obligations with respect to their disjuncts. For example, one's being obligated to either go to Paris or London does not seem to imply one's being obligated to go to Paris. Similarly, one's being obligated to donate $10 either to charity 1 or to charity 2 does not seem to entail that one is obligated to give $10 to charity 1 and that one is obligated to give $10 to charity 2. On the basis of such examples and a plethora of ones like them it is difficult really to see how one might reject this, but this possibility will be explored subsequently nevertheless. That point aside, that there are any plausible examples where $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p$ holds means that SDL implies a host of contradictions (for there are surely many such cases) and this also suggests the problem raised here is not merely some artifact of the SDL formalism. Rather, it is a problem that arises in virtue of what appears to be the inadequacy of the SDL analysis/explication of obligation.

[8] Consider what the denial of P1 and the affirmation of P2 involves. If $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}(p \,\&\, q)$ but it is neither the case that $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p$ nor is it the case that $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}q$, then it follows that $\mathbf{O}(p \vee q) \rightarrow \mathbf{O}p \,\&\, \mathbf{O}q$. This seems totally at odds with out intuitive notion of disjunctive

that the obligation to do p or q entails only that it is not obligatory to do both p and q. So, it still seems to be the case that when faced with a disjunctive obligation an agent could discharge such an obligation by doing one of the disjunctive actions, but the agent is not obligated to do both. Since, given this line of thinking, P1 is being denied and P2 affirmed it is not the case that $\mathbf{O}p$ implies $\neg\mathbf{O}p$ and the simple weakening paradox is eliminated. Thus it would appear to be the case that satisfying an obligation to p could be accomplished as per SDL without immediately falling prey to the simple weakening paradox by rejecting P1 and adopting P2 instead of P1.

However, this response to the simple weakening paradox does not capture Donagan's insight at all, it is rather artificial and this solution still leaves us with a related and considerably more difficult problem related to T1. First, it should be clear that rejecting P1 and replacing P1 with P2 does not capture the insight that having an obligation to do p or q seems to imply the lack of an obligation to p and the lack of an obligation to q.[9] It is also an entirely arbitrary solution in that the replacement of P1 by P2 is solely motivated here by paradox avoidance at the expense of the intuitively well-grounded Donagan insight. More importantly, even if we ignore these charges, an additional problem arises when we recognize that if one's legitimate obligation to p obeys the deontic weakening principle T1 and P2 but not P1 applies to disjunctive obligations, then for any p the obligation to p entails an additional obligation to do one of an infinite or very large disjunction of possible options. In infinitary contexts (i.e. those where there are infinite possible alternatives to p) having an obligation to p, by weakening, implies having an obligation to do one of p or an infinite number of alternatives to p. In realistic finitary contexts (i.e. those where there are merely finite but very many possible options to p) having an obligation to p, by weakening, implies having an obligation to do one of p or one of the very large number of possible alternatives to p, many, or even most, of which will be unknown to the agent. In both cases P2 implies that it is not the case that one must do all of the disjunctive options. So, one must be able to satisfy such an obligation in principle by either doing one of an infinite number

---

obligation as evinced by Donagan's quote referenced earlier and as discussed in the previous footnote. One may be obligated to do p or q without being at all obligated to specifically do p and without being specifically obligated to do q. This may just indicate that SDL runs together more than one distinct notion of obligation. It has also been suggested that in modal contexts disjunction and conjunction are not used in the same manner as they are in non-modal contexts and this may be related to some of the problems raised here. See, for example, Geurts 2005 and Zimmermann 2000.

[9] Again, what this may suggest is that there are two sense of obligation at work in standard deontic logic.

of alternatives or by doing one of a very large number of alternatives many of which will be unknown to the agent. But, as we shall see, neither of these things can be done rationally.

All of this essentially means that agents have bona fide *options* with respect to satisfying disjunctive obligations–at least provided the options are individually satisfiable in the sense that the agent can do each of them–and, as per T1, every non-disjunctive obligation implies disjunctive obligations. What is curious then is that satisfying any disjunctive obligation is then a matter of choice.[10] This broadly comports with orthodox decision theory and, according to orthodox decision theory, an agent faces a choice when she has options that are within her power to perform upon choosing. As Levi points out,

> Having a choice presupposes having options. Having the option to perform some action entails having the ability to perform the action upon choosing it. Hence, having a choice presupposes having abilities to perform various actions upon choosing them (Levi 1986, 47).

Let us suppose then that the ability to perform each option from among the options facing our agent who is disjunctively obligated is within that agent's power. It follows then that the agent with any such disjunctive obligation must, as a result of this obligation, *choose* to do one of $(p \vee q_1 \vee q_2 \vee q_3 \vee \ldots \vee q_i)$ and importantly this holds even where the set of i alternatives to p is infinite or large and includes many options unknown to the agent. To discharge a disjunctive obligation *rationally* then, one must be able to rationally choose from amongst the options. In virtue of this requirement one must, of course, be able to identify the option that is the rational option. According to orthodox decision theory the rational decision is the one that maximizes expected utility with respect to the options an agent faces (i.e. one should rationally choose the option with the greatest utility value). In other words, the action that one *ought to do* is the one that maximizes expected utility. From all of this it follows that in such a situation it must be possible for the agent to choose to do one of p or the alternatives to p, one ought to decide by applying MEU and one's choice is rational just in case it actually maximizes expected utility. In the infinitary case this means then that if the agent is to rationally discharge the obligation to p or q–where there are an infinite number of q-alternatives to p–the agent must be able to efficaciously choose from amongst p and the infinite number of alternatives to p

[10] The relationship between deontic logic and decision theory, or course, involves controversial and difficult issues. However, the important and close relationship between the two types of formalisms has been explored in various ways and in terms of various conceptions of deontic logic in, for example, Hansson 1997; Torre and Tan 1997; Fusco 2015; Dietrich and List 2017; and Cariani forthcoming.

and the agent must choose the utility maximizing option from that set of possible options.

But this is impossible in the context of orthodox decision theory and the theory of preference on which it is based. Standard utility theory is based on the idea that if an agent's preferences obey the axioms of preference theory, then they can be represented as a utility function that exhibits certain supposedly desirable algebraic features. These axioms are introduced on the basis of their supposed intuitive (i.e. a priori) plausibility. Let "$x \preccurlyeq y$" mean "x is weakly preferred to y", "$x \prec y$" mean "x is strictly preferred to y" (i.e. x is weakly preferable to y but x is not indifferent relative to y) and "$x \backsim y$" mean "x is indifferent relative to y" (i.e. x is weakly preferred to y and y is weakly preferred to x). Let O_i, O_j and O_k represent distinct outcomes and **p**, **q**, **r**,… represent distinct probability values. Finally, let $u(O_i)$ be a function representing a real numbered valuation of O_i. Given these basic representations we can then represent a gamble with a probability p of winning O_1 and a probability q of winning O_2 as [**p**O_1, (1 – **p**) O_2]. In terms of these representations, the axioms are used to characterize what is intuitively taken to be rational preference orderings are as follows.[11] First we have the ordering axiom:

(U1) The preference relation $\succcurlyeq$ is a total ordering that is reflexive and transitive.

Second, we have the better prizes axiom:

(U2) For a fixed probability, prefer the gamble with a greater prize.

Third, we have the better chances axiom:

(U3) For a fixed prize prefer the gamble with a greater probability.

Fourth, we have the reduction of compound gambles axiom:

(U4) Compound gambles are to be evaluated in terms of the probability calculus.

Finally, we have the Archimedean or Continuity axiom:

(U5) For any outcome that is ranked between two others there is a gamble between the more preferred and less preferred outcomes such that the agent is indifferent between it and the outcome ranked in between the more preferred and less preferred outcomes.

Formally, in terms of $\preccurlyeq$ these axioms can be presented as follows:

(U1.0) For any O_i and O_j either $O_i \preccurlyeq O_j$ or $O_j \preccurlyeq O_i$,

[11] This is the standard presentation of this representation theorem and it closely follows Bartha 2007. See Resnik 1987 and Gaus 2008 as well.

(U1.1) For any O_i, $O_i \preccurlyeq O_i$.

(U1.2) For any O_i, O_j and O_k, if $O_i \preccurlyeq O_j$ and $O_j \preccurlyeq O_k$, then $O_i \preccurlyeq O_k$.

(U2) $O_i \preccurlyeq O_j$, iff, for any $0 \leq \mathbf{p} \leq 1$ and any O_k, $[\mathbf{p}O_k, (1 - \mathbf{p})O_i] \preccurlyeq [\mathbf{p}O_k, (1 - \mathbf{p})O_j]$ and $[\mathbf{p}O_i, (1 - \mathbf{p})O_k] \preccurlyeq [\mathbf{p}O_j, (1 - \mathbf{p})O_k]$.

(U3) If $O_i \preccurlyeq O_j$, then for any $0 \leq \mathbf{p}, \mathbf{q} \leq 1$, $\mathbf{p} \geq \mathbf{q}$ iff $[\mathbf{p}O_i, (1 - \mathbf{p})O_j] \preccurlyeq [\mathbf{q}O_i, (1 - \mathbf{q})O_j]$

(U4) For any O_i and O_j and $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$ such that $0 \leq \mathbf{p}$, $\mathbf{q}$, $\mathbf{r} \leq 1$, $[\mathbf{p}[\mathbf{q}O_i, (1 - \mathbf{q})O_j], (1 - \mathbf{p})[rO_i, (1 - \mathbf{r})O_j]] \backsim [\mathbf{r}O_i, (1 - \mathbf{t})O_j]$ for $\mathbf{t} = \mathbf{pq} + (1 - \mathbf{p})\mathbf{r}$.

(U5) If $O_i \preccurlyeq O_j$ and $O_j \preccurlyeq O_k$, then there is a $\mathbf{p}$ such that $0 \leq \mathbf{p} \leq 1$ and $O_j \backsim [\mathbf{p}O_i, (1 - \mathbf{p})O_k]$.

If an agent's preferences satisfy these axioms then those preferences can be represented by a real valued utility function $u(O_i)$ obeying the following two important conditions:

(C1) $O_i \preccurlyeq O_j$ iff $u(O_i) \leq u(O_j)$.

(C2) $u([\mathbf{p}O_i, (1 - \mathbf{p})O_j]) = \mathbf{p}u(Oi) + (1 - \mathbf{p})u(O_j)$.

The Expected Utility Theorem, the core idea behind utility theory, is then simply this claim that if one's preferences satisfy U1-U5, then those preferences can be represented as a real valued utility function satisfying C1 and C2.[12] In other words, formal utilities are a real-valued measure of preference and the value V_i of an outcome O_i is just $u(O_i)$ and according to the principle of maximizing expected utility (MEU)–the fundamental principle of rational decision making–one should rationally opt for the option with the highest utility value.

This second problem related to T1 appears to be that in SDL having any obligation to p implies having an obligation to do one of p or an infinite number of alternatives. This, in turn and in the context of orthodox decision theory, implies that one must be able to rationally make choices with respect to infinite sets of alternatives and according to orthodox decision theory we are obligated to maximize expected utility in such situations. But, this is impossible because one cannot rationally assign utilities to the outcomes of an infinite set of options. This is simply because one cannot have coherent decision-theoretic preferences with respect to an infinite set of outcomes that could be measured in terms of such utilities. This, in turn, is the case because one cannot have preferences with respect to an infinite set of outcomes. So, we have a deeply problematic paradox arising from the principles

[12] This is just the standard way of introducing utility theory via a representation theorem. This approach takes it as given a priori that U1-U5 are true. Recently, this approach to legitimizing decision theory has been challenged in Easwaran 2014 and by Meacham and Weisberg 2011.

of SDL, P2 and orthodox decision theory. Moreover, replacing P1 with P2 in order to avoid the simple weakening paradox that appeared to afflict SDL and P1 does nothing to help with this more complex paradox. Finally, even if we could circumscribe the set of alternatives to p in such cases and render the set of q-alternatives to p finite (say by arbitrarily assuming a finitary language involving a finite but very large set of alternatives for each p), the obligation to do p or one of a finite but very large number of q-alternatives to p–very many of which will be unknown to the agent–is impossible. Again, one cannot have coherent decision-theoretical utility assignments with respect to the outcomes of unknown alternatives. This is because an agent cannot have preferences with respect to such outcomes. So, this is not a viable response to the complex weakening paradox. One cannot maximize expected utility with respect to a choice involving unknown alternatives with unknown outcomes because one cannot have preferences with respect to unknown outcomes. So, we have our reductio of SDL. What should we make of all this? The most reasonable response seems to be that T1 is an objectionable principle. But, rejecting T1 requires revising SDL and it remains to be seen what such a deontic logic would look like.

4. Conclusion

SDL is a logical system based on modal logic that is supposed to capture the logic of permission and obligation, as well as other related concepts. So deontic logic is just the study of the logical properties of these notions. But SDL is limited in that it fails to capture at least some of our commonsense notions of these concepts. The catalog of deontic paradoxes that has been amassed since the inception of deontic logic shows this. Here two new paradoxes for SDL were introduced: the simple weakening paradox and the complex weakening paradox. Both paradoxes arise in virtue of the underlaying logic of SDL and are consequences of the fact that SDL incorporates the principle of weakening. These two paradoxes show that SDL has counter-intuitive implications related to disjunctive obligations that arise in virtue of deontic weakening and in virtue of decisions concerning how to discharge such disjunctive obligations. The main result here is then that theorem T1 is a problematic component of SDL that needs to be addressed.

References:

Bartha, P. 2007. "Taking Stock of Infinite Value: Pascal's Wager and Relative Utilities." *Synthese* 154: 5-52.

Cariani, F. Forthcoming. "Deontic Modals and Probabilities: One Theory to Rule Them All?" In *Deontic Modality*, edited by N. Charlow and M. Chrisman. Oxford University Press.

Dietrich, F. and C. List. 2017. "Choice-theoretic Deontic Logic." Manuscript: http://www.franzdietrich.net/Papers/DietrichList-ChoiceTheoreticDeonticLogic.pdf.

Donagan, A. 2014. "Consistency in Rationalist Moral Systems." *The Journal of Philosophy* 81: 291-309.

Easwaran, K. 2014. "Decision Theory without Representation Theorems." *Philosophers' Imprint* 14: 1-30.

Fusco, M. 2015. "Deontic Modality and the Semantics of Choice." *Philosophers' Imprint* 28: 1-27.

Gaus, G. 2008. *On Philosophy, Politics and Economics*. Belmont: Wadsworth.

Geurts, B. 2005. "Entertaining Alternatives: Disjunctions as Modals." *Natural Language Semantics* 13: 383–410.

Hansson, S. 1997. "Situationist Deontic Logic." *Journal of Philosophical Logic* 26: 423-448.

Hilpinen, R. 2001. "Deontic Logic." In *The Blackwell Guide to Philosophical Logic*, edited by L. Goble, 159-182. Malden: Blackwell Publishers.

Levi, I. 1986. *Hard Choices*. New York: Cambridge University Press.

Meacham, C. and J. Weisberg. 2011. "Representation Theorems and the Foundations of Decision Theory." *Australasian Journal of Philosophy* 89: 641-663.

McNamara, P. 2019. "Deontic Logic." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. URL = https://plato.stanford.edu/archives/fall2018/entries/logic-deontic/.

Resnik, M. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis: University of Minneapolis Press.

Ross, A, 1941. "Imperatives and Logic." *Theoria* 7: 53–71.

Torre, L. and Y. Tan, "Prohairetic Deontic Logic and Qualitative Decision Theory." *AAAI 1997*: 1-9.

von Wright, G. H. 1951. "Deontic Logic." *Mind* 60: 1–15.

Zimmermann, T. 2000. "Free Choice Disjunction and Epistemic Possibility." *Natural Language Semantics* 8: 255–290.

NOTES ON THE CONTRIBUTORS

Cătălin Bărboianu is mathematician (MSc) and philosopher of science (PhD), author of over 30 academic publications. He is independent researcher with research interests in philosophy of mathematics and of science, formal epistemology, philosophy of language, philosophy of education, and philosophy of mind. Contact: cb@infarom.ro.

Domingos Faria holds a FCT junior researcher position (CEECIND/01066/2017) at the Center of Philosophy of the University of Lisbon (CFUL), where he develops a project entitled "The Epistemology of Complex Disagreement Among Social Groups" at the LanCog group. Affiliation: LanCog, Centre of Philosophy, University of Lisbon. Contact: domingosfaria@campus.ul.pt.

Mario Günther is assistant professor at the Munich Center for Mathematical Philosophy, LMU Munich. His areas of specialisation include Epistemology, Logic, and Metaphysics. He is the author of several articles among which: "Learning Conditional Information by Jeffrey Imaging on Stalnaker Conditionals" (*Journal of Philosophical Logic*, 2018), "Difference-Making Causation" (with Holger Andreas, *The Journal of Philosophy*, 2021), and "Ramsey's Conditionals" (with Caterina Sisti, *Synthese*, 2022). More information is available on his website www.mario-guenther.com. Contact: Mario.Guenther@lmu.de.

Mohammad Mahdi Hatef is a research collaborator at the Iranian Institute of Philosophy, where he received his Ph.D. in Philosophy of Science and worked as a Post-doc researcher. He also made a research stay at the LOGOS Research Group, University of Barcelona, in 2018. His research interests include metaphysics of science, social ontology, philosophy of medicine, and science & politics. He has published another recent work entitled "Revisiting Hull's evolutionary model of science" (*Balkan Journal of Philosophy*, 2021). Contact: hatef.mehdi@gmail.com.

Benjamin W. McCraw is instructor of philosophy at the University of South Carolina Upstate. His main research interests are within epistemology and philosophy of religion. He has published papers in *Acta Analytica, Philosophy and Theology, Social Epistemology,* and the *International Journal for Philosophy of Religion* as well as edited collections with *Palgrave Macmillan, Routledge,* and *Lexington Books*. Contact: mccrawb@uscupstate.edu.

Michael J. Shaffer is a professor at Gustavus Adolphus College and an external member of the Munich Center for Mathematical Philosophy. He is also a fellow of the center for formal epistemology at Carnegie-Mellon University, a fellow of the Rotman Institute for Science and Values at the University of Western Ontario, a Lakatos fellow at the London School of Economics, a fellow of the University of Cologne's summer institute for epistemology and an NEH fellow at the University of Utah. His main areas of research interest are in epistemology, logic and the philosophy of science, and he has published more than fifty articles and book chapters on various topics in these areas. He is co-editor of *What Place for the A Priori?* (Open Court, 2011) and is the author of *Counterfactuals and Scientific Realism* (Palgrave-MacMillan, 2012), *Quasi-factive Belief and Knowledge-like States* (Lexington, forthcoming) and *The Experimental Turn and the Methods of Philosophy* (Routledge, forthcoming). Contact: shaffermphil@hotmail.com.

NOTES TO CONTRIBUTORS

1. Accepted Submissions

The journal accepts for publication articles, discussion notes and book reviews.

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

There are no submission fees or page charges for our journal.

2. Publication Ethics

The journal accepts for publication papers submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The submitted papers should be the author's own work. All (and only) persons who have a reasonable claim to authorship must be named as co-authors.

The papers suspected of plagiarism, self-plagiarism, redundant publications, unwarranted ('honorary') authorship, unwarranted citations, omitting relevant citations, citing sources that were not read, participation in citation groups (and/or other forms of scholarly misconduct) or the papers containing racist and sexist (or any other kind of offensive, abusive, defamatory, obscene or fraudulent) opinions will be rejected. The authors will be informed about the reasons of the rejection. The editors of *Logos & Episteme* reserve the right to take any other legitimate sanctions against the authors proven of scholarly misconduct (such as refusing all future submissions belonging to these authors).

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*rtf) or Microsoft Word document (*docx) and must be double-spaced, including quotes and footnotes, in 12

point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author`s affiliation and professional postal and email address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles which pass the editorial review, with the exception of articles from the invited contributors, will be subject to a strict double anonymous-review process. Therefore the authors should avoid in their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it.

The submissions will be sent to at least two reviewers recognized as specialists in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

The review process is intended to be as quick as possible and to take no more than three months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors.

The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.

The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations.

11. Checking Proofs

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice şi Sociale „Gh. Zane" Academia Română, Filiala Iaşi, Str. Teodor Codrescu, Nr. 2, 700481, Iaşi, România. The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The

editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Copyright & Publishing Rights

No manuscripts will be returned to their authors. The journal does not pay royalties.

14. Electronic Archives

The journal is archived on the Romanian Academy, Iasi Branch web page. The electronic archives of *Logos & Episteme are* also freely available on Philosophy Documentation Center web page.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears at the end of March, June, September, and December. Its fundamental mission is to support philosophical research on human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal publishes articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal promotes all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the Continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

Logos & Episteme is published and financed by the "Gheorghe Zane" Institute for Economic and Social Research of The Romanian Academy, Iasi Branch. The publication is free of any fees or charges.

For further information, please see the Notes to Contributors.

Contact: logosandepisteme@yahoo.com.