Really? You're a Philosopher?

An Explanation of What I Do, for Non-Philosophers

Agustín Rayo

My research lies in the intersection of the philosophy of language and the philosophy of logic and mathematics. I have worked on understanding the relationship between our language and the world it represents, on clarifying certain connections between logic and mathematics, and on investigating the limits of communicable thought.

If you'd really like some additional detail, read on...

Linguistic Representation

Many philosophers think that language-use is governed by rules, in much the way that a game of chess is governed by rules. In chess, a bishop is only used correctly (i.e. in accordance with the rules of chess) when it is moved diagonally; in conversation, the sentence ‘I am hungry’ is only used truthfully (i.e. in accordance with the rules of assertion) when the speaker is hungry at the time of utterance.

This way of thinking about language is useful for certain purposes. But I believe it is inaccurate as a description of our everyday discourse, which is inevitably vague. Suppose I utter ‘John is bald’. How many hairs must there be on John’s head in order for my assertion to count as true? The vagueness of ‘bald’ makes it implausible to think that there is a precise answer to be given. And in the absence of a precise answer, it is not clear what a rule governing the usage of ‘John is bald’ would look like.

Some of my research has focused on developing an alternative to the rule-based conception of language. The basic idea can be explained with an analogy. Suppose you are asked to determine Susan’s age by looking at her face. Your examination will allow you to rule out certain possibilities. It might make you extremely confident, for instance, that Susan is more than five years old, and fairly confident that she is somewhere in her sixties. But your examination won’t put you in a position to determine Susan’s exact age: you are not in possession of a rule for transforming your evidence into a precise statement of Susan’s age-in-years (to say nothing of a statement of Susan's age-in-days). If pressed to produce such a statement, you would have no choice but to go beyond what you are able to ascertain on the basis of your evidence. You would make a *decision* which is responsive to your evidence without being settled by your evidence.

What my research suggests is that language-mastery should be modeled in a similar sort of way. When you say ‘John is bald’ I am able to eliminate certain possibilities. But this is not because I deploy a ‘rule of English’ to eliminate possibilities in which the sentence would have been used incorrectly. It is because I make a *decision*: I use your utterance to discriminate amongst the options at my disposal in a way which is responsive to, but not settled by, our community's linguistic practice.

One might worry that this would lead to an unacceptable degree of arbitrariness. But I argue for a certain brand of *localism*: the idea that the possibilities that are treated as live options for the purposes of a conversation tend to be fairly restricted, and that when they are restricted enough, competent speakers are usually in a position to make fairly principled decisions by making intelligent use of contextual information.

In my papers 'Vague Representation' and 'A Plea for Semantic Localism', I develop a rigorous model of linguistic communication based on this idea. ('Vague Representation' was selected by *The Philosopher's Annual* as one of the ten best papers published in philosophy during 2008.)

I have also shown that the proposal can be used to give a unified treatment of two seemingly disparate problems that have been of great concern to philosophers: (1) the problem of vagueness (i.e. the problem of explaining how a word like 'bald' can be used in communication even though there is no sharp distinction between the bald and the non-bald); and (2) the Liar Paradox (i.e. the problem of modeling the semantic behavior of paradoxical sentences such as 'what this very sentence says is not true'). My work suggests that these problems are only intractable when one assumes that they should be addressed by identifying 'rules of English' to govern the behavior of vague expressions like 'bald' and semantic vocabulary like 'true'. If, on the other hand, one thinks of communication as an exercise in sensitivity to context and common sense, then the phenomena can be modeled in natural and satisfying ways.

The proposed model of linguistic communication also turns out to have interesting applications in the field of artificial intelligence. Dustin Smith has used some of the central components of the model as part of his doctoral work at MIT's Media Lab, in which he designed a procedure for improving the ability of computers to make sensible use of contextual information when interpreting human inputs.

Logical Space

Think of *logical space* as the set of possibilities that we treat as 'live options' when we theorize about the world.

Some philosophers have assumed that our conception of logical space can be developed *a priori* (that is, independently of an empirical investigation). Scientific inquiry can therefore be neatly divided into two distinct components. There is, first, the *a priori* task of determining which possibilities to work with, and, second, the *a posteriori* (or empirical) task of determining which of these possibilities corresponds to the way the world actually is. One of the main claims of my book, *The Construction of Logical Space*, is that the right way of thinking about the matter is far more interesting. I argue that it is impossible to draw a clean separation between the task of deciding which possibilities to work with and the task of determining which of these possibilities correspond to the way the world is.

On the view I defend, our conception of logical space is shaped by our acceptance or rejection of ‘just is'-statements: statements like 'to be composed of water *just is* to be composed of H2O' (in which one identifies the property of being composed of water with the property of being composed of H2O), or 'to be in psychological state P *just is* to be in brain state B' (in which one identifies a given psychological state with a particular brain state).

When one accepts a 'just is'-statement one restricts one's conception of logical space. Someone who accepts ‘to be composed of water *just is* to be composed of H2O', for example, will think it *absurd* to suppose that a glass could contain water without containing H2O, since she will see no difference between being composed of water and being composed of H2O. And when one treats a supposition as absurd one treats it as lying outside logical space: one does not consider it a 'live option' in one's theorizing about the world.

The acceptance of a ‘just is'-statement comes with costs and benefits. The benefit is that there are fewer possibilities to rule out in one's search for the truth, and therefore fewer explanatory demands on one's theorizing; the cost is that one has fewer distinctions to work with, and therefore fewer theoretical resources. In deciding whether to accept a ‘just is'-statement one strives to find a balance between these competing considerations. Different ‘just is'-statements can be more or less hospitable to one's theoretical goals. So the decision to accept a 'just is'-statement should be grounded on its ability to combine with the rest of one's theorizing to deliver a fruitful tool for scientific inquiry. And because one's conception of logical space is shaped by the ‘just is'-statements one accepts, this yields the promised result that one's conception of logical space cannot be cleanly separated from the rest of one's scientific theorizing.

The resulting picture of logical space can be used to shed light on an elusive philosophical notion that is sometimes referred to as 'metaphysical possibility'. One of the theorems I prove in the book is that there is a precise sense in which one can use the set of true 'just is'-statements to fix the limits of metaphysical possibility.

Logic and Mathematics

In ordinary discourse the sentences 'the number of dinosaurs is Zero' and 'there are no dinosaurs' are used more or less interchangeably. In philosophy, however, these sentences are sometimes thought to have very different subject-matters. Whereas the former is thought to be partly about the number Zero (and therefore about a causally inert realm of abstract mathematical objects), the latter is thought to be entirely about the non-mathematical question of whether there are any dinosaurs.

This way of thinking leads to some embarrassing problems. Amongst other things, it leaves room for the view that one could know that 'there are no dinosaurs' is true, without knowing that 'the number of the dinosaurs is Zero' is true, on the grounds that we have no good way of investigating the causally inert realm of abstract objects that the number Zero inhabits.

One of the objectives of *The Construction of Logical Space* is to argue that this is a pseudo-problem, brought about by faulty philosophical assumptions. I argue, in particular, that it is a mistake to think that 'the number of the dinosaurs is Zero' and 'there are no dinosaurs' have different subject-matters. The right thing to say is that the two sentences are different descriptions of a single feature of reality. In other words: for the number of the dinosaurs to be Zero *just is* for there to be no dinosaurs.

On the resulting view, there is no clear distinction between mathematical and non-mathematical facts. For instance, the non-mathematical fact that there are no dinosaurs can also be thought of as the (partly) mathematical fact that the number of dinosaurs is Zero. One of the key results of the book is that this idea can be generalized. I show that there is a precise sense in which the mathematical fact described by an arbitrary sentence in the language of arithmetic (or, indeed, in the language of set theory) can be 'reconceptualized' as a non-mathematical fact. (A bit more precisely, what I show is that there is a compositional assignment of non-mathematical truth-conditions to arbitrary mathematical sentences.)

This is a surprising (and welcome!) result. It might, however, lure one into thinking that something stronger is true. One might think that it is in principle possible to *paraphrase* arbitrary mathematical sentences as sentences containing no mathematical vocabulary. For if mathematical *facts* can be reconceptualized as a non-mathematical facts, shouldn't it also be the case the mathematical *sentences* that are used to describe these facts can be rewritten so as to avoid mathematical vocabulary?

Unfortunately not. Although it is easy to come up with non-mathematical paraphrases for simple sentences like 'the number of dinosaurs is Zero' or '2 + 2 = 4', some of my earlier work shows that, under reasonable assumptions, it is logically impossible for there to be a paraphrase-method that works in every case.

Taken together, these formal results help clarify a philosophical position that is sometimes referred to as 'logicism': the view that mathematics is ultimately reducible to logic. What the results show is that there is a sense in which logicism is true, and a sense in which it is not. Logicism has a grain of truth in that there is a precise sense in which mathematical facts could always be reconceptualized as non-mathematical facts (and, in certain special cases, as purely logical facts). But logicism is false in that we cannot entirely escape mathematical vocabulary when we use language to describe the relevant facts. (Conspicuously, the proof that mathematical facts can always be 'reconceptualized' as non-mathematical facts makes essential use of mathematical vocabulary; one uses a mathematical theory to specify non-mathematical truth-conditions for arbitrary mathematical sentences.)

These formal results lay the groundwork for a philosophical account of mathematical knowledge that I develop in *The Construction of Logical Space.* According to this view, the content of a theorem of pure mathematics will always be *trivial*; there is, in other words, no particular way the world needs to be in order for the theorem to be true. A consequence of this view is that our knowledge of such a theorem cannot be modeled as the possession of information about some particular way for the world to be. I suggest, instead, that cognitive accomplishment in mathematics ought to be modeled as a particular kind of information-transfer ability: an ability to deploy information that was available for the purposes of one kind of task in the service of another.

One of my current research projects is a collaboration with Princeton philosopher Adam Elga, and is aimed at developing this idea into an exciting new model of cognitive accomplishment for subjects with limited computational abilities.

Higher-Level Languages

In English, we use the word ‘elephant’ to talk about elephants. But the word might have been used differently. It might have been used to talk about cabbages, say. A *model*, in the logician’s sense, is a possible assignment of meanings to words. According to some models, ‘elephant’ is assigned its actual meaning; according to others it is assigned a deviant meaning (the meaning of ‘cabbage’, as it might be).

Although models have proved to be an extraordinarily fruitful tool in the study of language, there is an important shortcoming implicit in the standard notion of a model. The problem arises because the meaning of a word is represented by a *set* of objects -- the meaning of ‘elephant’, for instance, is represented by the set of elephants -- and it is a consequence of Russell’s famous paradox that there are limits to how many objects a set can contain. The elephants are few enough form a set (and so are the natural numbers, even though there are infinitely many of them). But it is standardly assumed that there are too many sets to fit into a set. The result is that certain models go missing; there is, for example, no model according to which ‘set’ is assigned a set corresponding to its actual meaning.

Some of my early work was geared towards overcoming this difficulty. In collaboration with Gabriel Uzquiano, I was able to characterize an improved conception of a model. The trick was to make use of the enhanced expressive capabilities of *second-level languages*. In a first-level language -- the default language of philosophers and mathematicians -- one is only allowed to speak of properties that are instantiated by objects individually (for instance, the property that one ascribes to someone when one says that she individually lifted a piano). In a second-level language, by contrast, one is able to speak of properties that are instantiated collectively by several objects at once (for instance, the property that one ascribes to the movers when one says that they together lifted a piano, even if no mover lifted the piano individually).

In spite of their tremendous expressive power, second-level languages have expressive limitations of their own. This led me to develop a generalization of the (improved) conception of model to study *third*-level languages, and, more generally, *n*th-level languages, for any finite *n*. In joint work with Øystein Linnebo, I was able to generalize the proposal further still, by applying it to languages of *infinite* order. The result is an infinite hierarchy of languages of ever-increasing expressive power.

An exciting consequence of our work is that it reveals a robust analogy between the expressive limitations of the standard language of set-theory and those of the 'super language' that results from combining languages from arbitrary levels of the hierarchy. This suggests that certain kinds of expressive limitations are a fundamental feature of language, rather than an artifact of particular systems of representation.