

KRIPKE

draft 2 April 2008

1 Introduction

Saul Kripke, born in 1940, was a child prodigy. He published his first paper, on the semantics of modal logic, at 16, and went on to study mathematics at Harvard. Much of his work since then has also been of a technical, mathematical nature. But Kripke has matched this work in formal logic with a philosophical oeuvre that is both accessible and widely influential. The latter has played a central role in rehabilitating metaphysics in the latter part of the twentieth century and setting the agenda for philosophical logic since. Kripke was a professor at Princeton University, having previously taught at Rockefeller University and Cornell; he is now at the City University of New York Graduate Center.

The historical significance of Kripke's work must be understood against a particular background. Twentieth century 'analytic' philosophy was dominated by a combination of empiricism, logical positivism in particular, and the linguistic turn in philosophy, especially the work of Ludwig Wittgenstein. To appreciate Kripke's importance not all the details of these lines of thought need to be understood, although some will be explained below. For now, I shall mention two important features of the prevailing philosophical climate. First of all, *metaphysics* was rejected as an important field of philosophical knowledge. Since Plato and Aristotle (and before) philosophers had been asking questions about the essential natures of things, and about the every general kinds of thing there are, and about which facts are *necessary* (must be so) and which are *contingent* (could have been otherwise). The growth of empiricism in the eighteenth century was antithetical to metaphysics, since empiricism's emphasis on all genuine knowledge coming from the senses leaves little room for substantive philosophical knowledge in the form of metaphysics. In the nineteenth century positivism actively rejected metaphysics; positivists regarded metaphysics as an intermediate stage of intellectual development coming after theology and before finally reaching a scientific, purely empirical mode of enquiry (i.e. theorizing about the world purely on the basis of experiment and observation). As a component of logical positivism, this rejection of metaphysics predominated for most of the half-century following the First World War.

Logical positivism was also deeply influenced by what has been called the 'linguistic turn' in philosophy. Very roughly, the linguistic turn can be characterized as the widespread conviction that the problems of philosophy must be addressed through the analysis of language. Many such problems were held to arise from the vagaries of everyday natural language. Expressions of a natural language may be ambiguous.

Or two sentences may seem to have the same grammatical structure yet have different logical implications: for example, ‘Bill doesn’t swim’ and ‘Father Christmas doesn’t exist’ are grammatically alike, but the truth of the first implies that Bill exists, whereas the truth of the second implies that Father Christmas does *not* exist. Following Gottlob Frege and Bertrand Russell, the logical positivists held that the replacement, for scientific and philosophical purposes at least, of natural languages by an artificial, formal language would lead to the elimination of many of the traditional problems of philosophy. While Wittgenstein, in his later work especially, was far from being a positivist, he too held that the traditional propositions of metaphysics were either the result of grammatical confusion or, in the case of acceptable assertions, merely reflections of rules of grammar.

2 Necessity and Essence

As mentioned, a central concern of traditional metaphysics has been with *modality*. Modality deals with *necessity*—what must be the case, *possibility*—what could be the case, and *contingency*—what could be the case under some circumstances but could also fail to be the case under other circumstances. Thus medieval philosophers asked whether God is a necessary being and whether God possesses certain properties, such as benevolence, necessarily: must God exist (or might He not have existed) and must God desire what is good (or could He desire what is bad)? Furthermore, the claims of metaphysics (and indeed other branches of philosophy) are themselves held to be necessary. For example, not only is it true that identical entities possess all their properties in common, but this is also necessarily the case. The deflation of metaphysics prevalent in 20th Century philosophy entailed a corresponding deflation of the concept of necessity. The necessary propositions, it was held, are just those that are knowable *a priori*. A proposition is known *a priori* if it is known by pure reflection alone and without recourse to experience of the world, e.g. observation and experiment. The known propositions of logic and mathematics, for example, are known *a priori*. So what had been a distinctive metaphysical category was assimilated to an epistemic category, a category concerning knowledge. Furthermore, the *a priori* knowability of such propositions was explained by their being *analytic*—propositions that are true in virtue of their meaning. A standard example of an analytic proposition is “all vixens are foxes”. Since ‘vixen’ just means ‘female fox’, that proposition asserts that all female foxes are foxes, which is a logical truth. And so a common line of thought asserted that the necessity of *any* necessary proposition (e.g. “the internal angles of an Euclidean triangle sum to two right angles”) is just a matter of its being possible to know that proposition by pure reflection. And the latter is possible because once the meanings of its terms were understood, it could be seen to be equivalent to a logical truth. (It was often added that the basic logical truths provide implicit definitions of key logical terms such as ‘all’.)

Kripke’s work played an important part in the rehabilitation of modal metaphysics. In his formal work Kripke developed an elegant model or semantics for modality based on Leibniz’s idea of a *possible world*—a way the world might have been. In the simplest case, what is possible is true in some possible world, what is necessary is true in

all possible worlds. A proposition is true but contingently so, if it is true in the actual world but false in some other possible world (i.e. a proposition that is one that is actually true but possibly false). In a series of lectures, published as the book *Naming and Necessity*, Kripke makes a corresponding philosophical case for rejecting several central tenets of the anti-metaphysical, linguistically inclined philosophy of the twentieth century. Foremost among the rejected claims was the identification of necessity with *a priori* knowability (and with analyticity). Kripke argued that there are necessary propositions that are not knowable *a priori* but which are known only *a posteriori*, which is to say that they are known only with the aid of some experience of the world, for example they are known only thanks to observational evidence. Conversely there are some propositions whose truth we can know *a priori* but which are not necessary but contingent, i.e. which could have been false.

To persuade us that a proposition can be necessarily true but not known *a priori*, Kripke asks us to consider Goldbach's conjecture that every even number can be expressed as the sum of two prime numbers. This conjecture has never been proven, but neither has anyone come up with a counterexample, an even number that is not the sum of two primes, or any other kind of disproof. But along with all mathematical propositions, Goldbach's conjecture is, if true, necessarily true. Likewise, if it is false, its negation is necessary. So either Goldbach's conjecture is a necessary truth, or its negation is. But neither is in fact known *a priori*. Perhaps there is a proof of the conjecture (or its negation) which would allow us to know it *a priori*, just as a proof was eventually found for Fermat's long-conjectured 'last theorem'. But there is no guarantee that every true mathematical proposition has a proof. Kripke points out that Gödel's first incompleteness theorem shows that there are infinitely many arithmetical truths that cannot be proved from the favoured Peano axioms for arithmetic. This makes it extremely implausible that every arithmetical truth can be proved from a set of axioms themselves knowable *a priori*. If that is right, infinitely many arithmetical propositions are necessary, but not knowable *a priori*.

Kripke provides other examples of propositions that are necessary but are known only thanks to worldly experience, and so are *a posteriori* and not *a priori*. He uses the example introduced by Gottlob Frege, "Hesperus is Phosphorus". The early Greeks used 'Hesperus' to name the planet that is also known as the evening star, that is the first planet to appear on the horizon at dusk; likewise 'Phosphorus' named the morning star, the last planet to remain visible on the horizon at dawn. The Greeks were eventually persuaded by the Babylonians that these were not two planets but one and the same planet; they thus came to know the truth of "Hesperus is Phosphorus". This knowledge is clearly *a posteriori*, not *a priori*—it had to be learned with the aid of the careful astronomical observations of the Babylonians, plus a bit of astronomical theorizing. However, argues Kripke, since Hesperus just is Phosphorus, this proposition that Hesperus is Phosphorus simply asserts of something that it is identical to itself and so is a necessary truth. Many similar examples arise with people who have pseudonyms, such as authors writing under a *nom de plume*. Thus the propositions "George Eliot is Mary Ann Evans" and "Mark Twain is Samuel Clemens" cannot be known *a priori*, but are necessary, since in both cases they assert the identity of a person with themselves.

Further examples of *a posteriori* necessities proposed by Kripke have proven to be more contentious. These concern the constitution, natures and origin of kinds of thing

and of individual things. Kripke starts from a discussion by Timothy Sprigge, in which he asks (in effect) whether the Queen, Elizabeth II, could have had different (biological) parents. Kripke acknowledges that it could be discovered that here parents are not the people we thought them to be (George VI and Elizabeth Bowes-Lyon). This shows that the proposition that Elizabeth II is the child of George VI and Elizabeth Bowes-Lyon is *a posteriori* and not something we can know *a priori*. But, asks Kripke, given that Elizabeth II is indeed the child of George VI and Elizabeth Bowes-Lyon could she in some other possible world have different parents, e.g. Mr and Mrs Truman? No, says Kripke, they could have had a child with many of the same qualitative properties as Elizabeth II. But that child would not be *this* woman, Elizabeth II. It is impossible for a person to have originated from a different sperm and a different egg from the ones she did originate from.

In a similar vein, Kripke asks of a particular wooden table, could this very table have been made from a different hunk of wood, or even from water from the Thames frozen and made to look like wood. We might have to look carefully to check whether those things are true, which shows the relevant propositions to be *a posteriori*. Even so, given that the table *is* in fact made from some specific hunk of wood, it is impossible, says Kripke, that it could have been made from a different hunk of wood or from ice. A similar looking table may be so constructed, but it would not be *this* very table.

These two examples suggest the principle that if a material object has its origin in a particular hunk of matter, it could not have had its origin in any other hunk of matter. If this principle is true, then we have another general source of propositions that are necessary but which can be known only *a posteriori*. An *a posteriori* investigation such as a paternity test, may be needed for us to come to know that Fred is Bill's father, but if Fred is indeed Bill's father, then Bill could not have had any other father.

Kripke then proceeds to discuss theoretical identities and the properties of substances and natural kinds of thing. Theoretical identities are cases where science tells us what something *is*. The examples Kripke uses are: light is a stream of photons; water is H₂O; lightning is an electrical discharge; gold is the element with atomic number 79.

Starting with the example of gold, Kripke asks us to consider an alternative view, that gold is defined in terms of its apparent observable properties, so that it is *a priori* that gold is a yellow metal (a view he finds in Kant). Kripke asks us to imagine that the yellowness of samples of gold is just an optical illusion. Would that mean that there is no gold? No, it just means that we were wrong all along in thinking that gold is yellow. Similarly, Kripke asks whether "A tiger is a large carnivorous quadrupedal feline, tawny yellow in colour with blackish transverse stripes and white belly" could provide the meaning of 'tiger'. No; once again it might be that all tigers are three legged but that an optical illusion makes us think otherwise. At the same time, would anything that does satisfy this description thereby necessarily be a tiger? Let us suppose that we discover some hitherto hidden population of animals meeting this description but which, on investigation, turn out to be reptiles. Would they also be tigers? No, they would not be tigers, despite having the superficial appearance of tigers. (Kripke takes 'feline' to mean 'having a cat-like appearance'—if it meant 'belonging to the cat family' then 'tigers are feline' would clearly be an *a posteriori* discovery and cannot be part of an *a priori* definition.) Returning to gold, would some substance that had the

superficial yellow, shiny appearance of gold be gold? Not necessarily—it would not be gold if its atomic structure were different from that of gold. Kripke discusses the other theoretical identities in a similar way. The natures of the kinds in question are not given by their superficial or observable qualities. A substance that had a completely different atomic structure from H_2O , however much it resembled water in its appearance, taste and so forth, would not be water—it would be ‘fool’s water’ (by analogy with ‘fool’s gold’—iron pyrites, which can look like gold in its natural state). While light is known to us primarily through its effects on our vision, that is not the nature of light. A world where all humans are blind is not a world without light.

If the natures of tigers, water, gold, light, and so forth are not given by their superficial, observable characteristics, what does determine their natures? Kripke makes it clear that, in most cases at least, it is certain key *microstructural* properties that determine the nature of these natural kinds. This is explicit in his endorsement of the identities ‘water is H_2O ’, ‘gold is the element with atomic number 79’, ‘light is a stream of photons’, where the identity is between a kind in a familiar, everyday guise (water, gold, light) and a kind characterized by its inner or micro-structure (H_2O , element with atomic number 79, stream of photons).

In such cases, the microstructure gives the *essence* of the kind. Kripke uses ‘essence’ to mean a property that something has necessarily. That is necessarily, water is H_2O , since, as we discussed, a substance that is not H_2O would not be water, however much it superficially seemed like water. Likewise a compound or an element, but with an atomic number other than 79, would not be gold, even if it had a convincing appearance of gold. These kinds have a microstructural essence.

It is not only kinds that have essences, nor are all essences microstructural. We saw above that Kripke argues that Queen Elizabeth II could not have had parents other than George VI and Elizabeth Bowes-Lyon. Thus having those parents is an essential property of the Queen. In this case it is an individual who has an essence, and her essence concerns her origin, not her microstructure. Similarly, we saw Kripke argue that necessarily a certain table could not have originated from a different hunk of matter; a table made from a different hunk might be qualitatively identical but it would not be *that* table. And so such objects, Kripke thinks, as well as other material objects, all have their original matter essentially.

Kripke does not explicitly say that biological kinds have microstructural essences. He does say that we could determine that a tiger-like reptile is not a tiger by examining its internal structure, and likewise for a cat-like automaton, but that does not imply that true tigers and true cats have microstructural essences. Philosophers of biology, those friendly to essences at least, have subsequently argued that the essences of species and other biological kinds (genera, families etc.) are historical. This is most obvious in the case of *cladistics*, where a taxonomic group is considered to consist of an ancestral population and all of its descendants. Such approaches regard origin as essential for biological kinds, just as Kripke does for people and for material objects.

Interestingly, while Kripke takes these propositions asserting identities to be true, he takes correspondences between sensations and states of the brain to be contingent, and so not identities at all. For example, let the sensation of pain be correlated with the brain state that is the firing of C-fibres. Could the pain sensation *be* the firing of the C-fibres? No, says Kripke, because it is possible to have the firing of the C-fibres

without there being any pain: there are possible worlds where brains have the firing of C-fibres but the persons with those brains feel no pain. So the relationship between pain and C-fibres firing is contingent and thus not an identity. Why cannot we treat this like this case of light, which could exist, even though it brought about no visual sensations? Kripke's reason is that our intuitions tell us that in such a case, the stream of photons would still be light. On the other hand, something that was not painful, that was not felt as pain, would not *be* a pain. In the case of light the relationship with the corresponding microstructure (stream of photons) is necessary whereas its relationship with the sensations it causes (warmth) is contingent. For pain things are the other way around. The relationship between pain and the sensation of pain is necessary while the relationship with the corresponding microstructure is contingent.

3 Naming and Reference

Kripke's metaphysical views were a radical departure from the orthodoxy that prevailed during the central part of the twentieth century. Although they are still the subject of much debate, those views, and in particular the claim that identity statements such as 'Hesperus is Phosphorus' are necessary, seem to many today to be just plain obvious. 'Hesperus is Phosphorus' just states of one thing that it is itself: surely that is necessary if anything is. So why should Kripke's claim be so revolutionary?

The reason is to be found in a philosophy of language, and in particular a theory of *reference*, that was taken for granted and which leads to a contrary view about the modal status (i.e. status as necessary or contingent) of identity statements. Kripke, of course, held that this philosophy of language was mistaken and much of *Naming and Necessity* is devoted to correcting the picture of reference that is part of it.

Not all identity statements, even on Kripke's metaphysics, are necessary. For example, 'Benjamin Franklin is the inventor of bifocals' is true but contingent. That is because this statement contains a *definite description* 'the inventor of bifocals'. A definite description is not a genuine name. In modal terms, the expression 'the inventor of bifocals' denotes different entities in different possible worlds. In this world, it denotes Benjamin Franklin; in many other worlds it denotes some other person (e.g. it might denote Spinoza, if Spinoza had been the first person to make bifocals), or it might denote no-one at all if bifocals were never invented or invented simultaneously by two or more people. In some of the other possible worlds Franklin still exists. So we may say that an identity statement containing a definite description such as this will be contingent, not necessary.

A definite description denotes some object in virtue of that object satisfying the description uniquely. 'The *F*' denotes *x* precisely when *x* is *F* and is the only *F*. As remarked, different objects may be the unique *F* in different worlds. Now let us return to names. How do they denote what they do? This where the dispute between Kripke and his predecessors lies. We will come to Kripke's answer in due course. For now we note that his view is that however names do get their meaning, a genuine name denotes one and the same object at all possible worlds where that object exists. That is, with respect to any world where the man Benjamin Franklin exists 'Benjamin Franklin' denotes precisely that man. Terms that do this, denoting the same entity in all possible worlds (if

any entity is denoted) are called *rigid designators*. Since ‘Benjamin Franklin’ is a rigid designator and ‘the inventor of bifocals’ is not a rigid designator, we can see that in some worlds ‘Benjamin Franklin is the inventor of bifocals’ will be false. On the other hand ‘Hesperus is Phosphorus’ is a statement of identity with two names, and so, on Kripke’s view, with two rigid designators. Since ‘Hesperus’ and ‘Phosphorus’ are rigid designators, if they denote the same object in any world, they designate the same object in every world (if they designate at all). Hence ‘Hesperus is Phosphorus’ is necessarily true or necessarily false.

But there is an alternative view about names, which takes names themselves to be disguised definite descriptions. Imagine that this were correct. Then Kripke’s claims would be false. Definite descriptions are not rigid designators and so names would also not be rigid designators. Identity statements involving two names would thus not, in general, be necessary but would typically be contingent, because they are really statements employing two definite descriptions. We saw that Russell had argued that definite descriptions have a particular logical analysis; in turn he held that names as we find them in everyday language (‘Benjamin Franklin’, ‘Mark Twain’, ‘Hesperus’) are not genuine names (‘logically proper names’) but are equivalent to definite descriptions. This view may also be attributed to another great logician, Gottlob Frege, for whom the denotation of a name is given by its associated *sense* which we may take to be something like a definite description. A more sophisticated view, found in the work of Ludwig Wittgenstein and laid out in detail by John Searle, proposes that a name need not be equivalent to a single definite description but may be understood in terms of a cluster of properties. Let ‘N’ be the name of a person; then Russell’s view would be expressed by ‘N \equiv the (unique) person who is X’. The cluster view suggests instead: ‘N \equiv the (unique) person who best satisfies most of the properties X, Y, . . . , and Z’. For example, a Russellian view might analyze ‘Moses’ as ‘the (unique) person who led the Israelites out of captivity in Egypt’, whereas the cluster view would analyze Moses as ‘the (unique) person who did most of the following things: led the Israelites out of captivity in Egypt; spoke to the burning bush; wrote down the Ten Commandments; . . .’.

Kripke advances a series of arguments against both the simple and cluster description views. We will concentrate mainly on the former, since his arguments against the cluster view are just elaborations of the corresponding arguments against the simple view.

The first set of arguments against the description view of names are modal in nature. The description view renders ‘Hesperus is Phosphorus’ contingent, which is an advantage, according to the traditional view that says that propositions that are known *a posteriori* are contingent. But, Kripke points out, the view has a very severe *disadvantage* in terms of modality. Let us imagine that the correct description associated with Moses is ‘the (unique) person who led the Israelites out of captivity in Egypt’. If so, the proposition ‘Moses is the (unique) person who led the Israelites out of captivity in Egypt’ is (a) necessary, (b) analytic, and (c) *a priori*. But clearly that proposition is none of those. Had Moses decided otherwise, he could have led a quiet and pleasant life in the courts of Egypt; it is a *contingent* fact that he embarked on leading the Israelites out of Egypt. Likewise, prominent candidates for the definite description equivalent to ‘Aristotle’ might be ‘the last great philosopher of antiquity’ or the ‘teacher of Alexan-

der'. If one or other *were* the correct description that is equivalent to 'Aristotle', then it would be necessary that Aristotle was a philosopher or teacher. And that is not correct, for Aristotle might not have taken up a philosophical or paedagogic career (Aristotle would not have taught philosophy if he had not been sent to Athens by his guardian or if he had been killed when the Persians overran his home in Mysia).

The cluster view might be thought to be able to resist such criticisms, since on that view it is not required that the person denoted by a name satisfy any *specific* member of the cluster, so long as enough of the remaining descriptions are satisfied. Thus it is contingent that Moses spoke to a burning bush. 'Moses' can still denote a man who never spoke to a burning bush, so long as that man did all or almost all of the other things associated with that name in the Bible. Nonetheless, it remains the case that according to the cluster view it is necessary that Moses did most of those things. And that also is false. Moses could have died in childhood (the baby Moses could have fallen out of the basket placed in the reeds), in which case all or almost all of the descriptions associated with Moses from the Bible would be false.

Kripke's modal criticism of the description view depends to some extent on his choice of putative descriptions to associate with the names 'Aristotle' and 'Moses'. Would a different choice of descriptions have saved the description view from Kripke's arguments? Kripke does not deny that some descriptions may hold necessarily of a person, and indeed the essentialism we discussed above shows that he thinks that there are many necessary truths about individuals, and these may even suffice to identify them uniquely. Kripke takes it as part of the description view that the descriptions in question must be *analytically* equivalent to the associated definite description. That is to say, the definite description gives the *meaning* of the name. That being so, the relevant descriptions must be ones that a user of the name must, in principle at least, be able to associate with that name. Since pretty well all that we know about Moses comes from the Bible, any description equivalent to 'Moses' must be one of those found in the Bible, or some combination of them. Likewise it is appropriate that Kripke choose as candidate descriptions to be equivalent to 'Aristotle' those facts about him that are best known.

We can easily see that this aspect of the description view also rapidly leads to trouble. For we often use a name and succeed in referring to the bearer of the name even though we do not have sufficient knowledge of the bearer to identify them uniquely. Kripke uses the example of the name of the physicist Richard Feynman. Kripke asserts, almost certainly correctly, that many people can use 'Feynman' to refer to Feynman, even though they may know nothing more about Feynman than the fact that he was a famous physicist. But that fact does not stop them referring to Feynman, and not, say, to Murray Gell-Mann, another famous physicist. (Imagine the following conversation. A: "Can you name any famous physicists?" B: "Yes, Feynman, he is a physicist." A: "Do you know anything else about Feynman, for example, what his theories are about?" B: "No, I only know that he is a physicist." B was indeed referring to Feynman, though B may also know of Gell-Mann that he too is a physicist, and only that)

Those problems arise because we may know too few facts about someone to be able to associate a unique description to their name. Related problems will arise when what we do believe about someone is false. Many of us believe that Kurt Gödel was the first man to prove the incompleteness of arithmetic, and if that is all we believe about him,

that should be the description that picks out the referent of 'Gödel'. Kripke now asks us to imagine that in fact some obscure Viennese called Schmidt, whose body was found in mysterious circumstances, was the first to prove the incompleteness of arithmetic and that Gödel somehow got hold of his manuscript. Now let us ask, to whom does the name 'Gödel' refer? According to the description theory, that name refers to the unique person who first proved the incompleteness of arithmetic. And that person is Schmidt. So, for many people, they will not be referring to the person who in fact was baptized 'Kurt Friedrich Gödel' by his parents, who stole Schmidt's manuscript, who became famous for publishing it, and who became a friend of Einstein's at Princeton; rather they will be referring to the obscure and unfortunate Schmidt. And clearly that is wrong. If the story about Schmidt's manuscript were true, it is natural and correct to think that people who believe of Gödel only that he was the first man to prove the incompleteness of arithmetic, have a false belief about the Princeton logician, not a true belief about Schmidt.

It looks then as if the definite description account of names has many and severe disadvantages. But the description view has at least this advantage: if names are just equivalent to definite descriptions, then we have a clear account of how the denotations of names are fixed. What is Kripke's alternative view of the denotation of names? Kripke does not give a fully worked-out theory of naming, but he does present a more general picture of how naming works. Here is the simplest case. A name is given to a person or object in a naming ceremony, such as a baptism, for example a baby is given the name 'Ichabod'. Jane was present at the baptism, which enables her to use the name to refer to that baby. She then talks about the baby to John who was not at the baptism. John now uses the name 'Ichabod'; he intends to use it to refer to the baby to whom Jane was referring. So John's use of 'Ichabod' is not dependent on his possessing some description of the baby in mind, but rather on the fact that he acquired the use of that name from someone who was present at the baptism. John can likewise pass on the use of the name to someone else, maintaining the same intended reference. Thus when someone says 'Napoleon was the Emperor of the French' their use of 'Napoleon' refers to Napoleon because there is a chain of reference preserving links going back to the baptism of Napoleon by his parents. This chain is a causal chain, since later use of the name is causally dependent on earlier use. Consequently the simple theory is often known as the 'causal theory of reference'. In the case of 'Napoleon' the causal chain takes us back to Napoleon himself. But the chain of reference preservation need not do that in every case (it is the chain of preservation between speakers that is causal, not the link to the object named). Consider the astronomer Leverrier who, along with Adams, discovered the existence of the planet Neptune by analyzing the anomalous orbit of Uranus. Without observing the planet Leverrier was thus in a position to name it 'Neptune'. In this case, the name is attached to the planet *via* the description 'the planet causing the anomalies in the orbit of Uranus', but the name is not equivalent to the description. Rather the description takes the place of the pointing that Napoleon's parents were able to use to pick out their son as the intended recipient of a new name. In the Neptune case there is still an indirect causal link between Neptune and the reference-preserving chain of use, insofar as Neptune caused the anomalies in the orbit of Uranus, which caused Leverrier to work out his theory and ultimately to give the planet its name (in actual fact Leverrier proposed this name somewhat later).

But in other cases a definite description can introduce the name of an abstract object, e.g. ‘the ratio of the circumference of a circle to its diameter’ for π , in which case there is no causal link to the object named.

As we saw above, Kripke extended his metaphysical views concerning identity and essence from individual to natural kinds and quantities. Likewise he extended his account of naming and reference to terms referring to such things. Thus terms such as ‘tiger’, ‘water’, and ‘gold’ function like names, naming kinds of thing or substance rather than individuals. These ‘common names’, like names of individuals, are rigid designators. In many cases their references are fixed by something akin to an initial baptism with samples: ‘Gold is the substance instantiated by the items over there’. In other cases (heat, electricity) the reference is fixed via certain experiences or experimental effects. The kind (species, substance, quantity) terms may be associated with certain characteristic properties that are commonly used to identify members of the kind (e.g. yellow, shiny, malleable for gold). But those properties do not provide an analytic definition of the kind term, for it could be that genuine members of the kind lack those properties. Typically it is scientific investigation that reveals the properties that do fix kind membership, the properties that characterize the nature or essence of the kind.

4 Rules and Meaning

The Cambridge philosopher Ludwig Wittgenstein put forward a cluster view of the meaning of names in his posthumous book *Philosophical Investigations*. Furthermore, Wittgenstein’s work in that book and in his later oeuvre of which it is a part articulates a strongly anti-metaphysical philosophy. The work of Kripke discussed above is thus a reaction against a set of philosophical pictures and predilections to which Wittgenstein is a major contributor. It comes therefore as something of a surprise to find Kripke articulating and building on a sceptical paradox and a ‘sceptical solution’ to it that he finds in Wittgenstein’s *Philosophical Investigations* and which Kripke takes “to be the central thread of Wittgenstein’s later work on the philosophy of language and the philosophy of mathematics”. According to Kripke, the paradox is central to Wittgenstein’s ‘private language argument’, which, in his view, is to be explicated in terms of the problem of ‘following a rule’. While there is considerable debate as to whether Kripke’s interpretation is at all true to Wittgenstein’s intentions, for us the important thing is to understand Kripke’s discussion in its own right.

Kripke asks us to imagine the following scenario. I am competent with ordinary arithmetic and I am asked to perform a straightforward addition sum. It is a sum that I have not computed before, and involves numbers larger than those in any calculation I have previously performed. Nonetheless, it is still well within my mathematical abilities. Let us say, for sake of argument, that the sum I have been asked to perform is ‘68 + 57’. I am of course confident that the answer is ‘125’. However a bizarre sceptic raises the following possibility. Consider, he says, the following function, which we will call ‘quus’ and which we symbolize by ‘ \oplus ’:

$$\begin{aligned} x \oplus y &= x + y, \text{ if } x, y < 57 \\ &= 5 \quad \text{otherwise} \end{aligned}$$

Perhaps, he suggests, in the past, when I used the word ‘plus’ and the symbol ‘+’ I did not mean the function plus (i.e. addition, the function yielding ‘125’ as the correct answer) but rather I meant the function quus (‘quaddition’). In which case, in giving the answer ‘125’ I am mistaken—I should have answered ‘5’. The mistake, Kripke emphasizes, would not be in arithmetic but is instead a mistake about my previous intentions and meanings. And of course, one can sometimes be mistaken about one’s previous intentions (failure of memory, influence of drugs, etc.).

Although the possibility of a mistake has been raised, I am of course sure that there is no mistake. Since error and correctness make sense here, there must be, so it would seem, some fact of the matter about which I am correct, a fact of the matter that makes it the case that the sceptical suggestion is not only bizarre but indeed false. For example, imagine that I had performed this particular calculation many times before. The fact that I have on previous occasions given the answer ‘125’ to ‘68 + 57’ would show that I did not then intend quus by ‘+’. However, we set the example up on the assumption that I had not performed this particular calculation before, and in any case there will always be some calculation that is a little bigger than any other I have performed before, and for *that* calculation we can construct a quus-like function concerning which we can pose the sceptical question. So my previous behaviour—what I have said and done in the past—will be enough to rule out many sceptical hypotheses about my intentions. But that behaviour will also be consistent with many other (indeed infinitely many other) sceptical hypotheses.

So one possible proposal for facts that determine that I meant plus not quus—my past behaviour—is inadequate. The reason is that my past behaviour has nothing to say about cases that differ from those I have come across before. Surely, one might think, *something* links my past behaviour and my current answer and explains both. This would be my tendency or *disposition* to behave in one way rather than another. The answers I have given in the past and the one I give now are all manifestations of the same disposition, and it is this disposition that determines that I mean plus not quus and thus that ‘125’ is the correct answer. Dispositions look like a good answer to the sceptical problem because dispositions can concern possible circumstances that have not yet arisen or indeed may never arise. Some vases (such as the Portland vase) are struck and their fragility is shown by their breaking; but an equally fragile, identical vase may be sufficiently well protected that it never exhibits its fragility; such a vase is still fragile. Here is another example: the elasticity of a rubber band will determine how much it stretches in response to varying degrees of force. It may never have been subjected to a force of 3.6 N, but, its disposition, the degree of elasticity of the band determines that it *would* stretch by 4 cm (and *not* by 5 cm). Likewise, my disposition might determine that in the past I *would* have answered ‘125’ had I been given the question ‘68 + 57’. So that fact determines that I didn’t mean quus by ‘+’, and more generally, my underlying disposition fixes it that I did mean plus.

Kripke has two principal objections to this dispositional answer to the sceptical challenge. The first is that even if my dispositions extend beyond my actual behaviour they are still not infinite in the way required to fix it that I mean plus rather than some quaddition-like function that starts to diverge from addition for very large number. For addition specifies a specific answer for *any* pair of numbers, however large. But some such numbers are so large that I cannot even *think* them, let alone attempt to add them

together. So I do not have a disposition to give the sum, the output of the plus function, in answer to an addition question concerning such large numbers. Such cases show that my disposition does not fix on precisely addition rather than some other function.

Kripke's second objection focuses on the fact that meaning is a *normative* relation while my dispositions are purely causal and are not normative. That is to say, what I mean fixes what I *should* say, not what I do say nor what I *would* say. But as we all know, what we ought to do and what we are disposed to do can be quite different things. If I intend to play chess and so to play by the rules of that game, I *ought* to move my king only two squares when castling. If I inadvertently move the king three squares when castling on the queen's side, then that shows that I failed to do what I intended to do. It does not show that I did instead intend to play some other game that is a variant on chess. But what was I disposed to do when I moved my king three squares? It looks as if I was disposed to move the king three squares—after all that is what I did do, and the reasons for so doing all came from within me (it is not as if something else forced my hand to move the king three squares). So my disposition, argues Kripke, differs from my intention: I intended to castle according to the rules, but I was disposed to do something different. If we insist that my disposition does show what I intended, then we must conclude that after all I did intend to play by a rule other than the standard rule for castling. Either way, it cannot be true *both* that the rule I intended is the same as the pattern corresponding to my disposition *and* that I can err, failing to follow the rule I intend to follow.

So how should we respond to the rule-following paradox? Kripke distinguishes *straight* solutions from *sceptical* solutions. A straight solution would aim to show that there really is no paradox and that the paradoxer has made some error in arguing that there is a paradox. The dispositional solution is an attempted straight solution. If it were correct then it would show that there is a definitive answer ('my dispositions') to the question: what shows that I was following the plus rule rather than the quus rule? A sceptical solution, on the other hand, does not attempt to get rid of the paradox but instead seeks to explain why it does not have quite the negative impact one might at first suppose it to have. In this case Kripke's Wittgenstein proposes just such a sceptical solution. The conclusion of the paradox must be accepted. No fact about me makes it the case that I meant plus not quus.

If there is no fact that makes it the case that someone means plus, then surely it would never be reasonable to say the perfectly ordinary things we do say, such as "Jones means addition by '+'". This is where the sceptical solution comes in. According to Kripke, Wittgenstein denies this inference. Wittgenstein was concerned to reject a conception of meaning that dominates his own early *Tractatus Logico-Philosophicus*. According to this conception, sentences get their meanings by picturing or corresponding to some fact in the world. This is known as the 'picture theory of meaning' and exemplifies a more general philosophy of meaning according to which the meaning of a sentence is given by specifying the conditions under which the sentence is true. Applying this 'truth-conditional' approach to the case in hand, the meaning of "Jones means addition by '+'" is to be identified with some possible fact that would make this sentence true. But the rule-following paradox tells us that there is no such fact. Hence "Jones means addition by '+'" has no meaning. But note that the rule-following paradox leads to the conclusion that assertions concerning meaning are themselves

meaningless only because it has been combined with the truth-conditional approach to meaning. If we drop our adherence to the latter, then we are no longer committed to denying the meaningfulness of statements about meaning.

If we reject the truth-conditional approach to meaning, what replaces it? The answer is *assertion* or *justification* conditions. That is to say, the meaning the sentence “Jones means addition by ‘+’” is given by the conditions under which an assertion of that sentence is justified. What are those conditions? According to Kripke’s understanding of Wittgenstein, they are the conditions under which the other members of my language community would concur with my assertion. I can justifiably assert a sentence when the other members of my community would agree with my assertion.

It is important to see that Kripke is *not* saying that a sentence is *true* just when my community would say that it is true. That would be to give another kind of ‘straight’ answer to the paradox. Earlier we looked at the view that what makes it the case that I mean addition by ‘+’ is my dispositions. The view now being considered, that what makes my assertion true is what my community would say, holds in effect that that what makes this the case is not *my* disposition but is rather *my community’s* disposition. But as a straight answer, the community disposition answer suffers from entirely analogous problems to the individual disposition answer. Just as I don’t have the appropriate disposition when it comes to enormously large numbers, neither does my community. Nor does my community’s being disposed to say that something is right make it right. Even communities can err, in which case there can be a mismatch between what rule a community intends to follow and what it is disposed to do. So the Kripke–Wittgenstein community response is importantly different from the community disposition answer, and that is because the community’s disposition makes my assertion justified but does not make it true. And what the assertion means is given by the former, not the latter.

When we say that someone means something by the use of a word, it is very tempting to see this as asserting something about a hidden inner mental life. And so one might think that when someone uses the word ‘yellow’ to describe a marigold that is because some inner process is going on: they recognize their visual experience as being of the kind that that have decided to call ‘yellow’; likewise ‘pain’ names a different kind of experience, and what I mean by ‘pain’ is given by my associating it with that experience, as if, at some young age I connected that word with an inner experience of that sort, my first toothache for example. If my words were to get their meaning in that way, that would be what Wittgenstein calls a ‘private language’. Wittgenstein regarded the possibility of a private language as an implicit but central part of the conception of the mind we have been working with ever since René Descartes in the seventeenth century. Wittgenstein’s ‘private language argument’ was intended to undermine that conception of the mind by showing that a private language is impossible. According to Kripke, the rule-following paradox is the key component of Wittgenstein’s private language argument. Let us say that I do attempt define a symbol ‘S’ by reference to some inner experience or sensation; I am in effect giving myself a rule: the rule that some subsequent experience is to be called ‘S’ if and only if it is like the original defining experience. It *seems* as if my act of inner definition is sufficient to fix that rule as the rule I intend and so fix the meaning of ‘S’. But if the rule-following paradox is right, then nothing *I* can do will in fact achieve that.

Since the paradox and its sceptical solution are presented by Kripke as “Wittgenstein’s argument as it struck Kripke, as it presented a problem for him” it is difficult to know quite which philosophical views to attribute to Kripke on the basis of this book. Presumably Kripke does think that the paradox is a genuine and interesting philosophical problem. And one might infer that Kripke does not think that there is any obvious straight solution—if there were, he would surely have mentioned it; and he would not have found the paradox so interesting in the first place. One might conjecture that Kripke at least thinks that the sceptical solution is worth considering, even if it is not a view we can attribute to him. On the other hand, clearly much of it takes us in a direction that is quite different from Kripke’s work on modality and the philosophy of language. The latter is quite antithetical to Wittgenstein’s later philosophy. There is one point of contact however. We have seen that Kripke’s account of reference is an externalist account—the reference of a name I use is not fixed solely by facts about me, rather it is fixed by the initial baptism plus the chain of communication between that baptism and my current use of the name, facts about which I may be completely ignorant. The rule-following paradox endorses a similar conclusion about meaning in general. Facts about me do not fix the meaning of my words and symbols, including ‘+’. But the similarities stop here. First, there is a difference in the kinds of fact that *are* relevant to meaning. According to the causal theory of names, the relevant facts are not (just) those in my head but also include causal and historical environmental facts—facts about what entities and kinds have been in my environment and facts about the causes of my current use of a name. But according to the sceptical solution, the relevant facts are social facts. More importantly, the causal theory of names implicitly endorses a truth-conditional account of meaning. The meaning of ‘Mark Twain’ is the man Mark Twain because it is that man who make a contribution to the truth-conditions of sentences involving the name ‘Mark Twain’—for example, the meaning of ‘Mark Twain was a Mississippi riverboat pilot’ is given by conditions under which the sentence is true, viz. its being the case that *that* man, Mark Twain, was a Mississippi riverboat pilot. But as we have seen the sceptical solution to the rule-following paradox rejects the idea of truth-conditions altogether.

5 Conclusion

Kripke’s book on the rule-following paradox stands apart from the rest of his oeuvre. But the fact that this paradox has generated so much discussion is further testament to Kripke’s enormous fertility. We have not had the opportunity to look at Kripke’s formal work, on account of its technical nature. But it is worth concluding with a few remarks on *why* it is so important. Some formal logic is best conceived straightforwardly as a branch of mathematics, generating most of its problems internally. Other areas of formal logic bear a closer relationship to philosophy, and their problems originate with philosophical problems. Sometimes the problems of philosophy are sufficiently subtle or complex that the relatively informal ‘prose’ techniques of philosophy are insufficient to express or resolve them. In many such cases representing the problem using a formal, mathematical symbolism will clarify the problem and the subsequent application of formal or mathematical techniques will provided answers to the problems

thus clarified. Kripke's own formal work is a contribution to this kind of formal logic, and is exemplified by his semantics for modal logic and by his work on the theory of truth. Although formal logic was actively pursued throughout the twentieth century, the increasing, if as yet incomplete, readiness of philosophers to accept its relevance to 'ordinary' philosophical problems, has much to thank Kripke for, and represents a further departure from the linguistic turn and the view of Wittgenstein and of the so-called 'ordinary language' philosophers that a careful attention to the ways in which terms are used in everyday contexts will suffice to dissolve the problems of philosophy.

Metaphysics is one of the most active areas of philosophy today. But as a result of the linguistic turn it was moribund for several decades in the middle part of the twentieth century. This rehabilitation of metaphysics and in particular an interest in essentialism owes much to the work of Kripke. A key component of this achievement was his forceful arguments that the *a priori* and the necessary can be divorced. But this truth was obscured by a philosophy of language that took names to be equivalent to definite descriptions (or clusters of them). So a crucial step in the rehabilitation of essentialist metaphysics was the reformation of the theory of reference. Many philosophers have taken essentialism to be a direct consequence of Kripke's views about reference. This isn't quite correct and somewhat undervalues Kripke's achievement, by suggesting that the metaphysics and the theory of reference are sides of one and the same coin. Kripke's account of reference *permits* essentialism, but does not require it. So his rehabilitation of metaphysics and his giving a new direction to the philosophy of language are two distinct if parallel contributions to philosophy whose significance can hardly be overstated.

Further reading

Fitch, G. W. 2004. *Saul Kripke*. Chesham: Acumen.

Hughes, C. 2004. *Kripke*. Oxford: Oxford University Press.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.

Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.