EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDK 101 | ISSN 1845-8475

Publisher, Editorial office
University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia
Phone: +385 51 265 794
Fax: +385 51 265 799
E-mail: eujap@ffri.hr

Web address
http://www.ffri.unuri.hr/phil/casopis/index.html

Printed by Impress, Opatija (Croatia)
200 copies available

CONTENTS

EuJAP | Vol. 10 | No. 1 | 2014
UDK 165.1/.2
1 Kripke, S.
1 Ryle, G.
1 Rawls, J.

THE PREOCUPATION AND CRISIS OF ANALYTIC PHILOSOPHY

MICHAEL LOSONSKY
Colorado State University

ABSTRACT

I propose to reconsider Gilbert Ryle's thesis in 1956 in his introduction to The Revolution of Philosophy that "the story of twentieth-century philosophy is very largely the story of this notion of sense or meaning" and, as he writes elsewhere, the "preoccupation with the theory of meaning is the occupational disease of twentieth-century Anglo-Saxon and Austrian philosophy." Ryle maintains that this preoccupation demarcates analytic philosophy from its predecessors and that it gave philosophy a set of academic credentials as a rigorous discipline with its own domain and method. I will maintain that Ryle, with some minor qualifications, was correct in his assessment of the nature of analytic philosophy at that time, and I will argue that the next 50 years continued to be, very largely, the story of meaning, exemplified by the groundbreaking work of Rawls and Kripke. However, I argue that this work also contains the seeds that contributed to the emergence of philosophies that represent a significant departure from analytic philosophy.

Keywords: analytic philosophy, logic, meaning, naturalism, Ryle, Rawls, Kripke, 20th-century philosophy

1. Introduction

What defines analytic philosophy, viewed as a historical phenomenon? In other words, echoing Gilbert Ryle, what is "the story" of analytic philosophy, or, what is its "preoccupation" or even "occupational disease"? This is the question I will try to answer. I will maintain that Ryle's characterization of analytic philosophy in terms of the "preoccupation with . . . meaning" (1957, 239) is correct, with some qualifications, and that philosophy naturalized is an exit from analytic philosophy that leaves analytic philosophy in a crisis. However, I suggest that there is still important work left to do for analytic philosophy, work that reclaims its identity and continues its project of understanding meaning.

2. Characterizing Analytic Philosophy

A challenge to any attempt to characterize analytic philosophy is that it does not have a unifying doctrine or set of doctrines. It is tempting to conclude from this that it only consists of "overlapping strands, with no usefully defining fibre or fibres running through its whole temporal length" (Hacker 1996, 4), a "family resemblance" (Stroll 2000, 7), or even that it is only "a trail of influence" (Soames 2003, 1:xiii). This suggests that the best we can do to characterize analytic philosophy is to find a starting point – typically a set of philosophers, such as Russell, Moore, Wittgenstein and perhaps Frege with a unifying agenda – but whatever

it is that unifies this starting point is wholly transformed and even dropped by succeeding generations, who themselves begin their own agendas that are equally overcome, and so on.

This suggestion is helpful only if we are able to individuate some unifying features of the starting agenda and then also identify some principles that allow us to trace the proper trails of influence. After all, in addition to the trails that remain within the territory of analytic philosophy, there are trails that lead out of it. For example, there is a trail that leads from Frege to Husserl and there is a trail that leads from Austin to Derrida (Derrida 1988, 38 and 130-1). If we are trying to determine the historical identity of analytic philosophy, we need to understand why some trails keep us in and others take us out of the analytic tradition.

This can be done in one of two ways. Appropriate trails of influence can be individuated intensionally, that is in terms of some overlapping concepts and attitudes that define the appropriate trails of influence. Or, defining content can be ignored and instead we can trace a history of reception and succession in terms of tutors, teachers, students, departments, institutes, journals, books, textbooks, and so forth. The latter would be a naturalized, empirical, and social scientific history of analytic philosophy. Hacker in Wittgenstein's Place in Twentieth-century Analytic Philosophy offers several brief but fascinating and enlightening sketches of such lines of influence (Hacker 1996, 148-182).

Hacker, however, is an exception. Analytic philosophers typically are not interested in such naturalistic histories of their own philosophy. Accordingly, my strategy will be intensional, that is, I aim to identify, in Soames' words, the "underlying themes or tendencies that characterize" analytic philosophy (Soames 2003, 1:xiii). Even if there are no unifying doctrines, there may be a set of concepts and attitudes that characterize analytic philosophy and distinguish it from other kinds of philosophy. I hope to characterize this content without begging the question, and in such a way that non-analytic philosophers might find my characterization instructive. Accordingly, I aim to avoid tendentious or 'churchy' characterizations, for example that analytic philosophy is defined by a commitment to "clarity, rigor, and argumentation" and that it aims at "truth and knowledge, as opposed to moral or spiritual improvement" (Soames 2003, 1:xiv).

3. Ryle's Thesis

My point of departure is Ryle's characterization of analytic philosophy in 1956 in his introduction to the publication of a series of eight lectures from BBC's Third Programme with the title The Revolution in Philosophy. An aim of this collection, Ryle writes, is to "trace our proximate origins" and to let the essays in this collection be like "memoirs" that "supply the future historian with those considered and marshalled reminiscences which they will need" (Ryle 1956, 1). This is exactly how I want to use this collection.

In this introduction Ryle distinguishes between "the vehicle and what it conveys," where vehicles are meaningful psychological or linguistic entities, and what they convey is their sense or meaning. It is in virtue of their meaning that the vehicles are "capable of being true or false...and capable of implying and being incompatible with other judgments." Ryle

then characterizes analytic philosophy in its roughly first fifty years as "very largely the story of this notion of sense or meaning" (Ryle 1956, 8). This claim is echoed in 1957 in his essay "The Theory of Meaning," namely that the question "What are meanings?" has "bulked large in philosophical and logical discussions," and that "preoccupation with the theory of meaning could be described as the occupational disease of twentieth-century Anglo-Saxon and Austrian philosophy" (Ryle 1957, 239).

Some clarifications and corrections are in order. First, Ryle does not intend to narrow the focus of analytic philosophy to what Frege isolates as "sense" or "Sinn" in contrast to "reference" or "Bedeutung". Frege's technical notions of sense and reference are part of his own analysis or distillation of meaning, that is, in Michael Beaney's useful phrase, his "splitting of content" into two components (Beaney 1996, 151-2). This distillation is only one answer, albeit a very fruitful and influential one, to the question "What are meanings?" and of course there are others.

Second, for Ryle it is essential that the vehicles are psychological entities, but meanings are not: instead they are abstract objects that belong to the domain of logic and philosophy. While the rejection of psychologism certainly defines early analytic philosophy, I do not believe that this is a necessary condition of analytic philosophy. As in the case of Fregean senses, anti-psychologism is just one strategy for answering the question "What are meanings?" It is preoccupation with this question that first and foremost characterizes analytic philosophy, not a particular answer.

Third, while Ryle's regional, national or linguistic references to Anglo-Saxon and Austrian philosophy might be useful as a rough and ready way to fix the referent of "analytic philosophy", these references include and also exclude too much. Austrian philosophy, strictly speaking, also includes Husserl, Hayek and Adler, none of whom are analytic philosophers. Moreover, Reichenbach, Hempel and Tarski were neither Austrian nor Anglo-Saxon. They were anglophone in that they spoke and wrote English, but English was not their native language and their early works that are already constitutive of analytic philosophy are not in English.

Fourth, Ryle is characterizing a major and perhaps even dominant trend of twentieth-century philosophy, but it certainly does not characterize all of twentieth-century philosophy. If we substitute "analytic philosophy" for "twentieth-century philosophy" or " twentieth-century Anglo-Saxon and Austrian philosophy," Ryle's thesis properly cleaned-up is that the history of analytic philosophy is 'the story of this notion of sense or meaning' and the 'preoccupation with meaning is the occupational disease of analytic philosophy'.

Fifth, while the publication of this collection coincides with what might be called the "second wave" of analytic philosophy or what is sometimes labeled "ordinary language philosophy" and some of the essays aim to locate this wave within the analytic tradition, it is a serious error to see this collection as a mere "manifesto" of ordinary language philosophy, and to assume that the revolution mentioned in its title refers to this second wave (Wright 1994, 16). Certainly this was not the intent of this collection, which aims to cover a movement that includes Frege, Russell, Moore, both late and early Wittgenstein, as well

the Vienna Circle. The collection does include an essay by Strawson that places particular emphasis on the significance of ordinary language philosophy. He characterizes two contrasting "courses", one a course of construction that he ties to the work of Carnap and Quine and the other a course of analysis that he ties to Austin and Ryle.[1] However, Strawson considered both trends as live options in analytic philosophy, and both, on his account, were devoted to the analysis of sentences and propositions (Strawson 1956, 100-1).

Finally and perhaps most importantly, Ryle suggests that there is a line between analytic philosophy and "the technical or semi-technical ideas" of the "new Formal Logic." Ryle writes that these ideas "were taken over by philosophers for the solution of their own problems," but he separates the philosophers' problems from those of the formal logicians (Ryle 1956, 9). This line of separation is also traced by Soames in his 2 volume history of analytic philosophy, which, by his own admission, treats the formal work of Frege and Tarski, among others, "rather sparingly" (Soames 2003, 1:xvii). For example, there are no individual chapters devoted to Frege, Tarski or Carnap. There are good practical reasons for this separation, but it must be underscored that this work in logic is not incidental or merely parallel to the development of analytic philosophy. This work is at its very core, as is evident from the fact that advanced formal logic and even set theory were requirements in analytic graduate programs.

The reason for this is that analytic philosophy's concept of meaning is the concept of something essentially constrained by truth-preserving inferential relationships between propositions. Ryle himself assumes this when he writes that it is in virtue of meaning that vehicles of meaning are "capable of implying and being incompatible with other judgments" (Ryle 1956, 8). Here Ryle is simply following Frege, who in his groundbreaking work on logic, the Begriffsschrift, identifies "conceptual content" in terms of logical inference, specifically deductive validity. For analytic philosophers, truth-preserving inferences and equivalences are essential to meaning, and hence a core feature of this preoccupation with meaning.[2]

Analytic philosophy's inferential conception of meaning is an important reason why Husserl is not a canonical figure in the history of analytic philosophy. While Husserl's Logical Investigations discuss, as the title indicates, logical issues, this is not a work on logic as Frege understood it, and it does not look like a book on logic or the philosophy of logic from our vantage point either. Frege's letter to Husserl written October 30, 1906 concerning logic explains this difference very clearly:

> In logic one must be determined to regard equipollent sentences as differing only according to their form. After the assertoric force with which they have been uttered is subtracted, equipollent sentences have something in common in their content, and this is what I call the proposition they express. This is the only thing that concerns logic. (Frege 1976, 101-3)

[1] He also calls one the "American School" and the other the "English School" (Strawson 1956, 101).

[2] This leaves completely open the question about the appropriate type of this logic.

This common content, or what in the Begriffsschrift he calls the "conceptual content" consists only of that "which affects the possible inferences" (Frege 1879, 3; also iv). Logic is concerned only with "whatever is needed for a valid [that is, truth-preserving] inference" (ibid.). While for Frege conceptual content is "the only thing that concerns logic," in Husserl's Logical Investigations this is at best a very peripheral topic. Central to the Logical Investigations is an attempt to give a theory of intentional objects and contents that play a role in knowledge and experience, as suggested by the subtitle of the massive second volume: "Investigation of of the Phenomenology and Theory of Knowledge" (Husserl 1992, vol. 3, 3). Accordingly, Husserl's notion of meaning is from the very start not constrained by inferential relationships, but instead by the contents of intentional objects that are subject to an "immanent description" of these psychic acts on the basis of reflection. This will include a theory of abstraction, attention, intentionality, among other things. Similarly, Husserl's Formale und Transzendentale Logik discusses the concept of truth as an epistemic and phenomenological category, not in terms of truth-preserving relationships (Husserl 1992, vol. 7, 46-8).[3]

With these five provisos in mind, I take for granted that Ryle's characterization captures the first fifty years. Frege, Russell, Moore, Wittgenstein, as well as the members of the Vienna Circle were all preoccupied with meaning. Just to illustrate this with a simple instance, consider that Frege's Foundations of Arithmetic is entirely devoted to meaning. In the very first sentence Frege raises the question "what the symbol 1 means" (Frege 1978, i) and in the concluding paragraph he declares that the foundational problems of arithmetic, including the treatment of positive whole numbers as well as the difficulties of fractions and negative and irrational numbers are "just . . . a matter of fixing the content of a . . . judgment." (Frege 1978, 119).

What I wish to do here is make a case that this characterization holds for the second fifty years as well, by focusing on the two towering figures of this period: Rawls and Kripke. They are particularly interesting for my purposes because while they are defining fixtures in the history of analytic philosophy in the second half of the twentieth century, their work also motivated important new directions in philosophy that constitute a break with the past that, I will argue, constitute a departure from the analytic tradition.[4]

4. Rawls

I focus on Rawls' discussion of civil disobedience because it is a microcosm of his theory of justice and at the same time it is arguably an essay that plays a role in the emergence of applied ethics, which has an ambiguous status for many analytic philosophers.

The very title of this discussion highlights Rawls' preoccupation with meaning: "The

[3] I thank Sandra LaPointe for challenging me on excluding Husserl from analytic philosophy.

[4] It is useful to distinguish analytical philosophy from analytic philosophy. The term "analytical philosophy" casts a much wider net than the term "analytic philosophy," a net that catches philosophers such as Husserl, who are analytical, but not part of the canon of analytic philosophers. However, the term "analytical philosophy" is an evolution from the narrower conception of analytic philosophy, and hence it is important to understand the narrower conception first.

Definition of Civil Disobedience" (Rawls 1971, 363). With this definition Rawls aims to, in his words, "illustrate the content of the principles of natural duty and obligation." His discussion of civil disobedience has two major components. The first is a definition of civil disobedience as a kind of dissent within the context of a "nearly just society," which Rawls believes conceptually requires a democratic regime. Accordingly, civil disobedience is defined as a certain kind of opposition, namely loyal opposition, to democratic authority. The details of Rawls' definition are not relevant here, but suffice it to say that Rawls maintains that "civil disobedience has been defined so that it falls between legal protest and the raising of test cases on the other side" (Rawls 1971, 367).[5] The second component of his discussion of civil disobedience is the stating of the necessary and sufficient conditions under which civil disobedience is justified. Within the confines of traditional analytic methodology, this is nothing less than a definition of a subset of civil disobedience, namely justified civil disobedience.

Of course, a definition of justified civil disobedience also serves to justify civil disobedience in those instances that satisfy the necessary and sufficient conditions of justified civil disobedience. Thus it is tempting to emphasize the justification Rawls offers at the expense of the analytic task of defining justification. This can give a popular but misleading impression that Rawls' theory of justice is a "revolutionary" departure from earlier analytic ethics because it pays "almost no attention to meta-ethics and instead pursued moral arguments directly."[6] But this understanding of Rawls' project fails to distinguish two distinct kinds of justifications. One assumes a standard of justification and applies it to a certain set of cases, arguing that these cases match or satisfy this standard. The other aims first to offer the definition of a standard, and then sets out to offer the justification. Rawls' argument falls into the second category, and while Rawls' work does depart from his predecessors in terms of emphasis and inspires a new wave of work in ethics, his work does not depart from analytic philosophy's preoccupation with meaning.

Accordingly, Rawls offers three "conditions" for justified civil disobedience (Rawls 1971, 371). First, an act of justified civil disobedience is addressed to substantial and clear violations of the principles of justice that define justice as fairness (Rawls 1971, 372). Second, "the legal means of redress have proved of no avail," and finally, the acts of civil disobedience do not threaten the breakdown of the nearly just society (Rawls 1971, 373-4). It is worth noting how Rawls argues for this third condition. He considers cases that are, in his words, "conceivable...even if...unlikely" that there are many groups that satisfy the first two conditions, so many that their acts of civil disobedience would lead to serious disorder that threatens the just constitution. In other words, in accordance with the analytic pursuit of a definition, Rawls considers logically possible scenarios to tighten his definition of justified civil disobedience (Rawls 1971, 374).

The case of Rawls' analytic discussion of civil disobedience mirrors his overall project of

[5] It should also be noted that Rawls acknowledges that he gives a "narrower definition to civil disobedience" because he distinguishes it from what he calls "conscientious refusal," which he also aims to define (Rawls 1971, 368).

[6] http://en.wikipedia.org/wiki/Normative_ethics Accessed March 31, 2012. and http://www.academicroom.com/topics/normative-ethics Accessed April 13, 2014

offering a theory of justice. Rawls maintains that "the concept of justice...[is] defined... by the role of its principles in assigning rights and duties and in defining the appropriate division of social advantages" (Rawls 1971, 10). The problem is that there are different interpretations of this role – competing principles for assigning rights and duties, and distributing social advantages. These different interpretations are alternative conceptions of justice, and Rawls' primary task is to offer a conception of justice as fairness as a "viable alternative" to then dominant conceptions, particularly utilitarian conceptions (Rawls 1971, 3). The demonstration of viability rests on a justification that rational individuals in the original position would choose the principles of the conception of justice as fairness. This is not intended to be a psychological claim. Rawls intends it to be the conclusion of a "strictly deductive" argument (Rawls 1971, 119 and 121) that follows from, among other things, certain propositions about what it means to be rational, which entail as a matter of meaning that rational individuals would make certain choices given certain epistemic states. While Rawls admits that he falls short of this analytic goal, it is nevertheless the pursuit of this goal that drives Rawls' justification of justice as fairness.

As indicated above, it is possible to abstract from the analytic component of Rawls' theory of justice and isolate a normative argument, for instance an argument for civil disobedience. This sort of abstraction inspired the development of applied ethics, that is, the project of using normative principles to argue for specific courses of action on important social and personal issues. However, minus the analytic preoccupation with meaning, this work ceases to be in the analytic tradition, and this is why much work in applied ethics strikes many readers as a "revolutionary" development in philosophy. In order to be a work in analytic philosophy, it has to be concerned with meaning. Applied philosophy that satisfies this criterion (for example, work on the concept of intrinsic value, in the case of environmental ethics, or on the concept of pain in bioethics) falls within the scope of analytic philosophy.

5. Kripke

Kripke is the other towering figure in the history of analytic philosophy of the second half of the twentieth century. Soames, for example, writes that "the two most important achievements that have emerged from the analytic tradition in this period [1900-1975] are (i) the recognition that philosophical speculation must be grounded in pre-philosophical thought, and (ii) the success achieved in understanding, and separating one from another, the fundamental methodological notions of logical consequence, logical truth, necessary truth, and a priori truth" (Soames 2003, 1:xi). I have doubts about the first claim and I think these two claims are actually incompatible, a case I wish I had time to make. Be that as it may, certainly the second claim is true insofar as it characterizes an important set of influential developments in the analytic tradition in the 1970s, and Kripke's work on meaning and modality is at the center of this development.

The locus classicus for this development is a series of three lectures Kripke gave in 1970 and that were first published in 1972 and then republished in 1980 in the form of a book under the title Naming and Necessity. These lectures begin with a massive critique of descriptive theories of the meaning and reference of proper names that is now a canonical part of analytic philosophy, but the background of this critique is equally important. As Kripke

notes in his Preface to the 1980 edition, these lectures grew out of his groundbreaking model theoretic semantics for modal logic (Kripke 1980, 3), and this is another example of how formal studies are at the very core of developments in analytic philosophy.

Kripke's model-theoretic or possible worlds semantics of modal operators led to questions about how to interpret the terms of identity statements, especially ones that appeared to be contingent identity statements, such as that Aristotle is identical to the tutor of Alexander. If there can be contingent identities, then identities are not necessary, and Kripke held that as a matter of logic identical objects are necessarily identical (Kripke 1980, 3-4). To clear-up the appearance of contingent identities, Kripke introduced the technical notion of a rigid designator, namely a term that designates the same object across all possible worlds in which it exists. In 1963-64 Kripke began to apply this concept to proper names in natural languages, and this move allowed Kripke to decouple the meanings of proper names from descriptions. Once descriptions were expelled from their meanings, it was natural to question whether descriptions are even needed to fix the reference of proper names. If not, then names can succeed in referring directly without some intervening descriptive content. Moreover, the content or semantic contribution a name makes to the meaning of a sentence can be limited to its referent. These theses about the meanings of proper names were also extended to natural kind terms, separating the content of natural kind terms from the descriptive content of our beliefs and theories about those kinds.

As already indicated, these new theories of meaning for proper names and natural kind terms motivated important revisions in a wide range of areas of philosophy. A received view was that a priori knowledge rests on what can be known on the basis of reflection on the descriptive contents of our concepts, and that necessary truths rest on these descriptive contents knowable a priori. But if we assume that the terms "water" and "H_20" both function as rigid designators and that identical things are necessarily identical, then it is necessary that water is identical to H_20, but this is known a posteriori, not a priori. The traditional view that Ruth Millikan calls "Meaning Rationalism" (Millikan 1984, 10), namely that we know a priori, simply relying on Cartesian reflection, what we are thinking and intending to do when speaking or writing also suffered a setback. If we are thinking, speaking or writing with rigid designators, whose meanings "just ain't in the head," as Putnam famously puts it, then meaning rationalism appears to be false (Putnam 1975, 227). In short, rigid designators ushered in various forms of externalism according to which meaning was not at all or at least not entirely determined by the content internal to us.

6. Consequences

Beneficiaries of these changes in analytic philosophy's conception of meaning were realism and naturalism, particularly in the philosophy of science and ethics. However, a closer look at this development also reveals an exit from analytic philosophy that is the source of what has been characterized as an identity crisis (Wright 1994, 4).[7] The fate of Moore's Open Question argument in recent philosophy is a case in point. To simplify, Moore argued that the property of being good cannot be identical to a natural property, say, to use Moore's

[7] Reference to a crisis in analytic philosophy are not uncommon: Baz (2014), Glock (2008, 1), Preston (2007, 7), Preston (2005, 294), Leiter (2004, 1), and Biletzky and Matar (1998, xi).

example, the desire to desire. The reason is that the question "Is it good to desire to desire?" is intelligible and asks for new information. It is not equivalent to the trivial question "Is the good good?". Moore also puts his case in another way: the mere fact that we can doubt that the desire to desire is good "shews clearly that we have two different notions before our minds" (Moore 2005, 16). This way of putting Moore's argument is particularly significant because it rests on the same principle Frege used in the opening paragraph of "Sense and Reference" regarding identity statements and that Gareth Evans called the "Intuitive Criterion of Difference" (Evans 1982, 19). Two sentences have distinct meanings if a competent speaker understands both but without being incoherent or irrational can affirm one while deny or remain agnostic about the other. More broadly, what Moore and Frege assumed is that in assigning meaning to a person's linguistic or mental states, we must take into account differences in their cognitive attitudes. I will call this "Frege's Constraint" (following White 2004, 213).

The canonical naturalist reply is to deny that the cognitive differences that play a role in Moore's argument are relevant to the individuation of properties. In Brink's terms, for example, the "Semantic Test for Properties" fails, and part of the reason is that terms or predicates can succeed in referring or denoting something without satisfying any of the cognitive content a person has in mind while using the term (Brink 1989, 163-6; 2001, 154-76). Just as the terms "water" and "H_20" can be associated with cognitive differences (a person can rationally wonder if water is indeed H_20 while water and H_20 are as identical), "good" and some naturalistic term can be co-referential despite cognitive differences. What in the end underlies this response is the idea that moral terms can function as rigid designators as well as proper names or natural kind terms.

This is an example of how the preoccupation with meaning led to an influential and important shift in analytic meta-ethics. But it also points to a trail that leaves analytic philosophy behind once naturalists cease to care about questions of meaning and turn to explanatory projects constrained by results from the biological or social sciences at the expense of Frege's constraint.[8]

To elaborate, the canonical naturalist reply to Moore's Open Question Argument is not at all a refutation of his argument, but simply a change of topics. Moore was interested in the meaning of "good" and he assumed meaning was subject to Frege's constraint. This is evident when he writes that his argument "shews clearly that we have two different notions before our minds" (Moore 2005, 16). He was not concerned with matters of fact – for instance, in what kinds of things turn out to be good – but what it is that we understand – what is "before our minds" – when we use evaluative terms such as "good". Accordingly, as Akeel Bilgrami argues, the canonical naturalist reply does not show "that there is any fault in Moore's argument itself" because it remains "effective (and is only intended to be effective) against naturalisms that are definitional" (Bilgrami 2004, 129-30). Even if in fact good is identical to some natural property F, it is possible for a person without being irrational to doubt or even deny that good is identical to F, and this shows that the concept

[8] I think Richard Boyd's essay "How to Be a Moral Realist" is an exemplar of this departure, beginning with a discussion of semantic issues and ending with a 'just so' story of empirically discoverable homeostatic property clusters.

of good and the concept of F are distinct.

In effect, a naturalist who is completely satisfied with the canonical naturalist reply and with no concern for the relevant cognitive differences ceases to care about meaning subject to Frege's constraint. But the concept of meaning that is essential to analytic philosophy's preoccupation is meaning that not only serves inferential relationships, but that is also subject to Frege's constraint. Hence, a loss of interest in this concept of meaning is a loss of interest in analytic philosophy.

Of course, there is a way to understand this naturalist turn in philosophy as still a kind of preoccupation with meaning, except that now " 'meanings' just ain't in the head". I think this move is only a Pyrrhic victory. Frege's constraint is now replaced by the results of the empirical and special sciences. Meaning now is assigned not on the basis of how people understand what they say and think, but on the basis of what science has to say about what people think they understand. In this sense of meaning, any special science is also concerned with meaning because its results determine what we mean, not what appears before our minds.

This is a Pyrrhic victory because it draws all blood from meaning as a special subject for analytic philosophy. Meanings that are subject only to the constraints of natural science leave nothing for analytic philosophy to do except perhaps to correct those who still think that philosophy has a special domain. This also leaves analytic philosophy without a clear identity. Ryle in his introduction to The Revolution in Philosophy maintained that analytic philosophy was partially a response to new institutional pressures on philosophers from secular colleagues, mostly natural scientists, to identify a domain of expertise and method for philosophy (Ryle 1956, 4-5). Meaning and its analysis was this special domain of expertise. But meaning without Frege's constraint drains this special domain of expertise, and this leaves analytic philosophy in crisis.

7. The Roots of Crisis

Von Wright suggests a different diagnosis of analytic philosophy's identity crisis, namely that it is caused by pair of competing commitments that can be found at the very roots of analytic philosophy (Wright 1994, 3-32). He explains these commitments in terms borrowed from Frederick Waismann's distinction between science and philosophy in his essay "What is Logical Analysis?" The scientist "searches for knowledge, i.e. propositions which are true, which agree with reality," and the highest goal of this attitude is the construction of theories about matters of fact, Waismann writes. Philosophers, in particular analytic philosophers, "cannot be satisfied with this" because they find "the very nature of knowledge and truth . . . problematic" and are interested in the "deeper meaning of what the scientist does". Consequently, analytic philosophers aim not at "propositions, but the clarification of propositions" (Waismann 1940, 265-6). While Waismann sets this as a line of demarcation between science and philosophy, von Wright argues that both of these "attitudes of mind" are found within analytic philosophy, and that these developed into an "unbearable contradiction" that "had to destroy its unity" (Wright 1994, 4).

The philosophies of Moore and Russell, von Wright argues, already are marked by these opposing poles. Russell, von Wright suggests, is motivated by the scientific search for true propositions about matters of fact, and this characterizes Russell's work on mathematics and logic as well as his work on the problem of induction. In Our Knowledge of the External World as a Field for Scientific Method in Philosophy Russell proclaims that "the time has now arrived" to correct the "unsatisfactory state" that philosophy "has achieved fewer results than any other branch of learning," and von Wright cites this as the clearest expression of Russell's scientific attitude (Wright 1994, 6-7). Moore, on the other hand, is concerned with meaning. For example, it is not the truth of commonsense propositions that concerns Moore, but the analysis of their meaning (Wright 1994, 7-8).

While I think that von Wright's diagnosis points in the right direction, first I wish to correct some elements of this diagnosis. First, von Wright's discussion of Waismann's distinction is misleading. Waismann's distinction is not simply between questions of truth and questions of meaning because from his answer to the question "What is Logical Analysis?" it is clear that philosophy also involves the pursuit of true propositions. Philosophy, Waismann maintains, involves the assertion of tautologies, and a tautology, in his words, "expresses agreement with all truth-possibilities, i.e., that it is always true," adding that the truth of tautologies (as well as the falsity of contradictions) "no longer depends on the behavior of the real world" (Waismann 1940, 268-9). So Waismann's distinction is better understood as a distinction between different kinds of truths. Needless to say, this assumes that we can draw a distinction between these two domains, and Waismann assumes we can.

This assumption that there is a distinction between different kinds of truths is shared by Russell and Moore, although they drew this distinction differently over the courses of their careers. While both began with a Platonist ontology of meanings and Russell departs from this commitment in his theory of descriptions, Russell's turn to logical constructions and logical form preserves the commitment to a special domain for philosophy. Waismann expresses this quite nicely in his account of logical analysis, namely that philosophers aim at the "clarification of the sense of . . . propositions," and they do this by "demonstrating the purely logical relations between propositions" (Waismann 1940, 266 and 269).

Second, Russell's pursuit of scientific philosophy in 1914 in Our Knowledge of the External World and the companion essay "The Relation of Sense-Data to Physics" does not quite fit von Wright's divide between questions of truth and questions of meaning. For Russell, part of what makes philosophy scientific is his principle of construction. He writes: "the supreme maxim of scientific philosophizing is this: Wherever possible, logical constructions are to be substituted for inferential entities" (Russell 1986, 11). This is the principle that guides Russell's theory of descriptions, which is intended as an analysis of both the meaning and denotation of definite descriptions, and when Russell turns to the concepts of physics in 1914, he uses this principle to guide his account of the meaning of physical terms such as "matter" "object" or "place".

This principle gives philosophy a certain scientific character that Russell describes quite clearly in the essay "On Scientific Method in Philosophy" published as a pamphlet in 1914. He distinguishes between philosophy that aims to be scientific by seeking to base itself

upon the "results of science" or by adopting the "methods of science". Russell maintains that "much philosophy inspired by science has gone astray through preoccupation with the results momentarily supposed to have been achieved" (Russell 1986, 57). Russell cites the 19th-century naturalist Herbert Spencer, but surely this applies to trends in contemporary naturalism as well. For Russell, the appropriate way of making philosophy scientific is to transfer "not results, but methods" from "the sphere of the special sciences to the sphere of philosophy." That is, philosophy for Russell has a special sphere. According to Russell, "philosophy is the science of the possible", and possibility is studied by enumerating the logical forms of propositions and facts, and this is not an empirical investigation (Russell 1986, 65).

Moreover, the propositions of philosophy "must be a priori", which for Russell means that they are neither confirmed nor disconfirmed by empirical evidence because they hold "however the actual world were constituted" (1986, 65). So it is not the appeal to empirical evidence or the pursuit of contingent truths that makes philosophy scientific for Russell. What makes it scientific is that with a well-defined set of problems and a method, philosophy is in a position to deal "with its problems piecemeal, and . . . obtain, as the sciences do . . . partial . . . results . . . [that] subsequent investigation can utilize even while it supplements and improves them" (Russell 1986, 66). In other words, philosophy is scientific insofar as it is "progressive", which means that it consists of "successive approximations to the truth," but these are truths about logical form and possibility. So Russell indeed is concerned with questions of truth, as von Wright observes, but not with truths about empirical and contingent matters of fact, but with truths about logical form and possibility.

Russell's increasing interest in psychological matters could be seen as blurring the line between Russell and naturalism. Hylton (1990, 244) argues that after Russell's Principles of Mathematics, published in 1903, Russell begins "to take some interest in questions which he might have dismissed as merely psychological." Hylton maintains that the fact that for Russell after the Principles "philosophical theories . . . appear to be answerable to the data of experience, to facts about what is or can be plausibly supposed to be present to our minds . . . is clearly a considerable concession to psychologism" (1990, 330). While Hilton is correct about Russell's increased interest in psychological matters, an appeal to psychological facts and psychological plausibility does not constitute a concession to psychologism. After all, Frege's appeal to cognitive value or significance [Erkenntniswert] and his appeal to the distinction between what can be known a priori and what is a "valuable extension of our knowledge" (1980, 40) is a psychological appeal, but it is not a concession to psychologism. Finally, Hylton concedes that "Russell continues to think that the subject-matter of philosophy is wholly independent of psychology" (283). This is precisely what Russell delineates in his 1914 discussion of the scientific method in philosophy and that distinguishes Russell from contemporary naturalism. In general, a turn toward psychological facts is not, ipso facto, a naturalistic turn unless these facts, to use Russell's terminology, are results simply transferred from empirical psychology rather than facts pertaining to a philosophical psychology that is distinct from empirical psychology.

While von Wright is mistaken that Waismann's division between science and philosophy is a division within the origins of analytic philosophy and a source of its eventual disinte-

gration, he nevertheless points in the right direction. In the theory of descriptions, Russell applied his maxim "Wherever possible, logical constructions are to be substituted for inferred entities" (Russell 1986, 11) against Frege's and Meinong's ontologies of intensional entities. An essential feature of this intensional approach is that it is driven in part by a conception of meaning that aims to capture intuitive and cognitive differences in the understanding of language. This, of course, is Frege's constraint.

Russell's approach, especially in his theory of descriptions, suggests that in questions of meaning, ontological constraints can trump Frege's constraint. As Russell repeats several times, his analysis is "in obedience to the feeling of reality," and a "robust sense of reality is very necessary in framing a correct analysis of propositions," even if it comes at a cost of "apparently excessive complication," in particular the familiar complexity of analyzing definite descriptions into parts that only make sense in the context of a whole sentence with nested quantifiers in which there is no single unit that can be identified as a substitution for the definite description (Russell 1919, 48).

It is this weakening of the commitment to Frege's constraint and letting ontological considerations take its place that are the roots of analytic philosophy's crisis in the wake of Kripke's work on naming and necessity. Kripke's own case against descriptivism is guided by Frege's constraint. A gloss on one of his arguments is that someone can consistently believe that Aristotle is the tutor of Alexander while this is, in Kripke's words "not a necessary truth for him" (Kripke 1980, 63). Hence, the meaning of the proper name "Aristotle" cannot be identified with the description of him being the tutor of Alexander. Another argument is that a person can have beliefs about Einstein and meaningfully use "Einstein" without having any beliefs about him involving definite descriptions; hence a name can be meaningful without backing definite descriptions.

However, the irony of Kripke's achievement is that it prepared the decommission of Frege's constraint and inspired a naturalized philosophy that is no longer preoccupied with meaning constrained by cognitive differences. Instead, questions concerning matters of fact – for example "just-so stories" appealing to (borrowing from Russell) "results momentarily supposed to have been achieved," say, in evolutionary biology – replace questions about truths of meaning. If Ryle is correct that the preoccupation with meaning is essential to analytic philosophy and I am right that the relevant notion of meaning is one that is subject to Frege's constraint, then these kinds of philosophical projects, while they might borrow some of the standards and methods of analytic philosophy, are not examples of analytic philosophy anymore.

8. Conclusion

Stephen White made a trenchant parenthetical comment about the status of Frege's constraint today, namely that "the profession is currently in denial" about it. Current philosophical theorizing simply ignores Frege's constraint, White claims, "but not on the basis of cogent arguments" (White 2004, 222). White makes this point regarding current work in the philosophy of mind, arguing that a satisfactory account of agency needs to account for the "agential perspective," and that this in turn needs an account of meaning that satisfies

Frege's constraint. Without Frege's constraint, the perspective of the agent is ignored, and by ignoring this personal perspective, we ignore agency.

I would like to put this point somewhat differently. Frege's constraint is ignored on account of a major shift in philosophy that was made possible by rigid designators and direct theories of reference and meaning. Ignoring Frege's constraint is an unintended consequence and the profession has not come to terms with the fact that this marks a departure from previous philosophizing that is at least as dramatic as the inception of analytic philosophy itself. In fact, as I have been arguing, it is a way of doing philosophy that is not analytic anymore. Analytic philosophy is, at its core, preoccupied with meaning, where meaning is subject to Frege's constraint, but contemporary "naturalizing" philosophy is not.[9] Additionally, insofar as naturalized philosophy now dominates our profession, replacing the preoccupation with meaning with a preoccupation with the results of the natural sciences, analytic philosophy is not a dominant force in contemporary philosophy anymore. Nevertheless, contemporary analytic philosophy still has an important and yet unfinished task ahead, namely to revive and revitalize the intuitive criterion of difference and defend its essential role in any adequate account of our own self-understanding, particularly as subjects and agents. The revival of Frege's constraint, that is, the recognition that differences in our own understanding of what we mean have an essential role in determining what we mean, reclaims the identity of analytic philosophy, turns to its proper tasks, and makes analytic philosophy relevant again to the philosophical project of understanding what it is that we mean when we think, speak and act.

REFERENCES

Baz, Avner. 2014. The Crisis of Method in Contemporary Analytic Philosophy, and Ordinary Language Philosophy. In *The Reemergence of Ordinary Language Philosophy*, ed. N. Shah and A. George. Oxford: Oxford University Press, forthcoming.

Beaney, M. 1996. *Frege: Making Sense*. London: Duckworth.

Biletzky, A. and A. Matar, eds. 1998. *The Story of Analytic Philosophy: Plots and Heroes.* London: Routledge.

Bilgrami, A. 2004. Intentionality and Norms. In *Naturalism in Question*, ed. M. De Caro and D. Macarthur, 125-51. Cambridge: Harvard University Press.

Brink, D. O. 1989. *Moral Realism and the Foundations of Ethics*. Cornell: Cornell University Press.

Brink, D. O. 2001. Realism, Naturalism, and Moral Semantics. *Social Philosophy and Policy* 18:154-76.

De Caro, M. and Macarthur, eds. 2004. *Naturalism in Question*. Cambridge: Harvard University Press.

Derrida, J. 1988. *Limited Inc.* Trans. S. Weber. Evanston: Northwestern University Press.

[9] I borrow the gerund "naturalizing" from Kornblith (1985), the root, of course, is Quine (1969). For critical discussions of the varieties of naturalism see De Caro and MacArthur (2004). The essays in this collection indicate a way out of the crisis of analytic philosophy.

Frege, G. 1879. *Begriffsschrift, Eine der Arithmetik Nachgebildete Formelsprache des reinen Denkens*. Halle: Louis Nebert.

Frege, G. 1976. *Wissenschaftlicher Briefwechsel*, ed. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart. Hamburg: Felix Meiner.

Frege, G. 1978. *The Foundations of Arithmetic/Die Grundlagen der Arithmetik*. German text with English trans. by J. L. Austin. 2nd ed. Oxford: Blackwell.

Frege, G. 1980. Sinn und Bedeutung. In Funktion, Begriff, Bedeutung, ed. G. Patzig, 40-65. Gottingen: Vandenhoeck & Ruprecht.

Glock, H. 2008. *What is Analytic Philosophy*. Cambridge: Cambridge University Press.

Hacker, P. M. S. 1996. *Wittgenstein's Place in Twentieth-Century Analytic Philosophy*. London: Blackwell.

Husserl, E. 1992. *Gesammelte Schriften*, 8 vols. Hamburg: Felix Meiner.

Kornblith, H., ed. 1985. *Naturalizing Epistemology*. Cambridge: MIT Press.

Kripke, S. 1980. *Naming and Necessity*. 2nd ed. Cambridge: Harvard University Press.

Leiter, B. 2004. Introduction: The Future for Philosophy. In *The Future for Philosophy*, ed. B. Leiter, 1-23. Oxford: Oxford University Press.

Preston, A. 2005. Conformism in Analytic Philosophy: On Shaping Philosophical Boundaries and Prejudices. *Monist* 88: 292-319.

Preston, A. 2007. *Analytic Philosophy: The History of an Illusion*. London: Continuum.

Quine, W. V. O. 1969. Epistemology Naturalized. In *Ontological Relativity and Other Essays*, 69-90. New York: Columbia University Press.

Rawls, J. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.

Russell, B. 1919. *Introduction to Mathematical Philosophy*. London: Allen and Unwin.

Russell, B. 1986. The Philosophy of Logical Atomism and Other Essays, ed. J. G. Slater. In *The Collected Papers of Bertrand Russell*, vol. 8. London: Allen and Unwin.

Russell, B. 1992. Logical and Philosophical Papers 1909-13, ed. J. G. Slater. In *The Collected Papers of Bertrand Russell,* vol. 6. London and New York: Routledge.

Ryle, G. 1956. Introduction. In *The Revolution in Philosophy*, ed. G. Ryle, 1-11. London: Macmillan.

Ryle, G. 1957. The Theory of Meaning. In *British Philosophy in the Mid-century*, ed. C. A. Mace, 239-64. London: Allen and Unwin.

Soames, S. 2003. *Philosophical Analysis in the Twentieth Century*. 2 vols. Princeton: Princeton University Press.

Strawson, P. F. 1956. Construction and Analysis. In *The Revolution in Philosophy*, ed. G. Ryle. London: Macmillan.

Stroll, A. 2000. *Twentieth-Century Analytic Philosophy*. New York: Columbia University Press.

Waismann, F. 1940. Was ist logische Analyse? *The Journal of Unified Science (Erkenntnis)* 8: 265-289.

White, S. 2004. Subjectivity and the Agential Perspective. In *Naturalism in Question*, ed. M. De Caro and D. Macarthur, 201-30. Cambridge: Harvard University Press.

Wright, G. H. v. 1994. Analytische Philosophie – eine historische-kritische Betrachtung. In *Analyōmen 1: Proceedings of the 1st Conference "Perspectives in Analytical Philosophy,"* ed. G. Meggle and U. Wessels. Berlin: de Gruyter.

Received: October 31, 2012
Accepted: June 28, 2014

Department of Philosophy
Colorado State University
Fort Collins, CO 80521-1781
USA
Michael.Losonsky@Colostate.edu

EuJAP | Vol. 10 | No. 1 | 2014
UDK 165.1/.2
519.2
1 Bayes, T.
53-05 Cox, R. T

WHAT IS PROBABILTY AND WHY DOES IT MATTER

ZVONIMIR ŠIKIĆ
University of Zagreb

ABSTRACT

The idea that probability is a degree of rational belief seemed too vague for a foundation of a mathematical theory. It was certainly not obvious that degrees of rational belief had to be governed by the probability axioms as used by Laplace and other prestatistical probabilityst. The axioms seemed arbitrary in their interpretation. To eliminate the arbitrariness, the statisticians of the early 20th century drastically restricted the possible applications of the probability theory, by insisting that probabilities had to be interpreted as relative frequencies, which obviously satisfied the probability axioms, and so the arbitrariness was removed. But the frequentist approach turned more subjective than the prestatistical approach, because the identifications of outcome spaces, the choices of test statistics, the declarations of what rejection regions are, the choices of null-hypothesis among alternatives, the contradictory choices between sizes and powers etc., depend on thoughts or even whims of the experimenter. Frequentists thus failed to solve the problems that motivated their approach, they even exacerbated them. The subjective Bayesianism of Ramsey and de Finetti did not solve the problems either. Finally Cox provided the missing foundation for probability as a degree of rational belief, which makes the Bayesian probability theory (which is based on this foundation) the best theory of probable inference we have. Hence, it is quite unbelievable that it is not even mentioned in recent philosophy textbooks devoted to the probable inference. The reason could be that it requires fairly sophisticated mathematics. But not even to mention it? We explain the history and prove Cox theorem in a novel way.

Keywords: probability, subjective Bayesianism, logical Bayesianism, Cox theorem

Probability has a mathematical aspect and a scientific aspect. There is a reasonable agreement about the mathematics of probability. Almost everybody accepts the same probability axioms and has no disputes about the truths of the mathematical theory of probability. Yet, when it comes to scientific applications of the theory there are different opinions about what probabilities are. Some identify them with degrees of (rational) belief, some with limiting frequencies, and there are other opinions. Why does it matter?[1] We explain why, starting with a simple problem of a coin fairness testing.[2]

Consider a hypothesis about the probability of a coin coming up heads. If we denote it by H, then $H = 0$ and $H = 1$ represent a coin which, respectively, produces a tail or a head on every flip. There is a continuum of possibilities between these extremes, with $H = 1/2$ indicating a fair coin. Now, if you had observed 3 heads in 12 flips, do you think it was a fair coin?

[1] Think about mathematics of numbers, i.e. arithmetic. In scientific applications it does not matter what numbers are. What is important are rules that numbers obey, not what they really are. Is it not the same with probabilities?

[2] The idea is to compare Laplacean (prestatistical) approach to a textbook problem, with the statistical approach to the same problem. It may seem that this is more appropriate for a college course than for a research article, but my experience is that such an introduction is eye opening (and surprising) even for the audiences that are highly trained in probability. The problem is taken from Sivia 1996, ch.2.

A Bayesian who thinks of probabilities as degrees of rational belief, will use Bayes' theorem to answer this question:[3]

$$pr(H \mid D, I)dH = \frac{pr(H|I)dH \cdot pr(D|H,I)}{pr(\mathrm{D}|I)}$$

Actually, he will use a simpler form:

$$pr(H \mid D,I) \propto pr(H \mid I)\, pr(D \mid H, I),$$

because he can evaluate the missing constant (which does not depend on H) from the normalisation condition

$$\int_0^1 pr(H|D,I)dH = 1.$$

The power of the theorem lies in the fact that it relates the probability that the hypothesis H is true, given the data D (e.g. 3 heads in 12 flips) and background information I (e.g. flipping is vigorous, coin is symmetric etc.), to the probability that the data would have been observed if the hypothesis is true, which is easier to assign.

Prior probability $pr(H \mid I)$ represents the degree of rational belief in H given I (with no data D available). It is modified by the data D, through the *likelihood* $pr(H \mid D, I)$, and yields the *posterior* probability $pr(H \mid D, I)$, which represents the degree of rational belief in H given I and the data D.

In our specific case of coin flipping, *prior* $pr(H \mid I)$ represents what is known about the coin before any data is taken into account. The state of ignorance is represented by the uniform probability assignment

$$pr(H \mid I) = \begin{cases} 1 & 0 \le H < 1 \\ 0 & \text{otherwise} \end{cases}$$

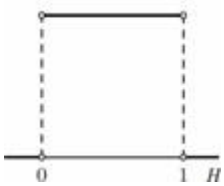

Fig. 1

This prior state of ignorance is modified by the data through *likelihood*:

$$pr(D \mid H, I) \propto H^R(1-H)^{N-R},$$

[3] In what follows *pr* is probability or probability density which, for short, we also call probability.

where H is the probability of obtaining a head and R is a number of heads obtained in N flips. (For simplicity, equality is again replaced by proportionality, since the omitted term does not depend on H.) By Bayes' theorem:

$$pr(D \mid H, I) \propto H^R(1-H)^{N-R}, \text{ for } 0 \leq H \leq 1,$$

otherwise it is 0. If the coin is flipped once and it comes up heads, the resulting posterior is:[4]

$$pr(H \mid \{h\}, I) \propto H$$

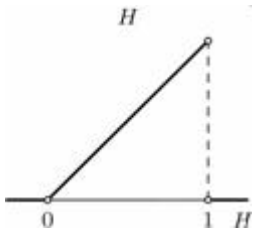

Fig. 2.

If the coin is flipped for a second time and again comes up heads, the resulting posterior is:

$$pr(H \mid \{h,h\}, I) \propto H^2$$

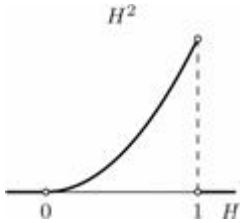

Fig. 3.

If the third flip comes up tails the resulting posterior is:

$$pr(H \mid \{h,h,t\}, I) \propto H^2(1-H)$$

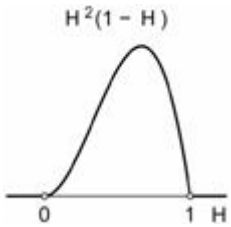

Fig 4.

If the forth flip also comes up tails the resulting posterior is:

[4] The graphs are normalized in such a way that the maxima are always 1.

$pr(H \mid \{h,h,t,t\}, I) \propto H^2(1\text{-}H)^2$

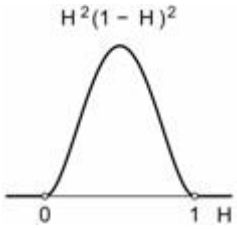

Fig. 5.

And so on. The following figures show how the posterior evolves as the number of data becomes larger and larger.[5] The position of the maximum wobbles around, but the wobbling decreases with the increasing amount of data. The width of the posteriors also becomes narrower with more data. For the coin in our example, the best estimate of H converges to 0.25 (of course this was the value used to simulate the flips).

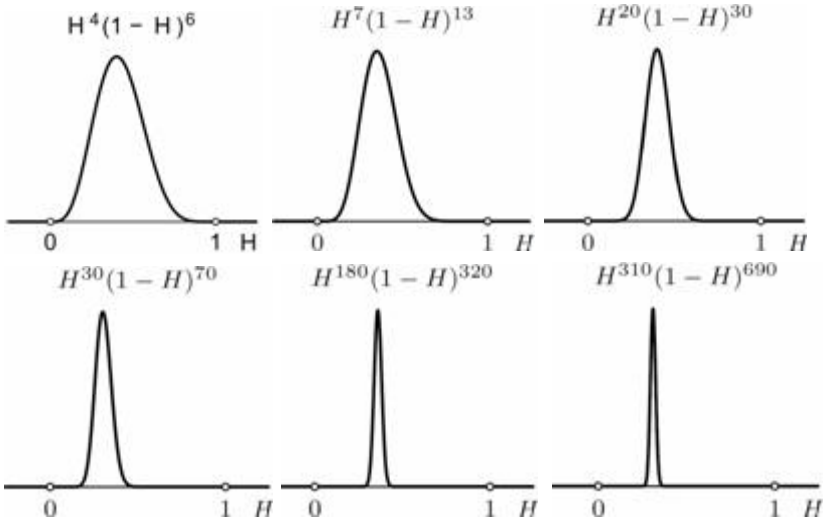

Fig. 6.

People tend to agree with the binomial distribution for the likelihood ($pr(D \mid H, I) \propto H^R(1\text{–}H)^{N\text{-}R}$) but worry about the prior: how would the inference about the coin have changed if a different prior was chosen? If we repeated the analysis of the date with different priors the results would have been the same, because the posterior is dominated by the likelihood, and the choice of the prior is largely irrelevant (cf. fig. 2.2. in Sivia 1996).

[5] It is easy to prove that it does not matter whether the data are introduced one by one or all of them simultaneously.

The result of the Bayesian analysis is the probability distribution of every possible hypothesis H, given one real data set D.

Often, we wish to summarize this distribution with just two numbers: the best estimate and its reliability. If we denote the posterior by P, i.e. $P = pr(H \mid D, I)$, then the best estimate of its value is the maximum point H_0 given by:

$$\frac{dP}{dH} \mid_{H_0} = 0 \quad \text{and} \quad \frac{d^2P}{dH^2} \mid_{H_0} < 0.$$

The measure of the reliability of the best estimate is the spread of the posterior about it. The behaviour of any function around a point can be estimated by its Taylor's expansion about the point. But, rather than dealing with posterior P, it is easier to deal with its logarithm $L = \ln P$. Expanding L about H_0, we get:

$$L \approx L(H_0) + \frac{1}{2}\frac{d^2L}{dH^2} \mid_{H_0} (H - H_0)^2,$$

(the linear term is 0 because L is monotone in P, so H_0 as the maximum point of P is also the maximum point of L).

If we define $1/\sigma^2 = -(d^2L/dH^2) \mid_{H0}$ we get:

$$L \approx L(H_0) - \frac{(H - H_0)^2}{2\sigma^2}$$

which by exponentiation yields:

$$P = pr(H \mid D, I) \approx P(H_0)\, e^{-\frac{(H - H_0)^2}{2\sigma^2}}$$

From the normalization condition:

$$1 = \int_0^1 pr(H|D,I)dH \approx \int_0^1 P(H_0)\, e^{-\frac{(H - H_0)^2}{2\sigma^2}}\, dH = P(H_0)\sigma\sqrt{2\pi}$$

it follows that $P(H_0) = 1/\sigma\sqrt{2\pi}$, i. e.

$$pr(H \mid D, I) \approx \frac{1}{1/\sigma\sqrt{2\pi}} e^{-\frac{(H - H_0)^2}{2\sigma^2}}$$

This normal or Gaussian approximation[6] is usually conveyed by the statement:

$$H = H_0 \pm \sigma,$$

with H_0 the best estimate and σ referred to as the *error-bar*. The integral properties of the normal approximation tell us that the probability that the true value of H lies within $\pm\,\sigma$ of H_0 is 67% i. e.

$$pr(H-\sigma \le H \le H_0+\sigma|D, I) = \int_{H_0+\sigma}^{H_0+\sigma} pr(H|D,I)dH \approx 0{,}67$$

Similarly, the probability that H lies within $\pm\,2\sigma$ of H_0 is 95%, and that H lies within $\pm\,3\sigma$ of H_0 is 99.8%.

In the coin example:

$$P = pr(H \mid D, I) \propto H^R(1-H)^{N-R},\ 0 \le \mathrm{H} \le 1,$$
$$L = C + R \ln H + (N-R)\ln(1-H),$$
$$\frac{dL}{dH} = \frac{R}{H} - \frac{N-R}{1-H} = 0 \quad \text{for} \quad H_0 = \frac{R}{N},$$
$$\frac{d^2L}{dH^2}\Big|_{H_0} = -\frac{R}{H_0^2} - \frac{(N-R)}{(1-H_0)^2} = -\frac{N}{H_0(1-H_0)},$$
$$\sigma = \sqrt{\frac{H_0(1-H_0)}{N}} < \frac{1}{\sqrt{N}}$$

Hence, the relative frequency of heads R/N is the best estimate of H, and its error-bar σ is less than $1\sqrt{N}$. So, the width of the posterior becomes narrower with the increasing number of the data N, as we have seen above (cf. fig. 6.).[7]

This prestatistical approach to our problem was the standard approach of Laplace and his contemporaries. As we have just seen, the approach is extremely successful. But nevertheless, it was rejected by the frequentists of the late 19th and the early 20th century. Why? The idea that probability is a degree of rational belief seemed too vague for a foundation of a mathematical theory. It was certainly not obvious that degrees of rational belief had to be governed by the probability axioms used by Laplace and others. The axioms seemed arbitrary in this interpretation.[8] To eliminate this arbitrariness, the mathematicians of the late

[6] The approximation is just the quadratic approximation of the logarithm and has nothing to do with coins or probabilities.

[7] The last formula also proves that it is easier to identify a highly biased coin than it is to be confidential that it is fair, because the nominator $H_0(1-H_0)$ is greatest when $H_0 = 1/2$.

[8] Furthermore, the probability rules described how to manipulate probabilities, but they did not specify how to assign the prior probabilities that were being manipulated. We will not address this

19th and the early 20th century drastically restricted the possible applications of the theory, by insisting that probabilities had to be interpreted as relative frequencies of occurrences in repeated random experiments (mostly imagined, rarely actual). The relative frequencies obviously satisfied the probability axioms, hence their arbitrariness was removed. Also, the frequentist approach, by its reference to observation of repeated experiments, seemed to make probability an objective property of "random phenomena" and not a subjective degree of the rational belief of Bayesians.

But, the frequency definition of probability made the concept of the probability of a hypothesis illegitimate, e. g. the prior $pr(H|I)$ and the posterior $pr(H|D, I)$ in the coin example make no sense. A hypothesis is either true or false, it is not a random variable. A consequence is that scientists are not allowed to use Bayes' theorem to asses hypotheses. So, how would a frequentist deal with the coin fairness problem? He can not calculate the probability of the fairness hypothesis (the hypothesis that $H = 1/2$), even less the probability distribution of every possible hypothesis H, given the data D, since hypotheses have no probabilities.

Hence, Fisher developed his system of significance tests for hypotheses testing. To perform the test, an experiment must be devised, in our example flipping the coin a predetermined number of times, say 12, and then the result analysed in three steps.

First, specify the outcome space. In our example 2^{12} possible sequences of 12 heads or tails. The result of the experiment should be summarised in some numerical form, e.g. the number of heads in the outcome. This summary is called *test-statistics*, and as a function of outcomes it is a random variable which has probability.

Second, calculate the probability of every possible value of the test-statistics, given the hypothesis you are testing (Fisher called it the *null-hypotheses*). This is the *sampling distribution* of the test-statistics. In our case it is $pr(R) = \binom{12}{R}(1/2)^R(1/2)^{12-R}$, with R the number of heads:

R	*pr R*	*R*	*pr R*	*R*	*pr R*	*R*	*pr R*
0	0.0001441406	3	0.053710938	6	0.225585938	9	0.053710938
1	0.0029296875	4	0.120849609	7	0.193359375	10	0.0161132813
2	0.0161132813	5	0.193359375	8	0.120849609	11	0.0029296875
						12	0.0002441406

Third, look at all results which *could have* occurred (given the null-hypothesis) and which,

question, although we have seen that at least in some cases, e. g. our coin example, it is irrelevant. Furthermore, probability is the logic of uncertainty, and as the standard logic does not tell us what are the factual truths, but only what follows from what, so probability does not tell us what are prior probabilities but only what probability follows from given probabilities.

as Fischer put it, are more extreme than the result that did occur. It means their probability is less than or equal to the probability of the actual outcome. Then calculate the probability *pr** that the outcome will fall within this group. For example, if our experiment produced 3 heads in 12 flips, the result with less or equal probabilities to this are $R = 0,1,2,3,9,10,11,12$; and the probability of at least one of them occurring (c.f. the shaded values in the table above) is $pr^* = 0.15$. Fisher's accepted convention is to *reject* the null-hypothesis just in case $pr^* \leq 0.05$. Hence our null-hypothesis of the fairness of the coin is not rejected.

Some statisticians recommend 0.01 or even 0.001 as the critical *pr**. The adopted critical probability is called the *significance level* of the test, and the null-hypothesis is said to be *rejected at this significance* level if *pr** is less than or equal to it.

"The null-hypothesis is rejected at a significance level" is a technical expression, which means that the result of the experiment fall in a certain region (declared "the rejection region"). But what does it really say about the null-hypothesis? Today the standard view (introduced by Neyman) is that a rejection or non-rejection of a null-hypothesis is not an inductive inference, but just an instruction for inductive behaviour. If we behave according to the instruction, in the long run we shall reject a true hypothesis H, i.e. we shall make a type I error, no more than once in a hundred times, when significance level is 0.01.

We may also worry, as Neyman and Pearson did, about accepting a false hypothesis *H*, i.e. making a type II error. The probability of type II error is the probability of rejecting a true alternative hypothesis, let's call it H_a[9], by accepting the false *H*. The complement of the significance level of rejecting H_a is called the *power* of a test and, in this context, the significance level of rejecting *H* is called its *size*. An ideal would be to maximize the power and to minimize the size of a test. But that ideal is inconsistent. In most cases a contraction in size brings with it an expansion in power, and vice versa.

Apart from the volatility of what is declared to be "the rejection region", the incoherence of contracting the size and expanding the power of a test, and considering only one or two hypotheses[10], there are other problems with the frequentist approach.

For example, different random variables may by defined on an outcome space, not all of them leading to the same conclusion when used as a test-statistics in a significance test. This is the notorious problem of "which test-statistics to choose?"

There is also the problem of "the stopping rule". Consider again that a coin has been flipped 12 times, giving 3 heads and 9 tails. Is this the evidence that the coin is biased? With the data provided, the frequentists cannot even begin to answer this question. Namely, from these data it is not clear what the outcome space for the data is. If the frequentist is told that the experimenter's plan was to flip the coin 12 times, then analysis can proceed as above. But this is not the only way for these data to be produced. The experimenter may

[9] It means that Neyman and Pearson approach considers (only) two possible hypothesis *H* and H_a.

[10] In a Neyman-Pearson test you have to choose which of the two is your null-hypothesis and that choice may change which of the two is rejected (cf. Howson and Urbach 2006, 156).

have planned to flip the coin until he produced 3 heads, or until he becomes bored with the flipping. In this case, the outcome space will be different, even infinite or ambiguous, and the final result of the significance test may also be different. (cf. Loredo 1990, 109-110).
It seems that the frequentist approach is more subjective than the Bayesian approach, because the identifications of outcome spaces, the choices of test statistics, the declarations of rejection regions, the choices of null-hypothesis among alternatives, the contradictory choices between sizes and powers etc., depend on thoughts or even whims of the experimenter. Frequentists thus failed to solve the problems that motivated their approach, they even exacerbated them.

The basic problem of frequentist analysis is that, in search of a rejection region, it evaluates a single hypotheses by taking into account data that *could have happened*. But what this *possible data* have to do with our problem? We have made our experiment, we have got *the real data* and we want to estimate hypotheses given *this real data*.

The result of the frequentis analysis is a behavioural attitude towards a single hypotheses, prompted by data that could have occurred but did not.

To be more specific, for Bayesians there is the probability of H being in an interval:

$$pr(R/N - 2/\sqrt{N} \leq H \leq R/N + 2/\sqrt{N}\ ) \approx 95\%$$

For frequentists there is no such probability. There is only the inductive behaviour according to which, when we prove that:

$$pr(H - 2/\sqrt{N} \leq H \leq R/N + 2/\sqrt{N}\ ) \approx 95\%$$

then if we behave so that we accept R/N as our estimate of H, we may expect to be correct in 95% of our repeated behaviours.

The simple Bayesian 95% probability that your hypothesis is true, is replaced by the convoluted frequentist 95% chance of being correct in your repeated "as if my hypothesis is being true" behaviours. Why on earth would anybody do that? Is there not a better answer to the frequentist critique that degrees of belief are subjective and therefore incoherent and (even if they have some sense) that we do not know whether they satisfy probability axioms. A lot of people thought there is.

For John M. Keynes a degree of rational belief is a degree of partial entailment. Sometimes a conclusion *follows* from premises, but more often it only *partially follows* from them. As Keynes used to say, a conclusion stands in a relation of probability with premises (cf. Keynes 1921, 52-3). The relation is logical, and probability is just an extension of classical "true or false" logic. But how do we asses this logical relation of probability and, more specifically, how do we establish the probability axioms from this logical point of view? Keynes thought we simply perceive them as true, with some kind of logical intuition. (cf. Keynes 1921, 52-3).

Harold Jeffreys held the same logical attitude towards probability. He was one of the earliest critics of the frequentist statistics, but he did more than criticize. In his Jeffreys of 1939, he solved many statistical problems completely inaccessible to frequentists. That should have been a clear indication that he was on the right track, even though his first hundred pages devoted to logical arguments for probability axioms were not very successful. His work was rejected on philosophical grounds, as was Keynes'.

The most famous critic was Frank Ramsey. His answer to Keynes' position (that there are logical relations of probability and that these can be perceived with some kind of logical intuition) was simple and final. He does not perceive the probability relations of Mr. Keynes and, moreover, he suspects that others do not perceive them either (cf. Ramsey 1926, 161-2).

I suppose Ramsey was referring to the probability axioms:

(1) pr: $\mathbf{P} \times \mathbf{P} \rightarrow [0,1]$

(i. e. probability is a real number from [0,1] assigned to an ordered pair of propositions[11] in $\mathbf{P} \times \mathbf{P}$, which measures how probable is the first proposition given the second),

(2) $\vDash A \Rightarrow pr(A|I) = 1$

(i.e. probability of a logically valid proposition is 1, whatever background information I),

(3) $I \vDash -(AB) \Rightarrow pr(A \vee B|I) = pr(A|I) + pr(B|I)$

(i.e. if A and B a contradictory given I, then the probability of their alternation, given I, is additive),

(4) $pr(AB|I) = \mathrm{pr}(A|B, I) \cdot pr(B|I)$

(i.e. the probability of conjunction is quasi-multiplicative).

I do not think that he or anybody else has problems with inferences of probability theory. For example, that from (1)-(3) it follows that $pr(A) = 1 - pr(A)$; or that from (1)-(3) and $A \vDash B$ it follows that $pr(A) \leq pr(B)$; or that from (1)-(4) follows Bayes' theorem; or that from (1)-(4) and $A \vDash B$ it follows that $pr(B|A, I) = 1$; or that from $pr(A|A \rightarrow B, I) \neq 1$ it follows that $pr(A|B, A \rightarrow B, I) > pr(A|A \rightarrow B, I)$[12]; etc.

Probability axioms are problematic, not probability inferences.

[11] Or statements, or sentences, here it is not important.

[12] Given $A \rightarrow B$; A does not follow from B, but A is more probable given B.

Instead of vague logical intuitions Ramsey (and de Finetti) offered a definition of probability and proved that it satisfies probability axioms (1)-(4). Namely, the probabilities were defined as betting quotients and it was proved that the betting quotients are coherent (i.e. do not allow unfair bets; which we define below) iff they satisfy (1)-(4). It was a great success and the logical theory was forgotten.

It is often declared that this is a very surprising result, and that it is by no means obvious that betting quotients, if coherent, should obey the probability axioms (c.f. Gillies 2000, 66). I think it is obvious, and I am offering an obvious proof bellow. Before that, let me present a more standard version of the proof that coherence implies axioms (1)-(4)[13].

Think of me as a bookie. If you are willing to pay me M' for prospect of getting M if A happens, then your net-gain $G(A)$ in this bet on A, is $M–M'$ if A happens and $–M'$ if it does not happen[14]. If we define the value of A as $V(A) = 1$ if A happens, and as $V(A) = 0$ if A does not happen, then

$$G(A) = MV(A) – M'.$$

If you are willing to pay me M' for prospect of getting M only if a condition C is fulfilled and A happens (i.e. the bet is cancelled if C is not fulfilled), then your net-gain $G(A|C)$ in this bet on A under condition C, is

$$G(A|C)) = V(C)(MV(A) – M')$$

(i.e. the bet is cancelled by $V(C) = 0$ and otherwise it is like before).

What I am offering you, i.e. M, is your possible brutto-gain or the value of the bet. What you are willing to pay for the bet, i.e. M', is your betting expectation. Your betting quotient, in this particular bet on A, is defined as

$$q(A) = M'/M.$$

In this definition it is presupposed that your expectation M' is proportional to the value of the bet M, i.e. that your betting quotient depends only on the proposition A you are betting on and not on M. Real bets are definitely not like that and this is the soft point of Ramsey-de Finetti's argument. But let's go further with the argument.

Since $M' = q(A)M$, your net-gains (c.f. above) can be reformulated as
$G(A) = M(V(A) – q(A))$
$G(A|C) = V(C)M(V(A) – q(A))$.

[13] The converse does not interest us here. In philosophical literature the proof of the converse is usually omitted, and if not omitted it is often incorrect (cf. Gillies 2000, 60-64; Hacking 2001, 165-168). Gilles "proves" that each of (1)-(4) taken separately, implies coherence. Of course, it is nonsense, because then each of them, taken separately, implies all of them, since coherence implies all of them. Hacking's proof is similar, although it is not clear whether he is claiming a proof or just an idea of it.

[14] Of course, a negative gain is a loss, as a negative loss is a gain.

Your betting quotients are said to be coherent (and your bets to be fair) if I can not choose my *M*s so that I win whatever happens. Or, for that matter, that I can not choose them so that I lose whatever happens.[15] It means that the gain or the loss must depend on what happens. If it does not depend on what happens there is no gain and no loss, i.e. the gain is zero. More formally, *your betting quotients are coherent (and your bets are fair) if, and only if, G does not depend on V only if G = 0*.

Now, that we have defined coherence (fairness) we may prove that the probability axioms (1)-(4) follow from it.

Suppose that $q(A) \notin [0,1]$, i.e. $q(A) < 0$ or $q(A) > 1$. If $M > 0$, then $G(A) = M(V(A) - q(A)) > 0$ or $G(A) = M(V(A) - q(A)) < 0$ independently of the value $V(A)$. (If $M < 0$ then $G(A) < 0$ in the first case and $G(A) > 0$ in the second case.) Hence, $G(A) \neq 0$ independently of V, which is in contradiction with coherence. So, it is impossible that $q(A) \notin [0,1]$, i.e. $q(A) \in [0,1]$. This is our axiom (1).

If A is logically valid, i.e. $\vDash A$, then $V(A) = 1$ and

$$G(A) = M(V(A) - q(A)) = M(1 - q(A))$$

does not depend on V. By coherence it must be zero, i.e. $M(1 - q(A)) = 0$, from which (for $M \neq 0$) it follows that $q(A) = 1$. This is our axiom (2).

If you bet on A with quotient $q(A)$ for brutto-gain M_1, and on B with $q(B)$ for M_2, and on $A \vee B$ with $q(A \vee B)$ for M; then your total net-gain is

$$G = M_1(V(A) - q(A)) + M_2(V(B) - q(B)) + M(V(A \vee B) - q(A \vee B)).$$

Now, if from your background information it follows that A and B are mutually contradictory, then $V(A \vee B) = V(A) + V(B)$. If furthermore, your bet is such that $M_1 = M_2 = -M \neq 0$ then, for this particular bet,

$$G = Mq(A) + Mq(B) - Mq(A \vee B).$$

This gain does not depend on V so, by coherence, it must be zero,

$$M(q(A) + q(B) - q(A \vee B)) = 0.$$

It follows that $q(A \vee B) = q(A) + q(B)$. This is our axiom (3).

If you bet on AB with quotient $q(AB)$ for brutto-gain M, and on B with $q(B)$ for M_1, and on A under condition B with quotient $q(A|B)$ for brutto-gain M_2; then your total net-gain is

$$G = M(V(AB) - q(AB)) + M_1(V(B) - q(B)) + V(B)M_2(V(A) - q(A|B)).$$

[15] Of course, changing the signs of my *M*s turns one into another.

If your bet is such that $M_2 = -M \neq 0$ then, since $V(AB) = V(A)V(B)$, your net-gain is

$$G = -Mq(AB) + M_1V(B) - M_1q(B) + V(B)Mq(A|B).$$

If furthermore $M_1 = -Mq(A|B)$ then, fort this particular bet,

$$G = -Mq(AB) + Mq(A|B)q(B).$$

This gain does not depend on V so, by coherence, it must be zero,

$$M(-q(AB) + q(A|B)q(B)) = 0.$$

It follows that $q(AB) = q(A|B)q(B))$. This is our axiom (4).

This is a standard, maybe not extremely obvious proof. Now I present basically the same proof, which is trivial and completely obvious. Instead from coherence, I start from its simple consequence: for same bets you should have same expectations.[16] I define two bets as the same, if your brutto-gain in every possible situation is the same for both bets.

Example I: if A and B are mutually contradictory, then "to bet on $A \vee B$ for M" is the same as "to bet on A for M and to bet on B for M". Namely, there are only three possible situations $A\bar{B}$, $\bar{A}B$ and $\bar{A}\bar{B}$ (because AB is excluded) and in each of them your brutto-gain is the same for both bets (M if $\bar{A}B$ or $A\bar{B}$ and 0 if $\bar{A}\bar{B}$).

Example II: "to bet on AB for M" is the same as "to bet on B for M and then continue to bet on A for what you have got". Now, there are four possible situations, AB, $\bar{A}B$, $A\bar{B}$ and $\bar{A}\bar{B}$. In both bets you brutto-gains are the same in every of the four situations. They are, respectively: M, 0, 0, 0.

According to the example I, what you are willing to pay for bet on $A \vee B$ with brutto-gain M (if A and B are mutually contradictory), must be the same as what you are willing to pay for two bets, one on A for brutto-gain M and another on B for brutto-gain M. It means that

$$q(A \vee B)M = q(A)M + q(B)M,$$

and (for $M \neq 0$) it immediately follows that $q(A \vee B) = q(A) + q(B)$. This is our axiom (3).

According to the example II, what you are willing to pay for bet on AB for brutto-gain M, must be the same as what you one willing to pay for bet on B for brutto-gain M, which continues with the bet on A for what you have got. It means that

[16] It is a simple consequence of coherence. Namely, if you bet on A for M, with different expectations M_1 and M_2, i.e. with different quotients q_1 and q_2, then I may offer you M for one quotient and $-M$ for another. Your total net-gain in this compound bet is: $G = M(V(A) - q_1) - M(V(A) - q_2) = M(q_2 - q_1)$, which is independent of V and different from zero (because $q_1 \neq q_2$ and we can take $M \neq 0$). Hence, your quotients q_1 and q_2 are not coherent.

$$q(AB)M = q(A|B)(q(B)M),$$

and (for $M \neq 0$) it immediately follows that $q(AB) = q(A|B)q(B)$. This is our axiom (4).

I think the arguments for the axioms (1) and (2) were obvious. If your $q(A) > 1$ you obviously lose whatever happens, and if your $q(A) < 0$ you obviously win whatever happens. For valid proposition A, whatever happens, you obviously win if your $q(A) < 1$, (because valid A happens, whatever happens).

So, betting quotients quite obviously satisfy the probability axioms. There are no surprises about that. I would even suspect that these simple arguments for axioms (1)-(4) were well known from the beginnings of probability theory, because they are really extremely simple. Perhaps the reason they were not published (if they were not) is that the betting quotients were problematic, because they were not well defined.

And still today, they are not well defined. Presumption that M', which you are willing to pay for a bet, is proportional to M, which is the brutto-gain you are hoping for, is completely unsubstantiated. Even Ramsey was aware of that when he unsuccessfully tried to overcome the problem by introducing "ultimate goods" bets, instead of money bets, cf. Ramsey 1926, 173-176.[17]

The subjective Bayesianism of Ramsey and de Finetti did not solve the problems of the logical (or objective) Bayesianism of Keynes and Jeffreys. But Cox in the 1940' (cf. Cox 1946 and Cox 1961) provided the missing foundation for logical Bayesianism, which is today known as Bayesian probability theory, or BPT for short.[18] The intuitive appeal of BPT[19], the huge amount of successful results and its rigorous mathematical foundation provided by Cox and others, make it the best theory of probable inference we have. Hence, it is quite strange that it is not even mentioned in the recent philosophy textbooks devoted to the probable inference (e.g. Gillies 2000, Hacking 2001 and Mellor 2005). It is mentioned in Bayesian textbooks, e.g. Howson and Urbach 2006 which explicitly declares it as the best approach ("which begs fewest questions of all"), but even then the Cox's mathematical foundation is omitted because "it requires fairly sophisticated mathematics".

Although mathematics is a bit sophisticated, I will present a variant of the crucial proof; especially because printed proofs are rare, quite often not completely correct and they almost always presuppose more assumptions then necessary (cf. Cox 1946, Cox 1961 and Jaynes 2003).

Cox's idea was to start from the notion of the *plausibility of a proposition A given a proposition I as known*, which is denoted by $A|I$, and from some properties that these plausibilities have to satisfy. I will use the following properties.

[17] Gillies proposes, following early de Finetti, that we should use money bets with appropriately selected stakes, with no real explanation, not to talk about a definition, of appropriateness, cf. Gillies 2000, 57.

[18] Jaynes call it "probability theory as the logic (of science)", cf. the title of Jaynes 2003.

[19] Just compare the Bayesian vs. frequentist analysis of the coin fairness problem above.

(P1) $\quad \bot|I = o \leq A|I \leq I|I = j$

(i.e. plausibilities are real numbers between the minimum o, which is the plausibility of a logical contradiction and the maximum j, which is the plausibility of a logical truth).

(P2) $\quad I \models -(AB) \Rightarrow A \vee B|I = AI(A|I, B|I)$

(i.e. if A and B are mutually contradictory given I, then the plausibility of their alternation "A or B", given I, is determined by the plausibility of A, given I, and the plausibility of B, given I; the determination function AI may depend on I).

(P3) Functions $\mathcal{A}_I$ are continuous and strictly increasing in both arguments.

(P4) $\quad AB|I = KI\,(A|BI, B|I)$

(i.e. the plausibility of the conjunction "A and B" given I, is determined by the plausibility of B, given I, and the plausibility of A, given B and I; the determination function $\mathcal{K}_I$ may depend on I).

(P5) $\quad AI|I = A|I.$

From these properties, using the logical rule of the replacement of equivalents (e.g. from $(A \vee B)C \equiv AC \vee BC$; it follows $(A \vee B)C|I = (AC \vee BC)|I$, from $I \equiv II$ it follows $A|I = A|II$ etc.) it is possible to prove that there exists a continuous and strictly increasing function $f(x)$ such that $f(o) = 0, f(j) = 1$ and that for every proposition I:

$$\mathcal{A}_I(x,y) = f(f^{-1}(x) + f^{-1}(y)), \qquad \mathcal{K}_I(x,y) = f((f^{-1}(x) \cdot f^{-1}(y)).$$

This is equivalent to:

$$f^{-1}\mathcal{A}_I(x,y) = f^{-1}(x) + f^{-1}(y), \qquad f^{-1}\,\mathcal{K}_I(x,y) = f^{-1}(x) \cdot f^{-1}(y).$$

Hence, if we define $pr(A|I) := f^{-1}(A|I)$ and substitute $A|I$ for x and $B|I$ for y, we get:

$$pr(A \vee B|I) = pr(A) + pr(B), \qquad pr(AB|I) = pr(A|BI) \cdot pr(B|I).$$

The conclusion is: if plausibility satisfies (P1)-(P5) then there is a measure of plausibility which satisfies our probability axioms (1)-(4). Namely, every continuous and strictly increasing function of plausibility $A|I$ could be a measure of plausibility, as any other. Out of all these possible measures we chose $pr(A|I)$, not because it is more "correct" but because it is more convenient, i.e. the quantities pr obey the simplest rules of combination: the normality condition (1), (2), the sum rule (3) and the product rule (4).

The situation is analogous to that in thermodynamics (cf. Jaynes 2003, 42), where out

of all temperature scales (which are continuously increasing functions of each other) we choose Kelvin scale because it is more convenient, i.e. the laws of thermodynamics take the simplest form in this scale. Similarly, in mathematics, out of all angle scales we choose the radians as the most convenient; e.g. $d\sin x / dx = \cos x$ only if x is measured in radians.

Before I present the proof of this crucial result (usually called Cox's theorem) I should address one more problem. Why plausibilities should obey the properties (P1) – (P5)?

Desideratum (P1) is that degrees of plausibility are represented by real numbers (with the minimum which represent the plausibility of contradictions and the maximum which represent the plausibility of tautologies). I believe it is possible to prove that this desideratum follows from more elementary desiderata that (i) degrees of plausibility should be linearly ordered (i.e. that they are transitive, antireflexive and universally comparable), and that (ii) continuous, strictly increasing, commutative and associative operations (representing degrees of plausibility of conjunctions and alternations c.f. below) are definable on these degrees.[20]

In the moment it is just a conjecture, and I will not further discuss (P1).

As a first point about (P2), note that, given the knowledge of I, the process of deciding that $A \vee B$ is true, can be broken down into elementary decisions about A and B separately:

(i) Decide that A is true. $(A|I)$
(ii) Decide that B is true. $(B|I)$

In each step I indicate (in the brackets) the plausibility corresponding to that step. These two decisions completely determine our decision about $A \vee B$. More formally:

$$A \vee B|I = \mathcal{A}_I(A|I, B|I),$$

which is our (P2). Of course, if the plausibility in any of the two steps is increased then the combined plausibility of $A \vee B$ is increased, which is our (P3).

As for (P4), note that, given the knowledge of I, the process of deciding that AB is true can be broken into elementary decisions about A and B separately, in the following way:

(i) Decide that B is true. $(B|I)$
(ii) Having accepted B as true, decide that A is true. $(A|BI)$

Equivalently

(i') Decide that A is true. $(A|I)$
(ii') Having accepted A as true, decide that B is true. $(B|AI)$

Regarding the first procedure, in order for AB to be true it is necessary that B is true. So, $B|I$ is to be decided. Further, if B is true it is necessary that A is true. So, $A|BI$ is to be decided,

[20] The proof would be on adaptation of Hölder-Cartan proof that every linearly ordered group without minimum is embeddable in $\mathbb{R}$, and isomorphic to $\mathbb{R}$ if it is Dedekind continuous.

too. These two decisions completely determine our decision about AB. More formally:

$$AB|I = \mathcal{K}_I(B|I, A|BI),$$

which is our (P4)[21]. Of course, (P5) is self-evident.

If we define $x{:=}A|II$ and take into account that $j = I|I$, $I \equiv II$ and $AI \equiv A$, given I, then

$$\mathcal{K}_I(x,j) = \mathcal{K}_I(A|II, I|I) = AI|I = A|I = A|II = x.$$

Similarly,

$$\mathcal{K}_I(j,x) = \mathcal{K}_I(AI|AI, A|I) = AIA|I = AI|I = A|I = x.$$

In other words j is a neutral element for $\mathcal{K}_I$ (for every I).

It is as easy to prove that o is a neutral element for $\mathcal{A}_I$ (for every I):

$$\mathcal{A}_I(o,x) = \mathcal{A}_I(\bot|I, A|I) = (\bot\vee A)|I = A|I = x,$$

and similarly

$$\mathcal{A}_I(x,o) = \mathcal{A}_I(A|I, \bot|I) = (A\vee\bot)|I = A|I = x.$$

That $\mathcal{A}_I$ is associative, is proved in the following way:

$$\mathcal{A}_I(\mathcal{A}_I(x,y),z) = \mathcal{A}_I(\mathcal{A}_I(A|I,B|I),C|I) = \mathcal{A}_I((A\vee B|I,C|I) = ((A\vee B)\vee C)|I = (A\vee(B\vee C)|I$$
$$= \mathcal{A}_I(A|I, (B\vee C)|I) = \mathcal{A}_I(A|I, \mathcal{A}_I(B|I, C|I)) = \mathcal{A}_I(x, \mathcal{A}_I(y,z)).$$

It is even easier to prove that it is commutative, i.e. that:

$$\mathcal{A}_I(x,y) = \mathcal{A}_I(y,x).$$

Furthermore, $\mathcal{K}_I$ (is distributive with respect to $\mathcal{A}_I$, i.e. for every I and C:

$$\mathcal{K}_I[\mathcal{A}_{CI}(x,y),z] = \mathcal{A}_I[\mathcal{K}_I(x,z), \mathcal{K}_I(y,z)].$$

Namely, $(A\vee B)C \equiv AC\vee BC$, hence

[21] In many discussions of uncertain reasoning (most prominently in AI discussions of fuzzy logics) it is quite common to suppose that $AB|I = \mathcal{K}(A|I, B|I)$, with various candidates for $\mathcal{K}$, although it is evident that no relation of this form is generally valid. (So, the discussions based on this assumption are completely futile.) For example, the plausibility of the next person being female and the plausibility of the next person being male could be about 50%, although the plausibility of the next person being male and female is zero. On the other hand the plausibility of the next person being older than 20 years and the plausibility of the next person being younger than 60 years could also be about 50%, although, in this case, the plausibility of the next person being older than 20 years and younger than 60 years should not be zero.

$(A \vee B)C|I = (AC \vee BC)|I$.

It follows that

$$\mathcal{K}_I((A \vee B)|CI,\ C|\ I) = \mathcal{A}_I(AC|I,\ BC|I),$$

which means that

$$\mathcal{K}_I[\mathcal{A}_{CI}(A|CI,\ B|CI), C|I] = \mathcal{A}_I[\mathcal{K}_I\ (A|CI,\ C|I), \mathcal{K}_I(B|CI,\ C|I)].$$

If we define $x := A|CI$, $y := B|CI$, and $z := C|I$, we finally have

$$\mathcal{K}_I\ [\mathcal{A}_{CI}(x,y),z] = \mathcal{A}_I[\mathcal{K}_I(x,z),\ \mathcal{K}_I(y,z)],$$

which was to be proved.

If we substitute $z = j$ in the above formula of distributivity we get:

$$\mathcal{K}_I\ [\mathcal{A}_{CI}(x,y),j] = \mathcal{A}_I[\mathcal{K}_I(x,j),\ \mathcal{K}_I(y,j)],$$

which simplifies to

$$\mathcal{A}_{CI}(x,y) = \mathcal{A}_I(x,y),$$

(because j is a neutral element of $\mathcal{K}_I$). We may repeat this while exchanging C and I and get

$$\mathcal{A}_C(x,y) = \mathcal{A}_{CI}(x,y) = \mathcal{A}_I(x,y).$$

The conclusion is that $\mathcal{A}$ does not depend on I. (That $\mathcal{K}$ does not depend on I, will follow from what follows.)

So far we have proved that:

$$o \le \mathcal{A}(x,y) \le j \qquad o \le \mathcal{K}(x,y) \le j$$
$$\mathcal{A}(x,o) = \mathcal{A}(o,x) = x$$
$$\mathcal{K}(x,j) = \mathcal{K}(j,x) = x$$
$$\mathcal{A}(\mathcal{A}(x,y),z) = \mathcal{A}(x,\mathcal{A}(y,z))$$
$$\mathcal{K}[\mathcal{A}(x,y),z] = \mathcal{A}[\mathcal{K}(x,z),\ \mathcal{K}(y,z)]$$

(where $\mathcal{K}$ could be any $\mathcal{K}_I$).

In what follows the binary operation $\mathcal{A}(x,y)$ is renamed $x \circ y$. This operation is defined on $[o,j]$, it is continuous, associative, commutative, strictly increasing in both arguments, and it has a neutral element o (in algebra usually called zero).

For any number u, such that $o < u < j$, we have

$$o < u < u\circ u < u\circ u\circ u < \dots$$

(because $\circ$ is strictly increasing). Hence, if we define $\underline{1}\,u := u$, $\underline{2}\,u := u\circ u$, $\underline{3}\,u := u\circ u\circ u$, etc. We immediately see that $\underline{m}u < \underline{n}u$, whenever $m < n$.[22] Furthermore, if $u < v$ then $u\circ u < v\circ v$, $u\circ u\circ u < v\circ v\circ v$ etc. (because $\circ$ is strictly increasing). Hence, $\underline{m}u < \underline{m}v$, whenever $u<v$. In other words, the two valued function $\underline{m}u$ is continuous (because $\circ$ is continuous) and strictly increasing in both arguments ($m\in N$ and $u\in[o,j]$).

If we fix the first argument, i.e. m, we get the strictly increasing function $\underline{m}u$, of one argument u. Because of

$$j = \lim_{u\to j} u \le \lim_{u\to j} \underline{m}u \le j,$$

it follows that this function maps $[o,j]$ onto $[o,j]$. And it makes it in 1-1 fashion, because it is strictly increasing. Hence, for every $u\in[o,j]$ there is exactly one $v\in[o,j]$ such that $\underline{m}v = u$. We symbolize this v with

$$v := \frac{u}{\underline{m}}.$$[23]

Now we are ready to define our function f. For every $m/n\in[0,1]$

$$f\left(\frac{m}{n}\right) := \underline{m}\,\frac{j}{\underline{n}}.$$

Of course, we have to prove that for every km/kn

$$\underline{km}\,\frac{j}{\underline{kn}} = \underline{m}\,\frac{j}{\underline{n}}$$

It is easy if we note that $j/\underline{kn} = (j/\underline{n})/\underline{k}$, (this follows immediately from $\underline{knz} = \underline{k}(\underline{nz})$, which is obvious). Namely, $\underline{km}(j/\underline{kn})$ is equal to

$$\frac{j}{\underline{kn}}\circ\dots\circ\frac{j}{\underline{kn}} \qquad (km \text{ times}),$$

Which is equal to

$$\left(\frac{j}{\underline{kn}}\circ\dots\right)\circ\dots\circ\left(\frac{j}{\underline{kn}}\circ\dots\right) \qquad (m \text{ brackets; } k \text{ times in brackets}).$$

---

22 Note that $\underline{n}u$ is not nu. By underlining n we stress the difference.

23 Note that $u/\underline{n}$ is not u/n. By underlining n we stress the difference.

But then, each bracket is equal to $\underline{k}((j/\underline{n})/\underline{k})$, which is j/n. Hence, the whole value is equal to $\underline{m}(j/\underline{n})$, which was to be proved.

Now it is easy to prove some important properties of f. First of all, f is strictly increasing,

$$m_2 > m_1 \Rightarrow f\left(\frac{m_2}{n}\right) > f\left(\frac{m_1}{n}\right),$$

because $\circ$ is strictly increasing and $j/\underline{n} > o$:

$$f\left(\frac{m_2}{n}\right) = \left(\frac{j}{\underline{n}} \circ \ldots \circ \frac{j}{\underline{n}}\right) \circ \left(\frac{j}{\underline{n}} \circ \ldots \circ \frac{j}{\underline{n}}\right) \circ (o \circ \ldots \circ o) = f\left(\frac{m_1}{n}\right),$$

(in the first and the third bracket we have m_1 times $j/\underline{n}$; in the second and the forth bracket we have $(m_2 - m_1)$ times $j/\underline{n}$ and o). Furthermore,

$$f\left(\frac{m_1}{n}\right) \circ f\left(\frac{m_2}{n}\right) = \underline{m}_1 \frac{j}{\underline{n}} \circ \underline{m}_2 \frac{j}{\underline{n}} = (m_1 + m_2) \frac{j}{\underline{n}} = f\left(\frac{m_1 + m_2}{n}\right)$$

i.e. f is $\circ$-additive.

So, f is strictly increasing, $\circ$-additive function defined on rational numbers from [0,1], such that $f(0) = o$ and $f(1) = j$. There is the unique continuous $\circ$-additive extension of this function to the real numbers from [0,1] (remember, we presupposed that $\mathcal{A}$, which means $\circ$, is continuous). This extension, which we continue to denote f, is also $\circ$-additive:

$$f(x) \circ f(y) = f(x+y).$$

If we substitute $u = f(x)$ and $v = f(y)$ this is equivalent to

$$u \circ v = f(f^{-1}(u) + f^{-1}(v)).$$

Let us pause and state what we have proved so far.

There is a continuous and strictly increasing function f, defined on [0,1], such that f(0) = o, f(1) = j and

$$\mathcal{A}(u,v) = f(f^{-1}(u) + f^{-1}(v)).$$

If we substitute this result into distributive law

$$\mathcal{K}_1[\mathcal{A}(x,y),z] = \mathcal{A}[\mathcal{K}_1(x,z), \mathcal{K}_1(y,z)]$$

(which we proved above), we get:

$$\mathcal{K}_1[f(f^{-1}(x)+f^{-1}(y)),z] = [f^{-1}(\mathcal{K}_1(x,z))+f^{-1}(\mathcal{K}_1(y,z))].$$

Further, if we denote $f^{-1}(x)$ by u, and $f^{-1}(y)$ by v, and apply f^{-1} to both sides of the above equations, then

$$f^{-1}[\mathcal{K}_1(f(u+v),z)] = f^{-1}[\mathcal{K}_1(f(u),z)] + f^{-1}[\mathcal{K}_1(f(v),z)].$$

We further simplify by defining $M(u,z) := f^{-1}[\mathcal{K}_1(f(u),z)]$, to get

$$M(u+v,z) = M(u,z) + M(v,z).$$

It means that M is additive in the first argument, from which it follows it is linear in the first argument (because it is continuous):

$$M(u,z) = k(z)u.$$

From the defining equation $M(u,z) = f^{-1}[\mathcal{K}_1(f(u),z)]$ it follows:

$$\mathcal{K}_1((f(u),z) = f(M(u,z)) \text{ i.e.}$$

$$\mathcal{K}_1(t,z) = f(M(f^{-1}(t),z)) = f(k(z)f^{-1}(t)).$$

Substituting j for t we get:

$$z = \mathcal{K}_1(j,z) = f(k(z)f^{-1}(j)) = f(k(z)),$$

from which it immediately follows that $k(z) = f^{-1}(z)$. Hence, (for every I),

$$\mathcal{K}_1(t,z) = f(f^{-1}(t)\cdot f^{-1}(z)).$$

Let us summarize what we have proved so far.

There is a continuous and strictly increasing function f, defined on [0,1], such that f(0) = o, f(1) = j and

$$\mathcal{A}(u,v) = f(f^{-1}(u)+f^{-1}(v)).$$
$$\mathcal{K}(u,v) = f(f^{-1}(u)\cdot f^{-1}(v))$$

or equivalently

$$f^{-1}(\mathcal{A}(u,v)) = f^{-1}(u)+f^{-1}(v)$$
$$f^{-1}(\mathcal{K}(u,v)) = f^{-1}(u)\cdot f^{-1}(v).$$

If we substitute concrete plausibilities for u and v we get:

$$f^{-1}(\mathcal{A}(A|I, B|I)) = f^{-1}(A|I)+f^{-1}(B|I)$$
$$f^{-1}(\mathcal{K}(A|I, B|AI)) = f^{-1}(A|I)\cdot f^{-1}(B|AI).$$

Now we can define probability function *pr* by $pr(A|I) := f^{-1}(A|I)$ and finally get:

$$pr(A \vee B|I) = pr(A|I) + pr(B|I),$$

$$pr(AB|I) = pr(A|I) \cdot pr(B|AI)$$

(of course, we presupposed that $I \models -(AB)$).

At the end we should address Halpern counterexample to Cox's theorem, cf. Halpern 1999. The crucial point is that the counterexample presupposes there is only finitely many probability values. But it is trivially true that for every *m/n* there is a proposition with probability *m/n*, e.g. "from urn with *m* white balls and *n–m* non-white balls a white ball will be drawn". Hence, it is as relevant to probability as any statement about finite structures is to arithmetic. You may explore finite structures and finite probability spaces and these are important subjects, but they do not provide us with counterexamples to arithmetic or probability. (For example, in a finite field of residues modulo 7 there is only finitely many primes but this has nothing to do with Euclid's theorem on infinitude of primes in ordinary arithmetic.)

REFERENCES

Cox, R. T. 1946. Probability, Frequency and reasonable Expectation. *American Journal of Physics* 14: 1-13.

Cox, R. T. 1961. *The Algebra of probable Inference*. Baltimore: John Hopkins Press.

Gillies, D. 2000. *Philosophical Theories of Probability*. London: Routledge.

Hacking, I. 2001. *Probability and inductive Logic.* Cambridge: Cambridge Univ. Press.

Halpern, J. Y. 1999. Cox's theorem revisited. *Journal of artificial intelligence research* 11: 429-35.

Howson, C. & Urbach, P. 2006. *Scentific reasoning – The Bayesian Approach.* Chicago: Open Court.

Jaynes, E. T. 2003. *Probability Theory – The Logic of Science*. Cambridge: Cambridge Univ. Press.

Jeffreys, H. 1939. *Theory of Probability*. Oxford & The Clarendon Press, 1961.

Keynes, J. M. 1921. *A Treatire on Probability*. London: Macmillan, 1963.

Loredo, T. J. 1990. From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophisics. In *Maximum Entropy and Bayesian Methods*, ed. P. F. Fongere. Dordrecht: Kluwer Academic Publishers.

Mellor, D. H. 2005. *Probability – A Philosophical Introduction*. London: Routledge.

Ramsey, F. P. 1926. Truth and Probability. In *The Foundations of mathematics and other Logical Essays*. London: Routledge, 1931.

Sivia, D. S. 1996. *Data Analysis*. Oxford: Oxford Univ. Press.

Received: January 23, 2013
Accepted: April 21, 2014

University of Zagreb
Faculty of Mechanical Engineering and Naval Architecture
Ivana Lučića 5
10002 Zagreb, Croatia
zvonimir.sikic@gmail.com

EuJAP | Vol. 10 | No. 1 | 2014
UDK 165.1/.2
519.2
1 Van Inwagen, P.
53-05 Hawking, S. W.

THE PROBABILITY OF THE POSSIBLE

RON WILBURN
University of Nevada

ABSTRACT

In "Why is There Anything at All?" Peter van Inwagen argues that even though it was never necessary that concrete beings existed, it was always maximally probable – just short of necessity – that they did (van Inwagen, 1996). I argue that van Inwagen's argument fails, albeit for an interesting reason which has remained so far unnoticed in the literature: there is a critical tension between two of its premises, both essential to its soundness, concerning the nature of comprehensively specified possible worlds. I summarize van Inwagen's argument, develop this objection, and then describe more problems which invariably accrue when we try to ascribe probability values to possible worlds.

Keywords: existence, possible worlds, probability, van Inwagen, Hawking

Introduction

In his paper "Why is There Anything at All?" van Inwagen argues that it is as improbable as improbable can be that the actual world might have been uninhabited by concrete objects (van Inwagen, 1996). Even though this argument fails, it fails for an interesting reason: an unhealthy tension obtains between two of its premises, both essential to its soundness, concerning the nature of fully specified possible worlds. I summarize van Inwagen's argument, develop the aforementioned objection, and then detail a more general objection to the project of ascribing probability values to possible worlds.

Van Inwagen's Argument:

Van Inwagen's argument invokes four premises. (99)

(1) There are some beings
(2) If there is more than one possible world, there are infinitely many.
(3) There is at most one possible world in which there are no beings.
(4) For any two possible worlds, the probability of their being actual is equal (I
will sometimes call this the "equi-probability assumption").
(5) It is "as unlikely as unlikely could be" that the world could have had no concrete
occupiers.

Herein, I treat premises (1) and (2) as unproblematic and deal with them only in passing. My critical scrutiny is reserved for (3) and (4). This scrutiny proceeds in two stages. First, I explicate these premises' meanings and motivations. Second, I explore a joint weakness that arises from an essential tension between them. Brief moralizing then follows.

First, let's consider the two unproblematic premises. Premise (1), the assumption that there are some beings, is certainly "a safe enough assumption," based on empirical observation: we know that the actual world is occupied because we can see that we and other items occupy it. (100) Premise (2), the claim that if more than one possible world exists, then infinitely many do, results from our ability to articulate an infinite number of alternative descriptions of reality, each corresponding to a distinct way that things might be. Van Inwagon writes that it would be "bizarre" to suppose that properties which vary in magnitude or across an indefinite range of dimensions define only a limited number of worlds. He writes, "if there is more than one possible world, then things can vary; and it seems bizarre to suppose, given the kinds of properties had by the things we observe, properties that seem to imply a myriad of dimensions along which these things could vary continuously, that there might be just two or just 17 or just 510 worlds." (56) The shape of my coffee cup, for instance, could presumably vary in an infinite number of ways, and each of these different ways would a difference-making feature of a distinct possible world.

As for the second two, much more problematic, premises, let's first consider (3), i.e., the assumption that there is at most one possible world in which there are no beings. This premise emerges from van Inwagen's contention that any two unoccupied worlds are identical, since all worlds share exactly the same abstract occupiers. Van Inwagen makes this assumption because he takes abstract objects ("numbers, pure sets, 'purely qualitative' properties and relations, possibilities, possible worlds themselves") to be common to all possible worlds, and thus not effective difference-makers between them. Moreover, he takes the actual concerns expressed by "why anything exists" queries to regard concrete, rather than abstract, occupiers of reality. Philosophers worry, he tells us, that there might have been "no physical things, no stuffs, no events, no space, no time, no Cartesian egos, no God [and presumably no fields, forces and the like], etc." It is this possibility, not one in which, say, the number 2 doesn't exist, that fuels our anxieties about existential contingency.[1] It is from this assumption that van Inwagen concludes that there can be but one empty possible world. For if there is only one way in which a world can be empty (by containing no concrete items), and being empty identifies a world as the world that it is, then no two empty worlds can be distinct.

Premise (4), the claim that the probability of any two worlds' obtaining must be equal,

[1] Even though van Inwagen does not claim that this taxonomy of concrete occupiers is comprehensive, it appears to be so. Indeed, it seems redundant. To the extent that we view physical objects relativistically, as extended series of events, or worldbraids, and the properties of space/time as determined by relational configurations of such items, "objects, space, time and stuffs" fail to exist independently of events. The alleged existence of abstract objects in all possible worlds can perhaps best be understood by reference to their contrast with physical objects in this respect. Abstract objects exist outside of spacetime in the sense that they are existentially independent of the worldbraid material occupiers of spacetime. Thus, their existence cannot vary across possible worlds as a function of differences in these worlds' physical (event) occupiers.

is the contention that van Inwagen spends the greatest time and effort defending. It is, he concedes, "the one that people are going to want to dispute" (101). Thus, van Inwagen's defense of (4) is considerably more complicated than those of (1) – (3), requiring more in the way of both setup and explication. By way of setup, he asks us to imagine some system of objects and associated abstract states. which the s is either in or not. These states, he suggests, behave logically very much like propositions, allowing us to treat possible worlds as state sets, the component states of which constitute the various details of the world in question. Van Inwagen asks us to think of possible worlds as sets of "fully consistent and maximal" possible states of affairs which, as a function of their maximality, remain "isolated" beyond the reach of "pre-cosmic selection machines." Let's pause briefly to explicate clarify these locutions.

To do this, let's simplify things, paraphrasing to eliminate technical formulation whenever doing so doesn't distort van Inwagen's fundamental intent. A worlds is "maximal," on his telling, if it consists of fully specified state sets, where said specification guarantees that *any* given possible state of affairs either obtains in that world or doesn't. A world is "consistent" if it does not include states which stand in logical tension with each other; if it involves no states which both do and do not obtain. A world is isolated "if no facts about objects external to it could influence it." Finally, a "pre-cosmic selection machine" (henceforth, "selector mechanism") is a principle which operates from outside all possible worlds, and functions to determine the likelihood of any given world's obtaining over its alternatives. [2] Talk of machines here is ambiguous. On one hand, these items might be construed as statistical laws, just as Turing machines are best construed as abstract programs. On the other hand, they might be construed as concrete machinery which functions to express or enforce such laws, just as (finite approximations of) Turing machines might be construed as desktop computers. For our purposes here, it seems best to understand selector mechanism in the former terms, since translation into the latter terms is always available (e.g., talk of a law which dictates that the best of all possible worlds must be actual can always be paraphrased, if one chooses, into talk about Leibniz' omnibenevolent God).

With these locutions in hand, the rationale behind premise (4) becomes clear. Because possible worlds are maximal, i.e., completely inclusive with respect to all possible yfacts, each is an isolated system in the sense that there are no unspecified details about it yet to be determined by other states that do not already partially constitute it.[3] In particular, no selector

[2] Van Inwagen's choice of locution ("pre-cosmic") is, of course, particularly puzzling when we ask what the tense designation (if that's what it's supposed to be) is intended to tell us, given that possible worlds encompass entire world histories.

[3] Note that in announcing my decision to treat premise (2) as unproblematic I am not committing myself to the claim that it is unproblematic. Jack Macintosh has pointed out a potential problem from which it suffers that we are only now in a position to articulate. As we have seen, van Inwagen argues for an infinitely large class of possible worlds by noting the multitude of descriptions that we can generate by imagining quantitative variations of various magnitudes across an indefinite range of dimensions (again, the shape of my coffee cup could seemingly vary in an infinite number of ways, and each of these different ways seemingly defines a distinct possible world.) Such a class is not merely infinite. It is non-denumerably infinite: it is uncountable because additional members, generated by the same imaginative procedure, always wait in the wings to be included within it. But mightn't this fact suggest that possible worlds cannot be "maximal" in the way that van Inwagen takes them to be? For if an infinity of possible worlds is demonstrable through a procedure which shows them to be non-denumerably infinite, then mightn't the specification of each such world re-

mechanisms exercise such determining influence. This is because selector mechanisms, if there are such, exist (by stipulation) only externally to particular worlds, even though such externality is rendered impossible by the maximality which van Inwagen supposes these worlds to have. For any given state, the maximality of each possible world guarantees that this world either is or isn't in that state. Because possible worlds are inclusive of all the ways the world might be, there is no neutral logical space between them from which selector mechanisms might operate.

Van Inwagen makes one final inference before announcing his main conclusion. "For any system of objects with maximal states," he writes, the maximal states of the system should be regarded as equally probable, provided that the system is isolated" (104). Van Inwagen's intended reason for this is also clear: if there can be nothing external to maximal possible worlds which might privilege some of them over others, we should treat them all as having an equal probability of obtaining. "If a system is isolated," van Inwagen tells us, "then any two of its maximal states are of equal probability. But then we have an argument for the conclusion that any two possible worlds are of equal probability: 'Reality' is an isolated system, and possible worlds are maximal states of Reality" (105-6).[4]

As an aside, I doubt that this defense of the equiprobability assumption is cogent. On the face of it, it is difficult to see why different possible worlds are equally probable simply by virtue of their mutual isolation. For, doesn't such equiprobability also require the operation of a selector mechanism able to dictate that isolation ensures equiprobability? Van Inwagen assumes here that equiprobability is a natural *default state* for mutually isolated possible worlds. However it is hard to see why this should be the case. The default state could be one in which, say, our own actual world enjoys probabilistic advantage; to assume that it doesn't requires us to assume that some selector mechanism be operative.

However, let's ignore this complication. For, as noted at the outset, my concern in this paper is not to argue against any of van Inwagen's premises considered in isolation; it is rather to argue for the existence of a critical tension between them.

From his claim that isolation ensures equiprobabity, here van Inwagen proceeds to his

quire us to specify, across a non-denumerable set of state description sentences, whether or not those sentences are true?

Is this a problem for van Inwagen? I am not quite sure. For the mere fact that a non-denumerable set of potential descriptions exist, each picking out a unique possible world, does not in itself imply that we can only specify the character of each such world by assigning truth-values across non-denumerable sets of state description sentences. On the face of it, we would seem to be able to conclude from the fact that the base of my cup might be round or square or oval or slightly differently oval, etc., that a non-denumerably infinite number of worlds exists without concluding anything at all concerning the number of assertions that we must make in order to specify the character of each of these particular worlds. Perhaps this issue is only decided when we take a position on the existence of negative properties (see footnote (6)). This would tell us whether or not we could specify the conditions defining a world as those in which, e.g., the base of my cup is round as opposed to those in which the base of my cup is round, and not square, and not oval, etc.

[4] Premise (1), which maintains that some beings exist, functions in the background of this argument, but in a quiet way. Van Inwagen's is concerned to make it clear from the outset that occupied worlds are possible. This is a reasonable precondition that must be satisfied for the other premises of the argument to do any effective work.

major conclusion. Because possible worlds are infinite in number (premise 2) and all are equally probable (premise 4) and only one of them is devoid of beings (premise 3), the world in which there is nothing gets swamped by the worlds in which there is something. There is only one empty possible world to be actualized; however there are an infinite number of equally probable occupied alternatives. Thus, while it is not impossible that nothing exists, it is as improbable as improbable can be.[5]

Problems

Van Inwagen's argument has been the object of criticism before, generally by way of objection to *either* of premises (3) or (4). Again, however, my strategy here is somewhat different. It exploits a crucial tension that arises *between* (3) and (4). These two premises, I maintain, are in conflict because they are motivated by contradictory intuitions regarding the nature of "maximal" possible worlds.

Remember again the rationale behind (4), i.e., the claim that all possible worlds are equiprobable. Because each possible world is maximal, there are no unspecified details about it that might be determined by other states that do not already help constitute it. This is what makes it isolated, i.e., beyond the influence of any conceivable selector mechanisms able to render some worlds more or less (or equally) probable than others. In contrast with this, however, the rationale behind (3), i.e., the claim that there is at most one empty possible world, is that all that matters to making a world the world that it is are its concrete occupiers. Thus, it is only through the differential existence of such concrete items that differences can accrue between worlds. But possible worlds, so conceived (by (3)), are surely not maximal in the sense required by (4). This is because premise (4)'s conception of maximality is one on which the distinguishing role, not merely of concreta, but of laws, including those very general laws defining selector mechanisms," must be taken into account.

To see why this is the case, let's think more closely about selector mechanisms. If they existed, what would they be like? We have construed them as very general statistical laws rendering some worlds more or less (or equally) probable than others. What might examples of such laws look like? One candidate van Inwagen considers is a variation of Leibniz' simplicity imperative: simpler worlds are more likely than complicated ones. Now, of interest to us here is the fact that in neither Leibniz' nor van Inwagen's hands is the criteria of such simplicity merely ontological; it is nomological also. In Leibniz' case,

[5] Manson, Neil A., 2011, "No Chance for Nothing," American Philosophical Association Pacific Division Conference, San Diego, CA. Manson poses a significant challenge to van Inwagen's argument by objecting to premise (4) on the grounds that it violates countable additivity, the principle dictating that the collective probabilities of each individual world's obtaining must equal the overall probability of some or another of these worlds obtaining. This is a compelling criticism, but not one that I will pursue here. The problem he highlights is simply this. Since these worlds exhaust all the ways that reality could be, the overall probability that one or another of them obtains must be 1. But this poses a dilemma once we note that the sum that we get when we add together the probability values of each individual world's obtaining is either less or more than 1. We cannot ascribe probabilities in such circumstances without giving up the idea that various possibilities sum together to make up 100 percent of the original probability space. Thus, we cannot ascribe probabilities across infinite ranges of possible worlds.

it is arguable that the pertinent criterion of "simplicity" is primarily nomological: there is little significant respect in which the self-probabilifying simplicity of a world consists in its containing fewer constitutive items or fewer types of items than its modal alternatives. Rather, Leibniz' best of all possible worlds is that which is "richest in phenomena," but "simplest in hypotheses," where these hypotheses are best understood as articulating the law-like regularities governing phenomena (Leibniz, 1992, sec. VI).

In van Inwagen's hands, the simplicity imperative assumes a different form. One world is simpler than another, he suggests, if it is fully specifiable by a more minimal description, if we need say less about it in the course of fully specifying its nature (106).[6] But here we also have a criterion of simplicity before us which is primarily nomological in character, since it must surely register the number of assumptions generally which a theory makes about the world, not merely the number of assumptions it makes concerning the number (or kinds) of entities which that theory posits. One theory of the world is not made simpler than another merely by virtue of positing fewer items, it is also made simpler by being regulated by fewer laws governing these items' interaction. Theoretical scope is of value largely for its capacity to simplify various domains of phenomena by unifying them under single, shared explanatory rubrics.

Now, it seems a safe bet that what applies to simplicity in this regard applies to candidate "selector mechanisms" generally. Simplicity stands in close relation to various other vir-

[6] It is worth noting that there may be problems with the very idea that the simplicity of possible worlds can be read off their alleged capturability by or concise descriptions. Van Inwagen's suggestion, once again, is that the simpler worlds are those about which we need say the least in the course of specifying their natures. But what must we presuppose for such a criterion to work? One thing we must presuppose is that we can objectively distinguish between featureless and cluttered regions of the universe. We must presuppose that in specifying the character of a worldly region, we can identify the default state which must obtain in order for that region to count as uncluttered. But why should we assume this? The intuition that motivates the assumption is certainly clear: worlds without concreta are simpler than worlds with concreta by virtue of being less cluttered. But, once again, suppose we approach things relativistically and view physical objects as ordered event sets or worldbraids, the relational configurations between which determine the properties of spacetime itself. On such an account, all of the concreta that van Inwagen posits, "objects, space, time and stuffs" are real, but dependent upon events. Such an account might even offer advantages to van Inwagen's account for its ability to clearly contrast abstracta with concreta, at least in the actual world: abstract objects are those items which exist outside of spacetime in the sense that they are existentially independent of the worldbraid material occupiers of spacetime.

If we construe our ontology in these terms, however, it becomes less clear that the worlds van Inwagen describes as ontologically Spartan admit of simpler description than the worlds he regards as (relatively) more cluttered. For a natural way to characterize fundamental reality on the model envisioned above is to attribute property exemplifications to either individual space/time points or to regions of space/time. To specify reality, on such a telling, is to describe for each space-time point or region the properties occurrent within it, so that concreta are construable as regions of space-time in which objects' essential and contingent properties are exemplified. But then the question arises as to whether or not negative properties exist. If they do, then our description of a space-time region in which a basketball exists (e.g., a region in which properties P, Q and R are exemplified) can be no less simple than our description of a region in which one does not exist (i.e., a region in which ~P, ~Q and ~R are exemplified). The issue of whether or not negative properties exist has much to do with the question of whether or not they have causal powers, which itself has much to do with the analysis of causation one decides to accept (on counterfactual theories, for instance, it is arguable that negative properties do have causal powers). But this is an issue beyond the scope of the present paper. Nick Zangwill offers a nice primer on these issues (Zangwill, 2011).

tues which we typically prize in theories. Comprehensiveness (construed as a measure of the number of phenomena explained and predicted) and scope (construed as a measure of the number of types of phenomena explained and predicted) both appeal, in part, to deeper simplicity considerations: the idea in the case of each is to unify diverse happenings under as few encompassing explanatory rubrics as possible. Thus, it is no accident that simplicity has generally been the historical poster boy for selector mechanisms, whether the question at issue has been that of determining which possible worlds are most likely, or that of determining which of competing empirically equivalent theories or models of nature are true or provide the "best" available overall explanation of things.[7]

To return to the issue of the tension between van Inwagen's premises (3) and (4), the significance of the above observations is that a world's simplicity (and thus any other candidate probabilifying feature) is most plausibly construed as a function of both its ontological and nomological attributes. Thus, to individuate worlds in terms of such probabilifying features, we must invoke criteria for distinguishing between worlds that mention, not merely the concreta constituting these worlds, but also the laws that govern these concreta. Moreover, premise (4) presupposes that the pertinent individuating laws in operation here include also those very general laws which constitute the pertinent selector mechanisms (so that the features individuating a world as the world it is must include, e.g., not only that world's simplicity, but also the selector mechanism which dictates that simpler worlds are more likely to exist than complicated ones). This is required by the maximality feature that van Inwagen finds it necessary to attribute to possible worlds in order to isolate them from the influence of any *external* selector mechanisms which might disturb their equiprobability with respect to each other.

We now have our essential tension between premise (4), which requires that we individuate worlds in terms of concreta and laws – including those laws which are the selector mechanisms, as required by the maximality condition – with premise (3), which requires that worlds are individuated only by their constitutive concreta, since empty worlds are identical only if this is the case. Without the former individuation criteria, we cannot guarantee the equiprobability of possible worlds. Without the latter individuation criteria, we cannot infer that there is but one empty possible world. We must instead conclude that there are infinitely many empty possible worlds, each distinguished from the others by its regulative laws.

Now, to say this we must make sense of the idea that worlds could remain distinct by virtue of being regulated by distinct natural laws rather than by being inhabited by distinct concrete denizens. To this end, we must presuppose certain features regarding natural laws. In particular, we must eschew Humean (or "systems" or "regularity") accounts on which laws of nature merely articulate uniformities in nature (e.g., actions are followed by equal and opposite reactions, the speed of light is *c*), and thus supervene on items in the world (Mill 1947; Lewis 1983, 1994). Instead, we must adopt some or other necessitarian account on which natural laws actually govern these constancies, effectively *making* them transpire.

[7] A notable example of this is to be found in Vogel (1990, 658-666). Here Vogel uses a criterion of explanatory comprehensiveness to argue that common skeptical hypotheses must always be worse explanations of experience than our everyday theory of prosaic physical reality.

Note, however, that having committed to a necessitarian view of this sort, we need not commit to any *particular* necessaritarian view. We may need to suppose that natural laws are describable as counterfactual relations between natural items, but we need not identify these relations as obtaining between objects and their properties, or between events, or between property exemplifications.[8]

Fortunately, I think that a necessitarian construal of natural laws is justified. John Carroll assembles a number of thought experiments which speak decisively in favor of such non-Humean accounts of law. (Carroll 2011) For instance, it seems perfectly reasonable to suppose that laws govern the ways in which particles would interact even in worlds in which they don't interact due to merely *contingent* circumstances which keep said interactions from ever occurring (Tooley 1977, 669). Similarly, consider a universe in which a single concrete object moves through otherwise empty space at a constant velocity of one meter/second. Carroll observes that there would seem to be a fact of the matter as to whether or not object velocity in this world is a constant or a non-constant feature of bodies. We can intelligibly pose the question of whether or not this velocity would change, for example, if the object were to collide with other inertial items (at least on an absolute conception of space)? The point is that no unique set of laws supervenes on the original imagined world, despite the fact that there still seems to be a fact of the matter as to what would happen if its circumstances were to vary. Humeanism looks to be false because the laws of a world are not uniquely determined by that world's total physical state. In particular, it is not uniquely determined by its number and kinds of concrete occupiers.

There may be other seeming complications for the account of possible world individuation described above, on which nomology floats free of ontology. But I suspect that in these cases also the problems at issue turn out to be either merely apparent or else easily addressed. For instance, we might worry that things get messy when we adopt a necessitarian view that calls for counterfactual analyses of laws, given that such counterfactual analysis invariably invokes modal alternatives which are likely to cross-entangle possible worlds from the outset in unwelcome and unforeseen ways. But this worry strikes to me illusory. For one thing, it gives us no reason to suspect that such entanglement, even if it did occur, would force us to describe possible worlds in terms of van Inwagen-style selector mechanisms. For another thing, it is at least arguable that any counterfactual assertions we might use to specify natural laws are reducible to claims referring only to the categorical bases which render these counterfactual assertions true, leading us to talk about causal propensities of substances and events. This point holds, I suspect, for most if not all modal claims. For instance, if it is an a posteriori necessary truth that water is H_20, then the modal fact that worlds containing H_20 must contain water is best fleshed out by reference to intrinsic and categorical features of water.

In short, I think it a safe assumption that the conflict between the individuation criteria for

[8] We might be obliged to eschew accounts, however, on which natural laws relate abstract universals like properties (Armstrong 1978, 1983; Dretske 1977, Tooley 1977, 1987). Once again, this is because both properties and relations, on van Ingwagen's account, exist (qua abstracta) across all possible worlds, including whatever relation of non-logical necessitation might serve to associate any two properties related in a law-like fashion.

possible worlds required by (3) and (4), respectively, keep van Inwagen's final conclusion from going through. He cannot depict a single empty possible world as being swamped by an infinite number of inhabited alternatives and claim in consequence that it is as unlikely as unlikely can be that there could have been nothing at all. This would be an arbitrary inference, no less than would be an inference to the conclusion that the probability that the world is empty is 1/2 and the probability that it is not empty is 1/2.

Conclusions and Morals

I have argued both that premises (3) and (4) are in conflict. I take this point to be important to the end of critiquing van Inwagen's argument. I have also argued against premise (3): on certain appealing accounts of natural law, van Inwagen can be seen as simply miscounting the number of empty worlds. Beyond these points, however, it is important to note that premise (4) is also questionable, and not merely because van Inwagen settles on the particular distribution of probabilities over possible worlds that he does. Rather, the problem is that invoking objective probability in our dealings with possible worlds is always a mistake. This is because such invocation can be nothing more than a veiled invocation of subjective (or epistemic or conditional) probability instead. The conclusion van Inwagen aims to derive concerns the way that the cosmos (as we may call the collection of all possible worlds) is. But this can hardly be done using the notion of conditional probability that he is working with. For irrespective of whether he construes possible worlds as collections of concreta (as suggested by premise (3)) or as collections of concreta and governing laws (as suggested by premise (4)), he can never be in a position to avow the equiprobability of possible worlds as anything more than a methodological principle of inference. What his reflections tell him is only that he has no a priori reason to privilege the probability of one possible world over that of another. Thus, the judgment that all worlds are equiprobable can reflect little more than a decision on his part. It can reflect nothing more than a procedurally dictated ascription of initial probabilities. This ascription may correctly reflect facts about human ignorance. But it does not reflect (for all we know) the way of the cosmos. This is seemingly the case for any interpretation of probability which is both plausible and potentially applicable to possible worlds.[9] [10] [11] [12]

---

[9] Classical interpretations are of no help here, as they would presumably presuppose that the probability of any given possible world's obtaining to be 1/infinity, precisely the kind of procedural assignment of values which, I have suggested, we have no reason to suppose gets at the metaphysical truth of the matter. Frequency interpretations look to be, at best, marginally applicable in this context, since the obtaining of a possible world isn't part of a sequence of events within which the relative frequency of that world's obtaining can be estimated (not that such a relative frequency could be determined in any case). Finally, Carnapian logical interpretations, on which syntactic features of premises generate degrees of confirmation for conclusions are, I take it, no longer regarded as plausible (Carnap 1950). Given the problem I have highlighted, i.e., the fact that the maximality of van Inwagen's possible worlds would seem to require that they include relevant selector mechanism, the interpretation of probability which would seem to be best fitted to entire possible worlds is some or other variant of Popper's, on which the probability of a world would be regarded as an objective feature of that world, i.e., intrinsic features of it which probabilify its obtaining (Popper 1959). But again, such accounts are not without their problems, and it is hard to see how we could know that such features obtain.

REFERENCES

Armstrong, D. (1978). *A Theory of Universals.* Cambridge: Cambridge University Press.

Armstrong, D. (1983). *What Is a Law of Nature?* Cambridge: Cambridge University Press.

Carnap, R. 1950. *Logical Foundations of Probability.* Chicago: University of Chicago Press.

Carroll, John W . Laws of Nature. *The Stanford Encyclopedia of Philosophy* (spring 2011 Edition), Edited by Edward N. Zalta. (ed.), URL =<http://plato.stanford. edu/archives/spr2011/entries/laws-of-nature/>. First published Tue Apr 29, 2003; substantive revision Sun Dec 26, 2010.

Dretske, F., 1977., Laws of Nature. *Philosophy of Science*, 44: 248-268.

Hawking, S., & Mlodinow, L. (2010). *The Grand Design*. New York: Bantam Press.

Horgan, J. (2010). Cosmic Clowning: Stephen Hawking's "new" theory of everything is the same old CRAP. *Scientific American*.

Leibniz, G. W. (1992). *Discourse on Metaphysics and the Monadology* (trans. George R. Montgomery). Amherst, New York: Prometheus., sec. VI.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61: 343-377.

Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, 103: 473-390.

Manson, N. A. (2011). No Chance for Nothing. *American Philosophical Association Pacific Division Conference*. San Diego, CA.

Mill, J. (1947). *A System of Logic*. London: Longmans, Green and Co.

---

[10] Neil Manson has argued that assumptions like van Inwagen's equiprobability premise (4) are philosophically tempting because the profession is presently so enamoured of Bayesian reasoning, perhaps because such reasoning seems to promise to make epistemology so very easy. (Manson 2011) I concur with this diagnosis.This beguilement has led to an extension of Bayesian aspiration into realms where it doesn't belong. Bayesianism promises us a method through which we can arrive at correct results even when our initial a priori probability assignments are relatively uninformed by evidence because pertinent relevant frequencies or propensities are not known (thus van Inwagen begins with the assumption that all possible worlds are equiprobable).

However, Bayesianism only offers this promise on the condition that we are able to update our initial probability assessments on the basis of incoming data. But when we are endeavouring to assess the probability of entire worlds, no such incoming data is forthcoming. Because no incoming data speaks in favour of the probability of the actual world vis-a-vis its alternatives, nothing we observe in the actual world could ever tell us anything of relevance more than what we originally knew (or didn't know) concerning the probability that this or any other particular world should have obtained. For this reason, the original assumption of possible world equiprobability can never get revised. Bayesianism interprets probability as a measure of perceived epistemic position, rather than as the objective frequency or a propensity that a given phenomenon might have to obtain.

Thus, for all we can ever know (intrinsic or extrinsic) selector mechanisms might very well operate to privilege some possible worlds over others; there might indeed be core elements of a single, or else common to a class of, metaphysically possible worlds that distinguish them from merely logically possible ones. Because we never have access to a broad enough background of statistical law to which we might appeal, it is impossible to make judgments concerning the objective or metaphysical probabilities of entire worlds in order to revise the equiprobability assumption we may have originally made in order to get a leg up on Bayesian reasoning. Objective judgments of probability must always be, to quote Derik Parfit, "grounded on facts about the world, so that such judgments cannot be applied either to how the whole of reality might be, or to how reality might be explained."

Parfit, Derek. (1998.) Why Anything? Why This? *London review of books*, vol. 20, no. 2 (January 22, 1998), 24-27; no. 3 (February 5, 1998), 22-25. Accessed at http://spot.colorado.edu/~heathwoo/phil3600/parfit.pdf.

Popper, K. R. 1959. The propensity interpretation of probability. *British Journal for the Philosophy of Science* 10: 25-42

Tooley, M. 1977. The Nature of Laws. *Canadian Journal of Philosophy* 7: 667-698.

Tooley, M. 1987. *Causation*. Oxford: Clarendon Press.

Van Inwagen, P. 1996. Why Is There Anything at All? *Proceedings of the Aristotelian Society, Supplementary Volumes*, 95-110. Oxford: Blackwell.

Vogel, J. 1990. Cartesian Skepticism and Inference to the Best Explanation. *The Journal of Philosophy* 87 (11): 658-666

Zangwill, N. (2011). Negative Properties. *Nous*, 1-29.

Received: January 31, 2013
Accepted: June 17, 2014

---

[11] The sort of story I've told above (in footnote 10) is also illuminating in certain epistemological contexts. In particular, it helps to explain the Cartesian skeptic's conception of knowledge as epistemic certainty. This conception should not be viewed as the result of some unmotivated and irrational decision to impose arbitrarily high standards upon our ordinary epistemic practice. Rather, it is better viewed as a result of the skeptic's attempts to talk about justification even when the generality of his inquiry robs the notion of degrees of justification of any possible purchase. Comprehensive skeptical scenarios are like maximal possible worlds. They jointly exhaust the whole of logical space, leaving no presuppositional material through the use of which one might judge common sense realism to be more likely than its various skeptical alternative scenarios. Thus, for the skeptic knowledge would have to be certainty to be anything at all. The range of epistemic states intermediate between absolute certainty and abject ignorance collapses like a broken accordion.

[12] Baysean issues aside, the problems we encounter whenever we endeavour to determine the relative probabilities of alternatives ways the world might have been are nicely illustrated by a recent and celebrated examples. Without pretending to understand the cutting-edge physics (Thank God), and looking instead merely at certain formal features of the proposal, we can see in Stephen Hawking's recent popular packaging of his "Grand Design" argument an attempt to identify intrinsic self-probabilifying characteristics of the actual world (Hawking and Mlodinow 2010). For his is an account on which the internal features of M-theory (which he alternately treats as a truly unified field theory and as a mere collocation of disparate accounts which collectively explain the whole of nature) breathe life into its own equations in a way that brings our universe into existence. Hawking's fundamental assumption is that the total (positive and negative) energy of the universe must remain constant. Thus, "on the scale of the entire universe the positive energy of matter can be balanced by negative gravitational energy," so that transformations from the latter to the former can create mass from the background energy vacuum alone. The fact of gravity is the thing that brings about such transformation, on Hawking's account. "Because there is a law such as gravity, the Universe can and will create itself from nothing." (Kindel edition, no pagination) Moreover, the universe so brought about must be replete with the fundamental laws and physical constants that characterize our own world. This allegedly follows from the requirement of supersymmetry, the counteractive balance

Dept. of Philosophy
University of Nevada
4505 Maryland Pkwy.
Box 455028
Las Vegas, NV 89154-5028
wilburn.ron@gmail.com

---

between force-making and matter-making particles, and the accompanying fact that M-theory is "the most general supersymmetric theory of gravity" and thus "the only candidate for a complete theory of the universe."

For our purposes, what is intriguing about Hawking's description is, again, not the details, but his summary claim that the genuine miracle in all of this "is that abstract considerations of logic lead to a unique theory that predicts and describes a vast universe full of the amazing variety we see" (Kindel edition, no pagination). For surely the assumption of supersymmetry is not a matter of abstract logic. Nor can M-theory be unproblematically identified as uniquely consistent in any interesting way. John Horgan has noted that M-theory is but one iteration of string theory, which has enjoyed two decades of popularity less for the its actual merits than for lack of decent alternatives (and possibly because of the near-religious enthusiasm of its adherents). Whatever the merits of the approach, no one is in a position to proclaim its logically inevitability. More interestingly, it has also been noted that M-theory comes in an almost infinite number of versions, each of which predicts a different universe. (Horgan 2010) This result is embraced by Hawking, who proclaims that all of these universes exist. But to say this is to seriously muddy the original proposition we set out to defend, i.e., that an examination of the intrinsic features of some particular world (say, the actual one) might provide us grounds for proclaiming its probabilistic privilege relative to its alternatives. What force can there be to the claim that intrinsic features of the actual world privilege its existence if a nearly infinite number of alternative possible worlds also "exist"?

Unrelatedly, the significance of Hawking's assertions is additionally muddled by his avowal of "model-dependent realism." When he asserts, for instance, "there is no model-independent test of reality [and] it follows that a well-constructed model creates a reality of its own," it becomes less than clear even what makes a "final theory of nature" worthy of the designation. It also reminds us that a role yet remains for philosophy to urge clarity on the part of scientists. I would be so bold as to suggest that Hawking's own lack of precision in such matters belies his assertion that "philosophy (i.e., metaphysics, philosophy of science) is dead."

EuJAP | Vol. 10 | No. 1 | 2014
UDK 165.21
1:81

WHY SEMANTIC UNSPECIFICITY IS NOT INDEXICALITY

DELIA BELLERI
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México

ABSTRACT

In this paper, I address the idea that certain sentences ("It's raining", "The leaves are green", "Sally reads an Italian book") suffer from what is generally called semantic unspecificity: their meaning is determinate, but their truth conditions are not. While there tends to be agreement on the idea that semantic unspecificity differs from phenomena such as ambiguity and vagueness, some theorists have defended an account which traces it to indexicality, broadly construed. Some authors have tried to vindicate the distinction between unspecificity and indexicality and, in this paper, I pursue the same cause, but with a critical stance towards previously employed strategies. I urge that the central argument employed by Sainsbury to trace this difference fails suitably to set unspecificity apart from indexicality and I propose a new argument, which helps to trace this distinction more perspicuously. The argument is based on embeddings of indexical and unspecific expressions within modal operators and on the ways in which the truth conditions of utterances of the resulting, complex expressions are thereby affected.

Keywords: semantic unspecificity, semantic underdeterminacy, indexicality, rigidity, intensional contexts

1. Some background on semantic unspecificity

The focus of this paper is the specific kind of semantic indeterminacy that afflicts sentences like the following:

(1) It's raining.
(2) The leaves are green.
(3) Sally reads an Italian book.

The problem with these sentences is that their linguistic meaning fails to determine a truth-condition for their utterances. For instance, the linguistic meaning of (1) is just that it is raining at the time of utterance (where the temporal information is supposed to be encoded in the present tense), but there is no word which stands for a locational aspect. The semantics of the sentence thus seems to fall short of specifying in which conditions an utterance of (1) is true. As to (2), it just says that the leaves are green, where this is unspecific between the leaves being green, for instance, on the outer surface or as their natural shade (perhaps covered by a red dye). Finally, (3) simply states that Sally reads an Italian book, but in what sense the book is Italian remains unspecified. Is it written in Italian? Has it been printed in Italy? And so on. Theorists regard sentences (1)-(3) as having a meaning which is "too poor" to specify a content for their utterances. Following Sainsbury (2002), I use the term *semantic unspecificity* to refer to this kind of phenomenon, which falls under the head-

ing of the broader phenomenon of semantic under-determinacy (cf. Carston 2002).

Unspecificity seems to be a peculiar phenomenon, quite different from other kinds of semantic indeterminacy. Firstly, there seems to be a difference between unspecific and *ambiguous* terms. Let us focus on the ambiguity that arises from polysemy: polysemous words such as "newspaper" consist in single lexical entries to which the dictionary assigns more than one meaning. Moreover, although the number of meanings can increase or decrease in time, it is generally definite and the meanings themselves are clearly identifiable (for instance, "newspaper publishing company", "newspaper editorial committee", "newspaper issue (type)", "newspaper issue (token)"). The latter conditions may be connected with facilitating language acquisition and memory, meaning that a single polysemous word may be more difficult to learn and competently apply if its meanings are indefinitely many and not clearly identifiable. Now, one could suppose that a word like "green" is polysemous too and therefore associated with many meanings: $green_1$, $green_2$, $green_3$,... $green_n$, according to what the speaker means with her use of "green" in different situations. However, since there are indefinitely many ways for something to be green and for it to be competently so called, this would lead to indefinitely many meanings for the word-*type* "green" which are not clearly identifiable (since we cannot clearly identify all the situations in which something can competently be called "green"). Thus, the constraint of meanings being definitely many and clearly identifiable would be violated. So, *prima facie* at least, we seem to have good reasons for resisting the claim that unspecificity is just polysemy (for further arguments, see Travis 1997 and Bezuidenhout 2002).

Unspecific expressions also seem to depart from vague expressions, although I believe one should concede that "to rain", "being green", "being Italian" etc. are *also* vague. If we assume that a necessary feature of vagueness is the presence of borderline cases, then there are certainly borderline cases of application for these expressions, for there are borderline cases of raining, of being green and even—arguably—of being Italian. However, the particular defectiveness that makes these expressions semantically unspecific is not, I argue, connected with their being vague. To illustrate, consider an utterance of "It's raining", which one assesses as not determinately true or false (maybe because drops of water are falling from the sky at too large a distance from each other). Now it seems that, in order for one to be able to say that the utterance doesn't have a definite evaluation in the first place, one needs to specify the location where the phenomenon is taking place—Paris, Venice, London etc. That is, in order to judge the vagueness of the statement "It's raining" one must previously resolve its semantic unspecificity as to the location of the rain. These considerations seem to make for the idea that semantic unspecificity doesn't coincide with vagueness, for the former needs to be dealt with "before" the latter is ascribed (for further arguments, cf. again Travis 1997, Bezuidenhout 2002). Moreover (and in connection with what has just been remarked), theorists often talk about vagueness as a matter of *degree*: for instance, they focus on what degree of hairlessness one must possess in order to count as bald; whereas often when theorists talk of unspecificity, they address it as a matter of multiple *dimensions*, each of which allows for degrees: for instance, where does an object have to be green in order for it to count as such? Or how does it have to be green? And so on.

Some authors trace the semantic unspecificity of (1)–(3) to the already familiar phenomenon of *indexicality* broadly construed, i.e. sensitivity of meaning to certain features, or parameters, of the context of utterance. Thus, for instance, Stanley (2000) posits a free variable in the logical form of (1), which occupies a hidden argument-place for locations in the predicate "to rain". "To rain" thus becomes a dyadic predicate *rain(t, l)*, expressing a relation between a time *t* and a location *l*. Resolving the semantic unspecificity of "It's raining" means simply filling in a covert, free slot in logical form, according to the location which is salient in context. Similarly, Szabò (2001) argues that the word "green" doesn't correspond to the monadic predicate *green(x)*, but to a two-place relation *green(c, p)*, in which *c* stands for a comparison class and *p* stands for a part in which the object is green. This way the predicate, as it occurs in "The leaves are green", is not unspecific, rather it is incomplete due to a failure to fill in the slot dedicated to the part in which the leaves in question are green. Once again, resolving semantic unspecificity becomes equivalent to resolving an instance of indexicality construed as sensitivity of content to certain features or parameters of the context of utterance.[1]

Accounts of unspecificity in terms of hidden indexicality encounter more than one difficulty. The most straightforward is the challenge arising from *overgeneration*. If the arguments of the proponent of hidden indexicality are correct, then they can plausibly be applied to more and more expressions which share the same characteristic features of "to rain", "being ready", "being Italian". This means that we may find ourselves with an unexpected amount of hidden indexicality in language. Most importantly though, the problem is that the hidden indexicality arguments seem to "work too well", to such an extent that we seem to lack a principled way of *ruling out* that further hidden argument-places may be "discovered" (see Cappelen and Lepore 2002, 2005, Hall 2008). There are also some methodological worries, pertaining to the data employed by the proponents of hidden indexicality in order to establish their theory. Since these authors aim to make an *empirical* point concerning the syntax of these expressions and yet the data they start from involve purely semantic intuitions, this way of proceeding has been indicated as methodologically flawed (see Neale 2007, Collins 2007, Pupa and Troseth 2011).

The objections just summarized are good insofar as they point to some undesirable implications for the hidden indexicality theory. Yet, they do not directly question the idea that unspecificity is indexicality. In this paper, I aim to provide a positive argument to counter the contention that unspecificity is really indexicality broadly construed. The most respectable way to achieve this result is by pointing at some properties that set unspecific and indexical expressions apart by virtue of how these expressions work within language and language use. An attempt in this direction is made by Sainsbury (2002), who explicitly challenges the hidden indexicality approach to the semantics of (1)–(3), with an argument aimed at distinguishing between unspecificity and indexicality. I find Sainsbury's argument flawed, though I sympathize with the point he wishes to make and the position he wishes

[1] Rothschild and Segal (2009) claim to offer an indexical account alternative to that advanced by Szabò (though this is dubious: see Clapp (2012)), while Stanley and Szabò (2000) work out an account in terms of hidden indexicality of quantifier phrases like "every bottle". Finally, King and Stanley (2005) argue for a methodology that favours hidden indexicality treatments over pragmatic intrusion strategies, using comparatives ("better than") and conditionals as case-studies.

to defend—namely, that there is such a thing as semantic unspecificity, which differs from indexicality. What I set out to do in this paper is to counter Sainsbury's original argument and eventually propose a new one, aimed at supporting the same conclusion—namely that semantic unspecificity differs from indexicality. The paper is organised as follows: In section 2, I will briefly reconstruct Sainsbury's central argument that semantic unspecificity doesn't coincide with indexicality. In section 3, I will highlight some problematic aspects of the argument, which stem from an unclear use of the notion of "comprehension", which ultimately prevents the argument from fully succeeding. In section 4, I will propose a different argument which will help to trace this difference more clearly.

2. The Comprehension Argument for Unspecificity

Sainsbury's central argument supporting the claim that sentences like (1)-(3) are unspecific rather than covertly indexical is essentially based on the idea that we "understand" these sentences even if no contextual information is available to us, contrary to the prediction of the indexicalist. Consider a sentence like (3) ("Sally reads an Italian book"). Let's suppose that the linguistic expression "Italian book" is covertly indexical. In Sainsbury's words:

> On the covert indexical theory, the logical form of "Italian book" could be represented by something like "Italian R book", where the interpretation of the relation variable R is to be supplied by the context. The form "xRy" could be interpreted so as to be true of the satisfiers of y which are written in a language which satisfied x, or so as to be true of the satisfiers of y manufactured in a place which satisfied the noun from which x is formed, and so on. [...C]ontext can point to sensible interpretations of R, and can place obstacles in the way of accessing interpretations which, in other contexts, would be natural. However, unless nothing better can be found, it seems an extraordinary account as applied to this kind of case. *It implies that you would not have understood an utterance like "Let's read an Italian book together" unless you had identified such an R; whereas in fact it seems you do understand even when you are in doubt about R.* You may go on to ask "Do you mean a book in Italian or a book about Italy?", but this no more shows that you did not understand the first remark than if, in response to "Let's go to the movies" you say "Do you mean let's go tonight or later?". In both cases, the proposal was fully intelligible but not fully specific. (Sainsbury 2002, 197-198, my emphasis)

Sainsbury's argument could be thus summarized: if "Italian book" were covertly indexical, thus having a logical form like "Italian R book", then the hearer who listened to an utterance of (3) would not understand it unless she had identified the right R. But, Sainsbury notes, hearers usually do understand uses of "Italian book" even if they are uncertain of what R is. So, expressions like "Italian book" are not indexical: they are just unspecific.

The comprehension argument introduces a distinction on which Sainsbury puts much stress in the subsequent lines. The distinction is between what the sentence may be taken to mean (its "reading"), as opposed to "what makes it true". As he states:

> It seems to me hard to deny (a) that we should distinguish different "readings" of a sentence from different ways in which it could be made true and (b) that this distinction has not always been scrupulously adhered to. (199)

This separation between "readings" and—let's say—"truth-conferring aspects" is essential to his notion of unspecificity. Unspecific expressions all have determinate, fully intelligible readings, i.e. they can be the object of comprehension. At the same time, they suffer from a lack of determinacy as to what in the world "makes them true or false". They are, so to speak, determinate in their "internalistic", psychologically relevant semantic aspects, while they are indeterminate as to their "externalistic" world-related semantic properties.

The comprehension argument plays a key role in Sainsbury's whole strategy. My aim in the following section will be to assess the claim that we have "comprehension" of uses of linguistic expressions like "Italian book" (but also "The leaves are green", "Sam cut the grass", "It's raining", and so on).

3. Assessing the Comprehension Argument

There are at least two ways of interpreting the word "comprehension" as Sainsbury uses it in the previously quoted passage. Firstly, comprehension may be equated with one's understanding of the "meaning-in-context" of a sentence, where the meaning-in-context of a sentence *s* results from the linguistic meaning of its non-context-sensitive components *plus* the content that context-sensitive expressions—typically, indexicals and demonstratives—acquire in the context of utterance. For instance, the meaning-in-context of an utterance of "I wear glasses" as uttered by Sally is that Sally wears glasses; the meaning-in-context of an utterance of "She is a photographer" as uttered while referring to Amy is that Amy is a photographer. The idea of meaning-in-context may be expanded once one accepts that sometimes speakers utter a sentence *s* which semantically expresses a proposition *p*, while meaning a richer (but still relevantly similar) proposition *p**. A case in point is, for instance, "I've had breakfast" (see Recanati 2004): while the sentence expresses the proposition that *x* has had breakfast at least once in the past, one may want to use this sentence to say something more specific (but still related to the sentence's literal meaning), perhaps that one has had breakfast *on the morning of the day of utterance*. If one accepts that the more specific proposition is still an instance of meaning-in-context (and not, to mention an alternative, an implicature), then comprehension in the first sense may include understanding of this kind of content as well.

If comprehension is understanding of (broadly construed) meaning-in-context, then it seems clear to me that there can be no comprehension in the sort of cases Sainsbury asks the reader to consider. Imagine we heard an utterance of "Let's read an Italian book" and were not sure of what relation counts for the book to be Italian. To the extent that the hearer has no access to this background information, she is in no position to understand the (broadly construed) meaning-in-context of the sentence—perhaps that the speaker and her audience should read a book *written in* Italian. For the comprehension argument to work, the notion of comprehension should not be understood as comprehension of meaning-in-context, whether in a restricted or broad sense of the term.

The second way in which "comprehension" could be understood is as *linguistic* comprehension, which may be characterised as knowledge of what a sentence means only in virtue of one's linguistic competence. This kind of comprehension is acquired by consulting one's own semantic competence or knowledge, in a context-independent fashion.

If comprehension is equated with linguistic comprehension, it seems clear to me that the proponents of hidden indexicals as well can claim that we have linguistic comprehension of expressions like "Italian book" in the same way that we have linguistic comprehension of words like "I", "today", "now". These expressions do after all have a linguistic meaning, which Kaplan (1989) called "character": the character of "I" may be captured as "the speaker in context", the character of "today" may be captured as "the day of utterance", and so on. The character of "Italian book" may well be "book which is Italian in a contextually relevant way". If the indexicalist is willing to accept this as the character of "Italian book", then there's no reason why she should predict, as Sainsbury maintains, that the hearer will not understand those words unless she identifies the contextually relevant relation that counts for the book to be Italian.

To sum up: either we interpret comprehension as an understanding of meaning-in-context broadly construed, in which case there is no comprehension in the cases relevant to Sainsbury's argument (since, by assumption, hearers have no access to relevant background information); or we interpret comprehension as linguistic comprehension, in which case indexicalists have no reason to predict that there will be no comprehension in the cases cited, contrary to what Sainsbury maintains. Either way, the comprehension argument seems to have problems. In the first case, the argument fails because the contention that language users understand the expressions at issue is false. In the second case, the argument may not fail if considered on its own: after all, hearers *do* seem to have linguistic comprehension of expressions like "Italian book". Yet they *do* have linguistic comprehension of indexicals as well, if we accept that indexicals have a linguistic meaning, or character. So, there is no difference between unspecificity and indexicality when it comes to this kind of comprehension.

4. Indexicality and Unspecificity: an Argument from (Modal) Embeddings

In this section, I would like to propose a way to clearly discriminate between unspecificity and indexicality.[2] The distinction I am about to make presupposes a *two-dimensional* semantics (see Kaplan 1989), in which sentences express propositions that are true at a context and a circumstance of evaluation. A context is a set of parameters which fix the

[2] The present proposal is in the same spirit as a number of remarks already set forth in the literature, mainly purported to show that the hidden indexicality strategy is misguided because alleged hidden indexicals fail to bear enough similarities to overt indexicals: for instance, Neale (2007) argues that hidden indexicals, unlike overt ones, express no perspective (for example, as to spatial and temporal distance, subjective or objective stance etc.); Cappelen and Lepore (2002) contend that hidden indexicals fail to enter in anaphoric relations and fail to give rise to a priori truths; Recanati (2010) notes that purported hidden indexicals have indefinite readings (e.g. the supposed hidden indexical in "It's raining" can receive an indefinite reading as in "It's raining *somewhere*") while overt indexicals trigger no such readings.

semantic value of context-sensitive expressions like "I", "here", "now", "she", "this". A circumstance of evaluation is a set of coordinates which serve for evaluation: typically, the index of the circumstances of evaluation contains possible worlds, but it may also include time and location coordinates, as well as coordinates on standards of precision, depending on the operators one is willing to admit into the language. In such a framework, a sentence *s* expresses a proposition *p* which is true at a context *c* and at a circumstance of evaluation *i* if *p* is true at i_c, i.e. at the circumstance of evaluation of the context (for instance, the possible world in which the utterance is performed).

The difference I will outline has to do with how the content of indexicals, as opposed to that of semantically unspecific expressions, is fixed when these expressions are embedded in intensional contexts, and in particular in modal contexts (operators like "It is possible that", "It could be the case that" and so on).

Let us start with indexicals. A characteristic feature of indexical expressions—especially of those described as "pure"—is that once their referent is fixed in a context, it is fixed for all circumstances of evaluation. As Kaplan states:

> When what was said in using a pure indexical in a context *c* is to be evaluated with respect to an arbitrary circumstance, the relevant object is always the referent of the indexical with respect to the context *c*. (Kaplan 1989, 500)[3]

It follows that, when an occurrence of "I" is embedded into a modal operator, whose role is characteristically that of "shifting" the world of evaluation for the embedded sentence, the referent of "I" doesn't switch as the circumstance of evaluation switches, since it is "rigidly" fixed at the context of utterance, as in:

(4) It is possible that I don't wear glasses.

The sentence expresses a proposition which is true iff there is at least one world w_i accessible from the actual world $w_@$ in which the *actual* speaker in context, *a*, doesn't wear glasses. Note that the referent of "I" is fixed, at the context of utterance, in such a way that that particular occurrence of "I" refers to *a* and such referent is fixed for all circumstances of evaluation.[4] This implies that, once the referent of that occurrence of "I" has been fixed in context, for every circumstance at which the embedded sentence may happen to be evaluated, one cannot reasonably find it indeterminate whether the expression has a reference or not—for, one would expect, either *a* exists in that circumstance (conceived as a possible world *w*), or she doesn't.[5]

---

3 See also Maier 2009, who defends the same Kaplanian claim.

4 With appropriate changes, the point seems to hold even for those theories that oppose the Kaplanian account in regarding indexicals as behaving like "monsters", that is as having their reference fixed also in contexts different from the context of utterance (see among others Nunberg (1993), Predelli (1998a, b), Schlenker (2003), Santorio (2010), Mount (2008), Parsons (2011), Rabern (2013)). These theorists could say that once the reference of "I" is fixed, *in the context of utterance or in any other relevant context*, its reference is fixed for all circumstances of evaluation.

5 With suitable modifications, I take it that this claim could be endorsed even by somebody who believes in counterparts. One could substitute "either *a* exists in that circumstance (conceived

To be completely fair, there could be circumstances in which it is not clear whether the actual speaker exists or not—maybe her body exists, but her brain has been transplanted into somebody else's skull; or maybe she exists, but she has a double personality and it is not clear which personality "I" refers to, and so on. In all these scenarios it would be unclear whether the counterfactual individual is identical to the actual speaker, however this could not be blamed on an indeterminacy of the referent of that particular occurrence of "I" in that particular context, since who the referent is in the context of utterance could be taken as clear (at least in ordinary cases); the indeterminacy is due, rather, to the obtaining of conditions that render it difficult to tell whether the individual in the imagined counterfactual scenario is the same as the individual who is actually doing the speaking. So I suggest that, even though indeterminacy could arise concerning the reference of "I", this would *not* be due to any indeterminacy at the semantic level, but to an indeterminacy that concerns a *metaphysical* question, namely whether a certain individual is identical to another individual.

This being said, the feature of "I" illustrated a few paragraphs above suggests the following thought: when the content of an indexical is assigned in a certain context *c*, e.g., when the referent of "I" in *c* is fixed as being identical to *a*, the content of the expression is fixed with respect to *all* circumstances of evaluation. This means that, once the required contextual information has been provided, no residual question need be asked as to what the referent of that occurrence of "I" is, for the referent will be identical with *a* in all worlds in which *a* exists (metaphysical issues aside).

Unspecific expressions, like "Italian book", seem to behave differently. Imagine Sally is reading a book which is written in English, but since she is bilingual, she could just as easily read a book which is written in Italian. She therefore utters the sentence:

(5) It is possible that I read an Italian book.

The sentence contains the unspecific expression "Italian book". With the help of background information as to what Sally means by "Italian book"—namely, a book written in Italian in a normal way—the unspecificity of the expression may be taken to have been resolved in the context of utterance: "Italian book" in Sally's mouth means "book written in Italian in a normal way". After the expression's unspecificity has been dealt with, we may be tempted to say that the utterance expresses the content that there is at least one world w_i accessible from the actual world $w_@$ in which Sally reads a book which is written in Italian in a normal way (and is true iff there is at last one world w_i accessible from the actual world $w_@$ in which Sally reads a book which is written in Italian in a normal way).

Yet, despite this completion, it is perfectly possible for (5) to be true even if circumstances different from the circumstances of the context of utterance obtain. The point deserves to be developed in more detail. To see it more clearly, first consider the simple sentence

as a possible world *w*), or she doesn't" with "either *a's counterpart* exists in that circumstance (conceived as a possible world *w*), or she doesn't". In this paper, though, I will assume a Kripkean take on modality.

(6) I read an Italian book.

Suppose that, in the context of utterance *c*, "Italian book" is to be understood as "book written in Italian in a normal way", in such a way that the expression's unspecificity is (presumably sufficiently) resolved in *c*. An utterance of this sentence may be true in a circumstance of evaluation *k1*, where the speaker reads a book which is written in Italian in the usual way—with words and sentences all in their place. However, it seems clear that the same expression "Italian book" could have been used in a different (counterfactual) context, to mean that the book in question is, for instance, written in Italian, but with a Cyrillic script. And in yet another (counterfactual) context, the expression "Italian book" could have been used to mean that the book in question has been bought in Italy.

The next step in the argument builds on the latter result and brings intentional operators into play, such as "It is possible that", "It could be the case that". These operators, as Kaplan (1989) and Lewis (1980) conceive them, combine with sentences in order to form more complex sentences; their role is to shift the circumstances of evaluation at which the sentence is to be assessed for truth or falsity.

Once we introduce intensional operators, we see that the way the expression "Italian book" is completed in the context of utterance need not affect the truth-conditions of the modal sentence. Consider the sentence:

(5) It is possible that I read an Italian book.

All that is required for this sentence to be true is that there is a world w_i accessible from the actual world $w_@$, where the speaker reads an Italian book, but note that the content of the words "Italian book" could be satisfied by different things from one circumstance of evaluation to another. The fact that, in the context of utterance, "Italian book" is used to mean "book written in Italian in the normal way" seems irrelevant for the purposes of evaluating an utterance of (5). So, for instance, (5) would be true even if there were a circumstance w_2 where "Italian book" is used to mean "written in Italian with a Cyrillic alphabet"; or if there were a circumstance w_3 where "Italian book" is used to mean "bought in Italy".

In order to see this, we may imagine a scenario in which the relevant interpretation of "Italian book" in the context of utterance is "book written in Italian in the normal way". Suppose Sally only reads the Cyrillic alphabet and, as a consequence, she cannot read Italian, because it is written in Latin alphabet. She utters: "I have never read an Italian book". Here we may suppose that the relevant interpretation of "Italian book" is such that the expression indicates books written in Italian in the normal way – obviously, with Latin alphabet. Even though this is the salient interpretation, Sally could go on and utter: "Though I *could* read an Italian book, if the words in it were Italian, but they were written in Cyrillic alphabet". As we can see, once the sentence "I read an Italian book" is embedded into a modal "could" (which I will consider as equivalent to operators like "It is possible that", "Possibly"), the contextually salient interpretation of "Italian book" has no role in fixing the truth-conditions in the modal sentence, which is true as long as there is a possible world w_i accessible from the

world of the utterance $w_@$ where Sally reads a book which is Italian *in some way*. Indeed, the way in which the book is Italian in the possible world w_i accessible from $w_@$ could even be different from "being written in Italian"; it could be "bought in Italy". To see this, imagine that Sally opens a book written in Italian and utters "I cannot read this Italian book". Here the salient interpretation is once again "written in Italian in the normal way". This interpretation need not affect "Italian book" when embedded in a modal context, since Sally could go on and say: "But I *could* read an Italian book, were it a book in Cyrillic bought in Italy". Here we see that the contextually salient interpretation of "Italian book" has definitely no bearing on the occurrence of the expression in the modal context, and on the truth-conditions of the sentence.

A contrast then emerges: on the one hand, when an indexical occurs in a modal embedding, once the content of that indexical is fixed in a context of utterance, it is fixed for all circumstances of evaluation. On the other hand, when an unspecific expression occurs in a modal embedding, even though the content of that semantically unspecific expression is fixed in the context of utterance, this content is not fixed for all circumstances of evaluation. To see this, it is enough to look at the truth-conditions of modal sentences that contain an occurrence of semantically unspecific expressions like "Italian book". These sentences could be true even with respect to circumstances of evaluation where the words "Italian book" describe something which doesn't satisfy the contextually salient interpretation.

A referee suggests that this contrast is ultimately due to the fact that indexicals are directly referential, while semantically unspecific expressions are not; this would imply that reference-fixing for a term like "I", which—at least according to the standard Kaplanian account—is tantamount to providing the value for a contextual parameter in a non-mediated way,[6] guarantees that the content that the indexical acquires in that context is the same for all circumstances of evaluation, while the way the content of an unspecific expression is fixed in context doesn't seem to give rise to the same guarantee. I take this contrast to be helpful in achieving a neat distinction between indexicality and semantic unspecificity.

A second example is worth analysing, which highlights further the contrast between indexicality and unspecificity. Consider a sentence like "It's raining" as opposed to "It's raining here". If one embeds the latter sentence within a modal operator like "possibly", thus obtaining

(7) Possibly, it's raining here

it's clear that the reference of "here" remains fixed at the location of utterance despite the

[6] One could disagree with the idea that indexicals are directly referential: For example, one may maintain that the reference of "I" is fixed through a definite description (Frege 1918/56, Searle 1983, Castañeda 1974, Loar 1976, Brinck 1997). This would of course drastically change the modal profile of sentences containing "I". I won't discuss descriptivism on the semantics of indexicals here and I will confine myself to assuming referentialism. Suffice it to say that I find it extremely difficult, at the level of linguistic intuition, to force a reading in which the reference of "I" is fixed by a definite description. In a sentence like the following, it's *very hard* to interpret the "I" as referring to Ralph: "Since we could be in a situation in which the person who is talking right now is Ralph, in that situation I could be talking right now".

fact that the circumstance of evaluation is shifted by the operator. If one is in Paris and utters (7), clearly the utterance is true iff there is a world w_i accessible from $w_@$ where it's raining *in Paris*. By contrast, if one embeds "It's raining" in a modal context, thus obtaining

(8) Possibly, it's raining

whether the location of utterance is Paris does not matter much. If Paris were salient in context, the uttered sentence could be interpreted as true if there were at least one world w_i accessible from $w_@$ where it rains in Paris. However, note that even if this were the salient interpretation, the utterance could still be true even in a world where it rains in London. This is testified by the following example. Suppose one utters "It's not raining here" in Paris, where in fact it's not raining; as a consequence of this, Paris is supposed to be particularly salient in the conversation. Still, it's perfectly appropriate to utter:

(9) It's not raining here [*Paris*]. But it could be raining, were we in London.

In the first sentence of (9), Paris is made highly salient by virtue of its being the location referred to by "here". Yet it is perfectly possible to interpret the second sentence in (9) as if the salient location were London. The fact that Paris is the place of utterance and is as a result particularly salient doesn't constrain the implicit reference to a location as being identical to the location of utterance. This phenomenon contrasts with how "here" would behave in a similar modal embedding. It would be infelicitous to say: "But it could be raining here, were we in London". In this case, the reference of "here" *has* to be the place of utterance (Paris), and it's very hard to extract a reading in which "here" refers to London. The fact that indexicals like "here" are directly referential fixes their reference in a much more "rigid" way than the way in which implicit reference to a place in an unspecific sentence like "It's raining" could be fixed.

To sum up: drawing from the argument just developed, indexicals appear as expressions such that, in modal embeddings, whenever their content is fixed in a context of utterance, it is fixed for all circumstances of evaluation. By contrast, unspecific expressions in modal embeddings are such that, even if their unspecificity is dealt with in a context of utterance, this contextual completion doesn't determine a content which remains the same in all circumstances of evaluation, as one can see by looking at the truth-conditions of modal sentences like (5), (8) or (9).

The role of *modal embeddings* is key in the argument, even though some might feel that the argument could be run without involving intensional contexts. It is only in such embeddings that we can fully appreciate the contrast between, for instance, the indexical "I" and the unspecific "Italian book". In "It is possible that I don't wear glasses", if the referent of "I" in context is Carl, whatever world w_i we consider, it will have to be Carl again. While in "It is possible that I read an Italian book", even though by "Italian book" the speaker means a book written in Italian in the normal way, the modal sentence will be true even in a world w_i where the speaker reads a book which is written in Italian but in a Cyrillic script, or in a world w_j in which the speaker reads a book which has been bought in Italy. The fact

that "Italian book" means "book written in Italian in the normal way" in the context of utterance doesn't seem relevant for the truth-conditions of the modal sentence. Similarly, it is only in modal embeddings that we can appreciate the difference between "It's raining" and "It's raining here"; in "Possibly, it's raining here", if the reference of "here" in the actual world is Paris, it has to be Paris in any other world w_i we would consider. In "Possibly, it's raining", even though the place of utterance is Paris and Paris is salient in conversation, the utterance could be true even in a world w_i where it rains in London.

One could object that the same phenomenon that I have associated with unspecificity also affects indexicals. Let us focus on so-called "pure" or "automatic" indexicals (Kaplan 1989, Barwise and Perry 1983, Perry 2001). One could object that expressions like "here", "now", "today", "tomorrow" might be such that, even if their content has been fixed in context, this is not, strictly speaking, fixed for all circumstances of evaluation. Let's suppose that the referent of "here" in "It is possible that it rains here" is fixed in context *c* as being identical to Paris. This doesn't prevent that an utterance of the modal sentence be true if there is a world w_i where it's not strictly speaking raining in Paris, but in the suburbs of Paris, or in a region that includes Paris but it's larger than its municipality area. This may seem to challenge the idea that, once the content of an indexical is fixed in a context of utterance, it is fixed for all circumstances of evaluation.

I do not believe this objection poses a serious threat to the strategy I employ. As already suggested with respect to "I", the mechanism of indexicality is *designed* in such a way that fixing the content for an indexical leaves no open question as to what its reference is across circumstances of evaluation. It is another matter whether, in some alternative situation, a certain counterfactual object or individual or space region would count as *the same* as the one actually referred to. This is a metaphysical problem. In the case just described of "here", although Paris (and, say, *only* the Paris city area) has been assigned as the reference of "here" in the context of utterance, one could allow a certain amount of flexibility and admit that Paris' suburbs, or a region larger than Paris' municipal area could *count as identical* across possible worlds (for current purposes) to the area actually referred to by that occurrence of "here". The same goes for other indexicals, like "now", "today", "tomorrow" and so on: the mechanism to which they respond is designed to fix their content across circumstances of evaluation; it is another matter whether we may be willing to count certain individuals, objects, space-regions or time-regions as the same as those actually referred to.

To conclude, the contrast that emerges between indexical and unspecific expressions as observed in modal embeddings is the following: on the one hand, when the content of an indexical is fixed in a context, it is fixed for all counterfactual circumstances of evaluation; on the other hand, when the content of an unspecific expression is fixed in a context, the expression's content is not fixed for all circumstances of evaluation. The argument just presented manages to draw attention on a clear difference between indexicality and unspecificity, a difference that can be brought to light once one looks at how these expressions behave in contexts such as modal embeddings, and could be traced to the fact that indexicals are directly referential, while unspecific expressions are not. In this respect, the argument I have presented does better than that proposed by Sainsbury, in that it contributes to a more accurate and sharp appreciation of the difference between unspecificity and indexicality.

Acknowledgments

I am grateful to an anonymous referee for comments and suggestions that helped to improve this paper in many ways. Work on this article has been in part subsidized by the Instituto de Investigaciones Filosóficas at the Universidad Nacional Autónoma de México (project CONACyT CCB 2011 1665), which I thank for its support.

REFERENCES

Barwise, J. and J. Perry. 1983. *Situations and Attitudes*. Cambridge, MA: The MIT Press.

Bezuidenhout, A. 2002. Truth-conditional Pragmatics, *Philosophical Perspectives* 16: 105-34.

Brinck, I. 1997. *The Indexical I. The First Person in Thought and Language*. Kluwer, Dordrecht.

Cappelen, H. and E. Lepore. 2002. Indexicality, Binding, Anaphora and a Priori Truth. *Analysis* 62 (276): 271-81.

Cappelen, H. and E. Lepore. 2005. *Insensitive Semantics: a defense of semantic minimalism and speech act pluralism*. Blackwell, Oxford.

Carston, R. 2002. *Thoughts and Utterances: the pragmatics of explicit communication*. Blackwell, Oxford.

Castañeda, H.-N., 1974. Thinking and the Structure of The World. *Philosophia* 4: 3–40.

Clapp, L. 2012. Indexical Color Predicates: Truth Conditinal Semantics vs. Truth Conditional Pragmatics. *Canadian Journal of Philosophy* 42 (2):71-100.

Collins, J. 2007. Syntax, More or Less, *Mind* 116 (464): 805-850.

Frege, G. 1918/1956. The Thought. *Mind*, New Series: 65, (259): 289-311.

Hall, A. 2008. Free Enrichment or Hidden Indexicals? *Mind and Language* 23 (4): 426-456.

Kaplan, D. 1989. Demonstratives. In *Themes from Kaplan*, ed. Almog, J., Perry, J., and H. Wettstein, 481-563. Oxford: Oxford University Press.

King, J. and Stanley, J. 2005. Semantics, Pragmatics, and the Role of Semantic Content. In *Semantics Versus Pragmatics*, ed. Z. Szabò, 111-164. Oxford: Oxford University Press.

Lewis, D. 1980. Index, Context, and Content. In *Philosophy and Grammar*, ed. Kanger, S. and S. Öhman, 79–100. Dordrecht: Reidel; reprinted in *Papers in Philosophical Logic*, ed. Lewis, D. 1998. Cambridge: Cambridge University Press.

Loar, B., 1976. The Semantics of Singular Terms. *Philosophical Studies* 30: 353–377.

Maier, E. 2009. Proper Names and Indexicals Trigger Rigid Presuppositions. *Journal of Semantics* 26: 253–315.

Mount A. 2008. The impurity of ‘‘pure’’ indexicals. *Philosophical Studies* 138: 193–209.

Neale, S. 2007. On Location. In *Situating Semantics: Essays on the Philosophy of John Perry,* ed. O’Rourke, M. and C. Washington, 251–393. Cambridge, MA: MIT Press.

Nunberg G. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16: 1–43.

Parsons J. 2011. Assessment-Contextual Indexicals. *Australasian Journal of Philosophy*

89 (1): 1-17.
Perry, J. 2001. *Reference and Reflexivity*. Stanford, CA: CSLI Publications.
Predelli S. 1998a. I am not here now. *Analysis*, 58(2): 107–115.
Predelli S. 1998b. Utterance interpretation and the logic of indexicals. *Mind and Language*, 13 (3): 400–414.
Pupa, F., Troseth, E. 2011. Syntax and Interpretation. *Mind and Language* 26 (2): 185-209.
Rabern B. 2013. Monsters in Kaplan's Logic of Demonstratives. *Philosophical Studies* 164 (2): 393-404.
Recanati, F. 2010. *Truth-conditional Pragmatics*. New York: Oxford University Press.
Recanati, F. 2004. *Literal Meaning*. Cambridge: Cambridge Univ. Press.
Rothschild, D. and G. Segal. 2009. Indexical Predicates. *Mind and Language* 24 (4): 467-493.
Sainsbury, R. M. 2002. Two ways to smoke a cigarette. In *Departing from Frege,* Sainsbury, R. M ., 386-406 . London and New York: Routledge.
Santorio P. 2012. Reference and Monstrosity. *Philosophical Review* 121 (3), 359-406.
Schlenker P. 2003. A Plea for Monsters. *Linguistics and Philosophy* 26 (1): 29-120.
Searle, J. 1983. *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
Stanley, J. 2000. Context and Logical Form. *Linguistics and Philosophy* 23: 391-424.
Stanley, J., and Z. Szabò 2000. On Quantifier Domain Restriction. *Mind and Language* 15: 219-61.
Szabò, Z. G. 2001. Adjectives in Context. In *Perspectives on Semantics, Pragmatics, and Discourse*, ed. Harrish R. and I. Kenesei, 119-147. John Benjamins Publishing Company.
Travis, C. 1997. Pragmatics. In *A Companion to the Philosophy of Language*, ed. Hale, B. and C. Wright , 87-107 . Oxford: Blackwell.

Received: February 5, 2013
Accepted: April 8, 2014

Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México
Circuito Maestro Mario de la Cueva s/n,
Ciudad Universitaria,
C.P. 04510, Coyoacán
México, D.F
deliabelleri@gmail.com

EuJAP | Vol. 10 | No. 1 | 2014
UDK 165.1/.2

ON THE RATIONAL IMPOTENCE OF URGES

SIMON RIPPON
Central European University

Can a Humean subjectivist – someone who believes that all our practical reasons are ultimately grounded in our desires – accommodate the intuition that certain of our basic (i.e., ultimate or non-instrumental) desires should have no weight at all in practical deliberation? Consider Harry Frankfurt's example of a loving father who finds himself with a sudden, unmotivated "murderous impulse" against his beloved son (2006, 12). We are inclined to say, intuitively – and Frankfurt agrees – that such a desire would provide the father with *no reason at all* to murder his son: it is "rationally impotent". The father would be making a serious mistake were he to treat the murderous impulse as having any normative weight whatsoever: it should be "extruded", "disenfranchised" or "silenced" in practical reasoning.

In his article in this journal, Regan Lance Reitsma accepts such intuitions about the "rational impotence" of some desires at face value, but doubts that Harry Frankfurt's well-developed subjectivist view can properly account for them (2013). Frankfurt wants to maintain both, on the one hand, the subjectivist claim that all our practical reasons are explained by our basic desires and, on the other, the realist-sounding claim that some basic desires are rationally impotent (as in cases like an addict's strong desire for heroin, or of the loving father who experiences an impulse to kill his son). Although these claims are not logically inconsistent, there is a certain tension between them, since their combination raises the question: Why do some desires ground practical reasons, while others do not?

Frankfurt is, of course, well-known for making distinctions between different kinds of desires. In particular he is known for having developed an account of *freedom of the will* in terms of higher-order desires (desires to desire), according to which I act freely when I act on a desire that I identify with (i.e., I have a higher-order desire that I act on the first-order desire, and I am satisfied with the higher-order desire, in the sense that I lack any inclination to change it). Frankfurt explains our caring and loving as essentially consisting in diachronic, reflective, higher-order desires of this kind, and identifies these as the sources of our weightiest practical reasons (1971; 2006). However, Reitsma thinks Frankfurt's distinctions on their own cannot explain the fact that some of our basic desires are rationally impotent. As Reitsma puts it:

> It wouldn't really help to appeal to Frankfurt's idea that some of the father's desires are "cares" or "loves" ... If caring about is a complex motivational state, a set of desires, then what we have is a case of desires (ruling passions) in competition with other desires (impulses), which, it seems, is to be handled by a trumping model ... What is needed for the disenfranchisement, the "categori-

> cal" rejection of a basic desire, it seems, is something with a different "shape" than another desire – something such as a norm or a rule or a set of criteria. (Reitsma 2013, 58)

Reitsma further characterises this objection to Frankfurt as a "new twist" on Watson's (1975) objection to Frankfurt (2013, 58n21). Watson argues that Frankfurt's view fails to explain why higher-order desires deserve special standing (specifically, he claims that there is no reason to think that higher-order desires should count as less "wanton" and more an agent's "own" desires than first-order desires). According to Watson, the fundamental work in Frankfurt's account is performed not, as Frankfurt suggests, by distinguishing between orders of motivation, but by positing (arguably by fiat) acts of "decisive commitment" or "identification with" a desire.

Reitsma's response to the puzzle of explaining rationally impotent desires is to introduce a new kind of practical consideration within Frankfurt's framework. We can describe Reitsma's proposal in three basic steps: (i) some of the things agents care about are *personal ideals*, (ii) so-called "norms of rational impotence" are constitutive elements of some personal ideals, and (iii) these norms of rational impotence require the agent not to treat certain kinds of basic desires as having normative significance.

How is Reitsma's solution supposed to improve on Frankfurt's original account of distinctions between higher-order ruling passions and first-order impulses? Can't Frankfurt just stipulate that what I will call "urges" – first-order impulses that we don't identify with in any way – have no normative weight? Reitsma's thought seems to be that, because Frankfurt's picture only contains competition between desires of different levels, a father's violent urge should still "count against" the reason to favour his son's welfare, even if it is trumped by the weightier consideration of the father's love. Pointing out that a desire is just an urge is not enough, Reitsma seems to think, to justify giving it zero weight in practical deliberation. So Reitsma seeks to provide an explanation of how the father's reason to act on this impulse is entirely extinguished by the father's conflicting concerns. He explains this by appealing firstly to the fact that the agent cares about the personal ideal of being a loving father, and secondly to the claim that the personal ideal of being a loving father *demands* that impulses to harm one's offspring are treated as utterly normatively insignificant.

Reitsma offers a sensitive account of the problem of rational impotence for subjectivists, and an intriguing proposed solution that is sympathetic to Frankfurt's broader framework. But I will argue that Reitsma's proposal holds an unstable position within Frankfurt's framework. If Reitsma's personal ideals can truly achieve the kind of categorical desire "extrusion" that Reitsma intends, then Frankfurt's account must already contain sufficient resources to explain why urges are rationally impotent. This is because, as I will argue, Reitsma's account rests on the presumption that cares necessarily outweigh urges *because of* their special place in a person's motivational structure. But once it is accepted that a desire's position within a person's motivational structure can affect its normative weight, there is no reason to deny that there is a certain position within a person's motivational structure – the position that urges occupy – that carries no normative weight whatsoever.

And this conclusion makes Reitsma's appeal to personal ideals superfluous to explaining the rational impotence of urges.

To see the problem, let us begin by asking the following question: On Reitsma's account, what determines whether one's desire to follow a particular personal ideal, such as the ideal of being a loving father, should be followed? Consider, for example, Reitsma's example of "a person without natural athleticism" who "takes on the daily routine of a striving and ambitious jock and suffers under the weight of his (predictable) failures" (2013, 62). Reitsma implicitly adverts to a prudential principle; something like the idea that one ought to order one's pursuit of one's ends so as to maximise the degree to which one achieves them overall. Reitsma does not state the exact form of the prudential principle he endorses. But as he explains his example, the would-be athlete has a strong practical reason to reject the personal ideal if it "places demands upon him that make it either impossible or exceedingly difficult for him to fulfil other ends that he cares about, or even cares about more," whereas it would be rational to submit to the ideal if the agent cares about it and if "some of his ruling passions favour it, none speak significantly against it" (2013, 62).

As Christine Korsgaard has pointed out, a prudential principle of the kind invoked here cannot be explained by the instrumental principle: the latter is completely silent about the ordering of one's ends overall (1997). In a footnote, Reitsma says that on his own subjectivist view, a principle of prudence only has rational validity for a person insofar as it can be derived from a person's own basic desires, such as a concern for "her own (long-term) welfare" (2013, 51n9). It is unclear to me how such an account could explain the normative force of the prudential principle without already presupposing it, as it raises the question: what gives the relevant basic desire(s) (e.g. the desire to look after one's own long-term welfare) primacy over all of one's other desires (e.g. an addict's desire for heroin, or the would-be athlete's desire to become an athlete)? But let us grant for the sake of argument that this question can be answered; and although it was not specified in the example, let us assume that Reitsma's would-be athlete also possesses whatever basic desires are considered necessary for grounding the application of Reitsma's preferred formulation of the prudential principle.

At this point the following difficulty arises: *If* Reitsma is right that the characterisation of ruling passions as higher-order desires provides no special resources for moving beyond a "trumping" model of competition between desires, then it is unclear how Reitsma's introduction of personal ideals can ever leave that model behind. This is because Reitsma accepts that any particular desire to follow a personal ideal (being a loving father, not being an addict) must be weighed against other desires in order to determine whether it ought to be followed. And although Reitsma, when he writes about the would-be athlete, only considers weighing the desire to follow the personal ideal against other ruling passions, it is unclear why *other* basic desires should not also be considered. If one must weigh one's desire to be an athlete against, for example, one's care that one is successful and respected in one's chosen field, then mustn't one also weigh it against one's impulse to lounge around? And if one must weigh the desire to be an athlete against the impulse to lounge around, then mustn't one similarly weigh the desire not to be an addict against the impulse to take heroin, and the desire to be a loving father against the impulse to murder one's child? My

point is this: according to Reitsma's account, a loving father might rationally "extrude" the impulse to murder the child on the grounds that he desires to live up to the personal ideal of being a loving father, and that the personal ideal demands extrusion of such impulses from practical deliberation. But like Reitsma's would-be athlete, he might now ask himself the question: *shall* I be a loving father? And when *that* question is asked, there appear to be no rational grounds for setting aside the very desires that the personal ideal would require him to extrude, or at least no grounds *other than* the fact that one desire is a personal ideal and the other is a mere impulse. Why can't it sometimes be prudent to give up a personal ideal in the light of a mere first-order desire that the personal ideal treats as a candidate for extrusion, rather than the other way around?

Reitsma faces a dilemma here. On the first horn, he might accept that giving up a personal ideal in the light of one or more first-order desires – even first-order desires that the personal ideal requires us to extrude – could sometimes be prudent. But if this is so, then the loving father's "categorical" rejection of the murderous impulse was not nearly as categorical as Reitsma's account first suggested, since the reason provided by his desire to kill might yet outweigh the reason provided by his desire to maintain his identity as a loving father. To find out whether it does, he needs to weigh one against the other. On the second horn, Reitsma could deny that rejecting a personal ideal in favour of satisfying a first-order desire is ever prudent. But since Reitsma accepts that rejecting a personal ideal in favour of other things you care about *is* sometimes prudent, this must be explained by some fundamental difference in significance between first-order and higher-order motivations; between impulses with which we don't identify and ruling passions with which we do. If Reitsma now accepts that there *is* some fundamental difference in significance between ruling passions and impulses, and that this is sufficient to explain why the reason provided by a ruling passion necessarily outweighs the reason provided by an impulse, then he must also accept the following claim: Something else is relevant to facts about what reasons one has *beyond* facts about what one's basic desires are for, together with the instrumental principle. In particular, whether a basic desire is a ruling passion or an urge will be held to make a normative difference. This difference will presumably be related to the higher-order nature of ruling passions, or to the fact that we identify with them. Whatever its explanation, exactly, the following question arises: If it is a fact about the nature of ruling passions that they provide us with weighty normative reasons, then why can't it be just as much a fact about impulses which we do not, in any respect, identify with, that they have no normative weight at all?[1] And if urges have no normative weight, then the machinery of personal ideals and norms of silencing is no longer needed to explain their rational impotence.

The view that urges have no normative weight will, of course, make different predictions than Reitsma's own rational impotence thesis, since it does not require a norm relating to a personal ideal to extrude each particular impulse deemed rationally impotent. But the differing predictions of the view I suggest here are intuitively plausible. Take Cohon's

[1] I am not primarily interested in Frankfurt interpretation here, but there is strong evidence that this is actually Frankfurt's view. For example, he writes: "the mere fact that a person has a desire does not give him a reason. What it gives him is a problem. He has the problem of whether to identify with the desire and thus validate it as eligible for satisfaction ... If he identifies with the desire, he acknowledges that satisfying it is to be assigned *some* position – however inferior – in the order of his preferences and priorities" (Frankfurt 2006, 11).

example, cited by Reitsma, of "the passing urge to stick my finger into a gooey substance" (2000, 63), where this act which has no objective value; or Quinn's example of a man who feels an urge to turn on radios for no reason whatsoever (Quinn 1994). It is intuitively plausible that these urges do not provide normative reasons for action, and this intuition seems not to depend on the existence of any personal ideal that specifically excludes such desires from providing reasons. It is perhaps telling that when Reitsma writes "Personally, I'm not convinced everyone would find a desire to stick a finger in goo rationally impotent", he then gives a *further* ground for performing the act one has a mere urge to perform, "there is, about it, something wonderfully antagonistic to bourgeois standards of cleanliness" (2013, 57). If mere urges were not intuitively rationally impotent in the first place, there should have been no need to advert to any other justification for sticking a finger in the goo. Reitsma mentions the example of "an afternoon's urge to eat an ice cream cone" as capable of generating, on Frankfurt's account, the "final 'ought' of practical reason." (2013, 53) Read in one way, it is plausible that it might. But it is difficult to see such a desire as a true "urge" in my sense; that is, as an impulse that we do not identify with in any way. After all, there are many good reasons to endorse acting on a desire to eat ice cream of an afternoon; in particular, ice cream will taste nice and bring us pleasure. If we take more care with our examples, it becomes clearer that mere urges do not give us reasons. An afternoon's urge to swallow a (harmless) stone does not seem, intuitively, to make it the case that one ought to do it.

Reitsma's concept of a personal ideal seems to offer a promising subjectivist explanation for why some basic desires are to be treated as rationally impotent. But a difficulty arises when we consider the question of which personal ideals one ought to follow, and why. Either those desires which occupy one place in a person's motivational structure, the ruling passions, are granted a special normative authority over other desires which occupy a different place, urges, in which case it is unclear why urges should not be said to lack normative authority altogether. This makes the concept of a personal ideal otiose to explaining the rational impotence of urges. Or all desires are on an equal footing, in which case reasons generated by our ruling passions to follow personal ideals need to be weighed against reasons generated by the very impulses that they were supposed to extrude from practical reasoning.

REFERENCES

Cohon, Rachel. 2000. The Roots of Reasons. *The Philosophical Review* 109 (1): 63–85.

Frankfurt, Harry G. 1971. Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68 (1): 5–20.

———. 2006. *Taking Ourselves Seriously and Getting It Right*, ed. D. Satz. Stanford, CA: Stanford University Press.

Korsgaard, Christine M. 1997. The Normativity of Instrumental Reason. In *Ethics and Practical Reason*, ed. G. Cullity and B. Gaut, 215–54. Oxford: Clarendon Press.

Quinn, Warren. 1994. Putting Rationality in Its Place. In *Morality and Action*. Cambridge, MA: Cambridge University Press.

Reitsma, Regan Lance. 2013. How a Modern-Day Hume Can Reject a Desire Categorically: A Perplexity and a Theoretically Modest Proposal. *European Journal of Analytic Philosophy* 9 (2): 48–66.
Watson, Gary. 1975. Free Agency. *The Journal of Philosophy* 72 (8): 205–20.

Central European University
Department of Philosophy
Nador u. 9
1051 Budapest, Hungary
rippons@ceu.hu

THE PREOCCUPATION AND CRISIS OF ANALYTIC PHILOSOPHY

PREOKUPACIJA I KRIZA ANALITIČKE FILOZOFIJE

MICHAEL LOSONSKY
Colorado State University

ABSTRACT

I propose to reconsider Gilbert Ryle's thesis in 1956 in his introduction to The Revolution of Philosophy that "the story of twentieth-century philosophy is very largely the story of this notion of sense or meaning" and, as he writes elsewhere, the "preoccupation with the theory of meaning is the occupational disease of twentieth-century Anglo-Saxon and Austrian philosophy." Ryle maintains that this preoccupation demarcates analytic philosophy from its predecessors and that it gave philosophy a set of academic credentials as a rigorous discipline with its own domain and method. I will maintain that Ryle, with some minor qualifications, was correct in his assessment of the nature of analytic philosophy at that time, and I will argue that the next 50 years continued to be, very largely, the story of meaning, exemplified by the groundbreaking work of Rawls and Kripke. However, I argue that this work also contains the seeds that contributed to the emergence of philosophies that represent a significant departure from analytic philosophy.

Keywords: analytic philosophy, logic, meaning, naturalism, Ryle, Rawls, Kripke, 20th-century philosophy

SAŽETAK

Predlažem reviziju teze Gilberta Rylea iz 1956. iz njegovog uvoda u *The Revolution of Philosophy* prema kojoj "priča o filozofiji dvadesetog stoljeća jest naširoko priča o toj ideji o smislu ili značenju" te, kao što piše drugdje, da "preokupacija teorijom o značenju jest profesionalna bolest anglosaksonske i austrijske filozofije dvadesetog stoljeća." Ryle prihvaća da ta preokupacija razgraničava analitičku filozofiju od njezinih prethodnika te da joj je dala skup akademskih zasluga kao rigorozne discipline sa svojom vlastitom domenom i metodom. Prihvaćam da Ryle, s omanjim kvalifikacijama, jest bio u pravu u svojoj ocjeni naravi analitičke filozofije toga vremena, te ću argumentirati da je sljedećih 50 godina nastavilo biti, uvelike, pričao značenju, oprimjerena inovativnim radom Rawlsa i Kripkea. Ipak, argumentiram da taj rad također sadrži sjeme koje je doprinijelo nastajanju filozofija koje predstavljaju značajni otklon od analitičke filozofije.

Ključne riječi: analitička filozofija, logika, značenje, naturalizam, Ryle, Rawls, Kripke, filozofija dvadesetog stoljeća

THE PROBABILITY OF THE POSSIBLE

VJEROJATNOST MOGUĆEG

RON WILBURN
University of Nevada

ABSTRACT

In "Why is There Anything at All?" Peter van Inwagen argues that even though it was never necessary that concrete beings existed, it was always maximally probable – just short of necessity – that they did (van Inwagen 1996). I argue that van Inwagen's argument fails, albeit for an interesting reason which has remained so far unnoticed in the literature: there is a critical tension between two of its premises, both essential to its soundness, concerning the nature of comprehensively specified possible worlds. I will summarize van Inwagen's argument, develop this objection, and then describe more problems which invariably accrue when we try to ascribe probability values to possible worlds.

Keywords: existence, possible worlds, probability, van Inwagen, Hawking

SAŽETAK

U "Zašto išta uopće postoji?" Peter van Inwagen argumentira da iako nikada nije bilo nužno da materijalna bića postoje, uvijek je bilo maksimalno vjerojatno — samo zakinuto za nužnost — da ona postoje (van Inwagen 1996). Također, argumentiram da je van Inwagenov argument neuspješan, iako je tome tako zbog zanimljivog razloga koji je dosad ostao nezapažen u literaturi: postoji kritična napetost između njegovih dviju premisa, obiju esencijalnih za njegovu valjanost, koje se tiču naravi sveobuhvatno specificiranih mogućih svjetova. Sažet ću van Inwagenov argument, razviti taj prigovor, te ću nakon toga opisati još neke probleme koji se počnu nagomilavati kada mogućim svjetovima pripisujemo vrijednosti vjerojatnosti.

Ključne riječi: postojanje, mogući svjetovi, vjerojatnost, van Inwagen, Hawking

WHAT IS PROBABILITY AND WHY DOES IT MATTER

ŠTO JE VJEROJATNOST I ZAŠTO JE BITNA

ZVONIMIR ŠIKIĆ
University of Zagreb

ABSTRACT

The idea that probability is a degree of rational belief seemed too vague for a foundation

of a mathematical theory. It was certainly not obvious that degrees of rational belief had to be governed by the probability axioms as used by Laplace and other prestatistical probabilityst. The axioms seemed arbitrary in their interpretation. To eliminate the arbitrariness, the statisticians of the early 20th century drastically restricted the possible applications of the probability theory, by insisting that probabilities had to be interpreted as relative frequencies, which obviously satisfied the probability axioms, and so the arbitrariness was removed. But the frequentist approach turned more subjective than the prestatistical approach, because the identifications of outcome spaces, the choices of test statistics, the declarations of what rejection regions are, the choices of null-hypothesis among alternatives, the contradictory choices between sizes and powers etc., depend on thoughts or even whims of the experimenter. Frequentists thus failed to solve the problems that motivated their approach, they even exacerbated them. The subjective Bayesianism of Ramsey and de Finetti did not solve the problems either. Finally Cox provided the missing foundation for probability as a degree of rational belief, which makes the Bayesian probability theory (which is based on this foundation) the best theory of probable inference we have. Hence, it is quite unbelievable that it is not even mentioned in recent philosophy textbooks devoted to the probable inference. The reason could be that it requires fairly sophisticated mathematics. But not even to mention it? We explain the history and prove Cox theorem in a novel way.

Keywords: probability, subjective Bayesianism, logical Bayesianism, Cox theorem

SAŽETAK

Ideja da je vjerojatnost stupanj racionalnoga vjerovanja činila se prenejasnom za temelje matematičke teorije. Zasigurno nije očito da su stupnjevi racionalnoga vjerovanja morali biti upravljani aksiomima vjerojatnosti koje su koristili Laplace i ostali predstatistički probabilisti. Aksiomi su se činili proizvoljnima u svojoj interpretaciji. Kako bi eliminirali proizvoljnosti, statističari ranoga dvadesetog stoljeća drastično su ograničili mogućnosti aplikacije probabilističke teorije, inzistirajući da su vjerojatnosti morale bitiinterpretirane kao relativne frekvencije, koje su očito zadovoljavale aksiome vjerojatnosti, te je tako arbitrarnost bila otklonjena. No pristup frekventista ispao je subjektivniji nego predstatistički pristup jer identifikacije izlaznih mjesta, odluke o testnim statistikama, deklaracije o tome što su regije za odbacivanje, odluke o nultoj hipotezi među alternativama, kontradiktorne odluke među veličinama i snagama itd., ovise o mislima ili čak o hirovima eksperimentatora. Frekventisti tako ne uspijevaju riješiti probleme koji su motivirali njihov pristup, već su ih pogoršali. Subjektivnibayesijanizam Ramseyjeva i de Finettijeva tipa nije također riješio probleme. Konačno, Cox je pružio temelj za vjerojatnost kao stupanj racionalnog vjerovanja koji je nedostajao, i koji je učinio bayesijansku teoriju vjerojatnosti (baziranu na tim temeljima) najboljom teorijom vjerojatne inferencije koju imamo. Stoga je vrlo nevjerojatno da to nije spomenuto u recentnim filozofskim udžbenicima posvećenima vjerojatnoj inferenciji. Razlog tome bi mogao biti to što zahtijeva značajno sofisticiranu matematiku. No da čak nije bilo ni spomenuto? Objasnit ćemo povijest i dokaz Coxova teorema na nov način.

Ključne riječi: vjerojatnost, subjektivni bayesijanizam, logički bayesijanizam, Coxov teorem

WHY SEMANTIC UNSPECIFICITY IS NOT INDEXICALITY

ZAŠTO SEMANTIČKA NEODREĐENOST NIJE INDEKSIKALNOST

DELIA BELLERI
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México

ABSTRACT

In this paper, I address the idea that certain sentences ("It's raining", "The leaves are green", "Sally reads an Italian book") suffer from what is generally called semantic unspecificity: their meaning is determinate, but their truth conditions are not. While there tends to be agreement on the idea that semantic unspecificity differs from phenomena such as ambiguity and vagueness, some theorists have defended an account which traces it to indexicality, broadly construed. Some authors have tried to vindicate the distinction between unspecificity and indexicality and, in this paper, I pursue the same cause, but with a critical stance towards previously employed strategies. I urge that the central argument employed by Sainsbury to trace this difference fails suitably to set unspecificity apart from indexicality and I propose a new argument, which helps to trace this distinction more perspicuously. The argument is based on embeddings of indexical and unspecific expressions within modal operators and on the ways in which the truth conditions of utterances of the resulting, complex expressions are thereby affected.

Keywords: Semantic unspecificity, semantic under-determinacy, indexicality, rigidity, intensional contexts

SAŽETAK

U ovome članku propitkujem ideju da određene rečenice ("Kiši", "Lišće je zeleno", "Sally čita talijansku knjigu") pate od nečeg što se općenito zove semantička neodređenost: njihovo značenje je određeno, no njihovi uvjeti istinitosti nisu. Dok postoji tendencija slaganja oko ideje da se semantička neodređenost razlikuje od pojava poput dvoznačnosti ili nejasnoće, neki teoretičari brane pristup koji se poziva na indeksikalnost, u širokom smislu. Neki autori su pokušali opravdati razliku između neodređenosti i indeksikalnosti i, u ovome članku, bavitću se istim razlogom, no s kritičkim osvrtom na prethodnoupotrebljavane strategije. Zauzimam se da središnji argument koji Sainsbury upotrebljava da bi pratio tu razliku ne uspijeva prikladno rastaviti nespecifičnosti od indeksikalnosti te predlažem novi argument, koji pomaže pratiti to razliku jasnije. Argument se bazira na ugradnji indeksikalnosti i nespescifičnosti izražene unutar modalnih operatora te na načinima na koji uvjeti istinitosti iskaza nastalih, složenih izraza jesu time zahvaćeni.

Ključne riječi: semantička neodređenost, semantička podređenost, indeksikalnost, rigidnost, intenzionalni konteksti

ON THE RATIONAL IMPOTENCE OF URGES

O RACIONALNOJ IMPOTENCIJI NAGONA

SIMON RIPPON
Central European University

Discussion, no abstract.
Diskusija, nema sažetka.

INSTRUCTIONS FOR AUTHORS

Submission

Submitted manuscripts ought to be:

- unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere. If it emerges that an accepted submission does not satisfy the previous conditions, the Editors reserve the right to reject it. If, due to the authors' failure to inform the Editors, already published materials will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue. In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors
- sent exclusively electronically to the Editors (eujap@ffri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format
- prepared for blind refereeing: authors` names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed
- accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 250 words), keywords, academic affiliation and full address for correspondence including e-mail address
- in American or British English
- no longer than 8000 words, including footnotes and references

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the *author-date Chicago style*, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book

T: (Nozick 1981, 203)

R: Nozick, R. 1981. *Philosophical Explanations*. Cambridge: Harvard University Press.

Chapter or other part of a book

T: (Fumerton 2006, 77-9)

R: Fumerton, R. 2006. The Epistemic Role of Testimony: Internalist and Externalist Perspectives. In *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa, 77-92. Oxford: Oxford University Press.

Edited collections

T: (Lackey and Sosa 2006)

R: Lackey, J. and E. Sosa, eds. 2006. *The Epistemology of Testimony.* Oxford: Oxford University Press.

Article in a print journal

T: (Broome 1999, 414-9)

R: Broome, J. 1999. Normative requirements. *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)
R: Skorupski, J. 2010. Sentimentalism: Its Scope and Limits. *Ethical Theory and Moral Practice* 13: 125-136. http://www.jstor.org/stable/40602550

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013. http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications.* Unpublished manuscript, last modified August 3, 1951.

Final proofreading
Authors are responsible for correcting proofs.

Permission to republish
For permission to republish material in print or electronic format please contact the Editors (eujap@ffri.hr).

Subscriptions

A subscription comprises two issues.
All prices include postage.

Annual subscription:
International:
individuals € 10
institutions € 15
Croatia:
individuals 40 kn
institutions 70 kn

Bank: Zagrebačka banka d.d. Zagreb
SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455
Only for subscriptions from Croatia,
please add: "poziv na broj": 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed in The Philosopher's Index.