A NOTE ON KRIPKE'S PARADOX ABOUT TIME AND THOUGHT*

Kripke's paradox about time and thought is reminiscent of Russell's paradox about the set of sets that are not elements of themselves. At a particular instant t_0, for example, during the writing of his paper, Kripke entertains the following set concept, while not thinking of any other set:

> *the set of instants t such that: (i) Kripke is thinking at t of exactly one set of instants; and (ii) t itself is not an element of the time-set Kripke is thinking of at t.*

Evidently, Kripke thereby thinks at t_0 of the particular time-set of which this is a concept, and of no other set: the set of times at which Kripke thinks of exactly one time-set, which time-set excludes the very time of thinking. If the very instant t_0 itself is an element of the Kripke set, then it is not. On the other hand, evidently, if t_0 is not an element, then it is.[1]

The second argument place of thinking-of is to be regarded here as fully extensional/referential: Where α and β are singular terms, if $\ulcorner\alpha$ thinks of $\beta\urcorner$ is true, then: (*i*) β designates something; and (*ii*) if β and γ co-designate, then $\ulcorner\alpha$ thinks of $\gamma\urcorner$ is also true. If there is an alternative, intensional notion of thinking-of ("Kripke is thinking at t_0 of a pink elephant," "Kripke is thinking at t_0 of the largest prime integer," and so on), it is not the notion of thinking-of invoked in Kripke's paradox. Furthermore, unlike *de re* belief, the relevant notion of thinking-of is *latitudinarian* with respect to sets. If one entertains the concept *the set of F's*, one thereby thinks of the set of F's, even if one is unable to provide independent specification and one does not know of some particular objects whether they are F's. (I use italics as a means of indirect quotation, to form a designator of the nonitalic expression's semantic content.) The relevant notion of thinking-of for sets may be regarded as the relative product of the relation of "entertaining" between a thinker and a set concept and the determination relation between a set concept and the set of which it is a concept.

* I am grateful to David Kaplan and the Santa Barbarians for discussion, especially C. Anthony Anderson and Teresa Robertson.

[1] Saul Kripke, "A Puzzle about Time and Thought," in *Philosophical Troubles: Collected Papers, Volume 1* (New York: Oxford, 2011), chapter 13, pp. 373–79.

0022-362X/13/1004/213–20

Where α is a singular term for an instant of time, let $\ulcorner S_\alpha \urcorner$ be an abbreviation for $\ulcorner$the only set S of instants of time such that Kripke is thinking at time α of $S\urcorner$, or more succinctly, $\ulcorner$the time-set Kripke is thinking of at $\alpha\urcorner$. Then the Kripke set would be designated as

D: $\{t \mid S_t \text{ exists } \& \ t \notin S_t\}$.

If the axiom schema of Separation is extended from the language of pure set theory to that of $\ulcorner S_\alpha \urcorner$, then the restricting ("separating") condition on instants, S_t *exists* & $t \notin S_t$, yields that there exists such a set as $\{t \mid S_t \text{ exists } \& \ t \notin S_t\}$.[2] But then t_0 is both an element and a nonelement. That is logically impossible.

Kripke's paradox may be regarded on the model of the paradoxes about designation, like the Berry paradox. Berry's paradox is generated by the description 'the smallest natural number not designated by any English description of fewer than 15 words', which is itself an English description of fewer than 15 words. (We assume that, *pace* Russell, a definite description designates that which it correctly describes uniquely.) I have proposed an alternative paradox of designation through the following description

d: the number that is 1 if d designates 0 (in English), and is 0 otherwise.[3]

To obtain the paradox attempt to determine whether d designates 0.

Suppose that at t_0 Kripke utters exactly one description: '$\{t \mid$ I utter exactly one set-theoretic description at t & the set-theoretic description I utter at t designates exactly one time-set & $t \notin$ the time-set designated by the set-theoretic description I utter at $t\}$'. Thinking of something by entertaining a concept of it is analogous to designating something by using a definite description that expresses a concept of it. Both crucially involve concepts; both might call for ramification. Kripke does not offer an official solution to his paradox about time and thought. He suggests that it might be correctly solved by ramifying the concept of thinking-of, or by otherwise assimilating thinking-of to a semantic relation like designation. I here offer another perspective, which Kripke does not explicitly consider. This perspective is compatible with Kripke's suggestion but more revelatory.[4]

[2] The Separation schema states that for any set S there is a subset whose elements satisfy the restricting condition ϕ_x, where ϕ_x is any condition expressible in the language of pure set theory (whose only nonlogical constant is '$\in$').

[3] The appellation 'd' within d can be replaced by a description like 'the term written on the blackboard in Nathan Salmon's campus office'.

[4] Kripke makes brief remarks in defense of extending the Separation schema to the condition, S_t *exists* & $t \notin S_t$. It may be built into the very conception of Kripke's

Kripke's paradox should also be regarded in the light of the paradox of the (alleged) village barber who shaves all and only those villagers who do not shave themselves. The barber paradox has nothing to do with cardinality issues or the like. For that matter, Russell's paradox also has little to do with cardinality or other properly set-theoretic issues.[5] Both have more to do with the following first-order-logical theorem, which I call "Russell's law":

$$\sim\exists x \forall y (Rxy \leftrightarrow \sim Ryy) .$$

This law obtains regardless of the universe of discourse and regardless of the binary relation R. It obtains if the discourse universe is the set of villagers and R is the relation, *x shaves y*. It obtains if the universe is a domain of sets and R is the relation, $y \in x$. It obtains if the universe is the set of English adjectives and R is the relation *x correctly applies to y*, and so on.[6] There can be no one who loves all and only those who do not love themselves, no list of all and only those lists that do not list themselves, no number-theoretic open formula that is provable of all and only those number-theoretic open formulas that are not provable of themselves, and so on.[7]

---

set that it is the *nonfuzzy* set of times at which Kripke is thinking of exactly one *nonfuzzy* time-set, which time-set excludes the very time of thinking.

It should be noted that there is a version of Kripke's paradox analogous to the recalcitrant liar paradox generated by 'For every level n, this very sentence is not true$_n$'. The notion of thinking-of can be replaced with thinking-of-at-some-level-or-other. The (partial) solution I propose is applicable to this strengthened version of Kripke's paradox. The thrust of Kripke's preferred solution to semantic paradoxes like that of the liar is to avoid ramification of such semantic notions as truth and designation. See his "Outline of a Theory of Truth," this JOURNAL, LXXII, 19 (Nov. 6, 1975): 690–716.

[5] It has been said that the defining condition involved in Russell's paradox does not determine a set because any collection or class that satisfies the condition is (or would be) "too big to be a set." This is like saying that the villager barber who shaves all and only those villagers who do not shave themselves does not exist because it is impossible for one person to shave that many people.

[6] British English evidently includes 'heterological'. The word has an entry in the *OED*. Paradoxically, British English nevertheless includes no adjective that correctly applies to all and only those British English adjectives that do not correctly apply to themselves. It cannot include such an adjective. What, then, of the word 'heterological'? If (as a matter of logic alone) there is no adjective that correctly applies to all and only those adjectives that do not correctly apply to themselves, then how exactly does the word (or the pre-existing adjectival phrase 'adjectival and non-self-applicable') work semantically? It is arguable (and maybe correct) that in the very attempt to fix application conditions for the word, the definition illegitimately presupposes that the word already has application conditions; therefore the definition misfires, leaving the word without application conditions—or at least with no fact of the matter concerning whether it applies to itself.

[7] On the other hand, as Gödel showed, insofar as the notion of proof is expressible, there is a number-theoretic open formula that is *satisfied by* all and only those

Applying Russell's law to Kripke's paradox, we let the discourse universe be the set of instants of time and let R be the relation, S_{t_1} *exists* $\rightarrow t_2 \in S_{t_1}$, that is, $\lambda t_1 t_2$[the only time-set Kripke is thinking of at t_1, if a unique such set exists, includes t_2 among its elements]. Then,

$$\vdash \quad \sim\exists t_0 \forall t[\,(S_{t_0} \text{ exists} \rightarrow t \in S_{t_0}) \leftrightarrow (S_t \text{ exists} \ \&\ t \notin S_t)\,].^8$$

This truth of first-order logic cannot be evaded. There is no such time as t_0 for the same reason that there is no such village barber: The supposition that there is generates contradiction.

But wait. At t_0 Kripke uses the set-theoretic expression *D*, entertaining the time-set concept expressed, without simultaneously thinking of any other set of times. It would then seem that Kripke is thinking at t_0 of $\{t \mid S_t$ exists & $t \notin S_t\}$ and of no other set, so that S_{t_0} exists and is $\{t \mid S_t$ exists & $t \notin S_t\}$. If so,

$$K\!: \quad \forall t[\,t \in S_{t_0} \leftrightarrow (S_t \text{ exists} \ \&\ t \notin S_t)\,].$$

Given that S_{t_0} exists, contradiction ensues. This appears to prove, paradoxically, that Kripke never thinks of $\{t \mid S_t$ exists & $t \notin S_t\}$ without simultaneously thinking of any other set of times. But he evidently did exactly that, in writing the very paper in which he sets out the paradox. The argument for this contradiction just *is* Kripke's paradox about time and thought.

In fact, Kripke never does think of $\{t \mid S_t$ exists & $t \notin S_t\}$, and for a simple reason: There is no set there for Kripke to think of. Kripke does entertain the relevant time-set concept, but there is no set of which it is a concept. The expression *D*, which Kripke uses at t_0, does not designate anything. Kripke writes, "We are simply dealing with a subset of the set of all times, defined by the axiom of separation....The only assumption made is that I am free to think of the set [$\{t \mid S_t$ exists & $t \notin S_t\}$] at a chosen time t_0. Not only

---

number-theoretic open formulas that are not provable of themselves. Assuming the open formula is not provable of any numbers that do not satisfy it (number-theoretic formulas provable of themselves), it is not provable of all those numbers that satisfy it.

[8] It might be clearer if we let the universe be the set of times at which Kripke thinks of exactly one set of times, and let R be the relation, $t_2 \in S_{t_1}$. So interpreted, Russell's law states that there is no time t_0 at which Kripke thinks of exactly one set of times such that $S_{t_0} = \{t$ for which S_t exists $\mid t \notin S_t\}$.

C. Anthony Anderson suggests a second-order version of Russell's law:

$$\sim\exists f \exists R \exists x \forall y[R(yf_x) \leftrightarrow \sim R(yf_y)].$$

This states that there is no triple consisting of a function *f*, a binary relation R, and an individual *x*, that stand in the indicated complex relation.

does this assumption seem quite unexceptionable, I have in fact fulfilled it."[9] On this point I must differ. The paradox itself is the eating of the pudding. One is tempted to take it for granted that

H: $\{t \mid S_t \text{ exists } \& \ t \notin S_t\} = S_{t_0}$.

This hypothesis, however, simply presupposes, on the authority of Separation and without independent proof, that Kripke's set exists. (See again note 8.) More cautiously, we may take it as given that

P: $\{t \mid S_t \text{ exists } \& \ t \notin S_t\} \text{ exists} \rightarrow \{t \mid S_t \text{ exists } \& \ t \notin S_t\} = S_{t_0}$.

Notice that P states a contingent, *a posteriori* fact. Kripke might have thought at t_0 of the set of times at which television is unknown, or of no set at all. It follows from P that

C_1: $\{t \mid S_t \text{ exists } \& \ t \notin S_t\} \text{ exists} \rightarrow \forall t[t \in S_{t_0} \leftrightarrow (S_t \text{ exists } \& \ t \notin S_t)]$.

The consequent of C_1 is K, which, as we have just seen, is inconsistent with S_{t_0}'s existence. Our premise P thus entails

C_2: $\{t \mid S_t \text{ exists } \& \ t \notin S_t\}$ does not exist.

It also precludes the paradoxical hypothesis H. Specifically, if $\{t \mid S_t \text{ exists } \& \ t \notin S_t\}$ exists, then t_0 both is an element and is not. Therefore $\{t \mid S_t \text{ exists } \& \ t \notin S_t\}$ does not exist. Since it does not, Kripke never thinks of it. There is no set there for Kripke to think of.

The expression D is an abbreviation for the following definite description:

D': $\iota S \forall t[t \in S \leftrightarrow (S_t \text{ exists } \& \ t \notin S_t)]$.

D' designates a set S if and only if that set uniquely satisfies the defining condition given by '$\forall t[t \in S \leftrightarrow (S_t \text{ exists } \& \ t \notin S_t)]$'. If any set uniquely satisfies this condition, then S_{t_0} exists and is the only set that satisfies the condition. It follows that no set uniquely satisfies the defining condition. The description D' is improper. Likewise, D does not designate.[10]

There is a wrinkle. Unlike the concept of a village barber who shaves all and only those villagers who do not shave themselves (also unlike the concept of the set of sets that are not elements of themselves, and unlike the concept of an English adjective that correctly

[9] Kripke, "A Puzzle about Time and Thought," pp. 373–75.

[10] Perhaps instead of saying that $\{t \mid S_t \text{ exists } \& \ t \notin S_t\}$ does not exist and D does not designate, it should be said more cautiously that *there is no fact that* $\{t \mid S_t \text{ exists } \& \ t \notin S_t\}$ exists or that D designates, or even something meta-meta-theoretic, for example, that D is not an element of the meta-extension of 'designates'.

applies to all and only those English adjectives that do not correctly apply to themselves, and so on), the defining condition for Kripke's putative set is itself perfectly consistent. It is merely inconsistent with a contingency: the fact that at t_0 Kripke entertains the concept that he does, that expressed by D. The defining condition is only accidently empty; there could easily have been a time-set of which it would have been a concept. (By contrast, where R is the relation, S_{t_1} *exists* $\rightarrow t_2 \in S_{t_1}$—or any other binary relation—the concept of a time that bears R to all and only those times that do not bear R to themselves is indeed inconsistent. There could not be such a time.) Although D does not designate, it is not a rigid *non*designator. Let w be a possible world in which Kripke never entertains the set concept expressed by D, nor any set concept that mathematically entails it. With respect to w, D designates a particular set of times: the set of times at which Kripke is thinking in w of exactly one time-set, which time-set excludes the very time of thinking in w. In w there does exist such a set—only it is not *Kripke's* set in w. (Someone else might think of it.) The fact that D does not designate with respect to the actual world is a byproduct of the fact that in the actual world, Kripke entertains the concept expressed by D. Had he never entertained any entailing concept, D would have designated a particular set. But he does; consequently, it does not. The defining concept of Kripke's set is jiggered in such a way that if he ever entertains it (without simultaneously thinking of any other set), then there is nothing of which it is a concept. Many other such concepts are there for the grasping, for example, *the smallest natural number that Kripke never thinks of.*[11]

The premise P evidently conflicts with the instance of the extension of Separation that yields Kripke's paradox. Whereas P is merely contingent and *a posteriori*, it is about as certain as any *a posteriori* truth can be. It evidently illustrates that the extension of Separation to some conditions that are not purely set theoretic (in particular to the condition, S_t *exists* & $t \not\in S_t$) is erroneous, even if only contingently. This should not be seen as the loss of a set, or as a mysterious limitation on our ability to generate subsets, or anything of the sort. Not a single set is lost in refraining from extending Separation as proposed in the paradox. In particular, there exists $\{t \mid S_t$ exists & Kripke does not entertain at t any concept that entails $\{t \mid S_t$ *exists* & $t \not\in S_t\}$ & $t \not\in S_t\}$. There also exists the union of this set with $\{t \mid$ Kripke entertains

[11] C. Anthony Anderson informs me that a remark of Russell's in *Problems of Philosophy* prompted him to consider the concept, *the smallest natural number that will never have been explicitly thought of.*

at t a concept that entails *{t | S_t exists & $t \notin S_t$}*, as do all the sets in between.[12] The defining concept of the Kripke set wishes to be a concept of one or another of these sets. That wish, like so many others, remains unfulfilled. But all the sets are there in place.

Kripke's putative set fails to exist not for set-theoretic reasons, but for a first-order-logical reason—like the reason that there is no village barber who shaves all and only those villagers who do not shave themselves. In order for such a village barber to exist, he would have to shave himself and also not shave himself, nothing less. Likewise, in order for such a set as Russell's putative set to exist, it would have to be an element of itself and also not an element of itself, nothing less. Given premise P, in order for such a set as Kripke's putative set to exist, t_0 would have to be both an element and not, nothing less. It is logically impossible for something to be both an element and not an element of the same set. Given the contingent fact that he at some point entertains its defining concept, Kripke's set does not exist.

The observation that Kripke's set does not exist provides only a partial solution. Kripke's paradox is more like the Berry paradox than that of Russell. Consider the following stronger variant of Kripke's paradox: Suppose that t_1 is the first instant at which Kripke entertains any concept of the form, *the time that is such and such*, and that he does so by entertaining the following concept, and without simultaneously thinking of any other time:

> *the earliest instant t such that: (i) Kripke is thinking at t of exactly one instant; and (ii) t itself is not the instant Kripke is thinking of at t.*

(The supposition is far-fetched, but possible all the same.) Let $\ulcorner T_\alpha \urcorner$ be an abbreviation for $\ulcorner$the time Kripke is thinking of at $\alpha\urcorner$. It would appear that if $t_1 = T_{t_1}$, then $t_1 \neq T_{t_1}$; but also conversely if $t_1 \neq T_{t_1}$, then $t_1 = T_{t_1}$. This variant of the paradox replaces Kripke's nonexistent set with the first time Kripke thinks of a different time and no other. The new paradox involves no set theory, just time and thought, or more accurately, time and thinking-of.[13] It is resilient. It is not put to rest by asserting that Kripke does not think at t_1 of

[12] Recall the use of italics as a means of indirect quotation.

[13] Kripke was undoubtedly aware of this variant. Compare it with the following variant of Berry: Suppose that t_1 is the earliest time at which Kripke utters a time description, and that the only time description he utters at t_1 is this:

> the first time I utter a time description that does not designate that very time.

(Again, this is far-fetched but nonetheless possible.) It would appear that if the description designates t_1, then it does not, and conversely.

any time, and that therefore T_{t_1} does not exist. If Kripke ever thinks later than t_1 of exactly one time, which time is not the instant of thinking ("that moment in kindergarten when I realized I was smarter than the teacher"), then there is an earliest such time, t_2, so that T_{t_1} does exist, even if this fact is merely contingent. (It can even be a further stipulated hypothesis. It replaces the extension of Separation to Kripke's putative subset of the set of times.) In that case, $t_1 \neq t_2 = T_{t_1}$. The paraconsistent spiral begins again. A complete solution requires some further idea.[14]

More accurately still, Kripke's paradox about time and thought is not really about time. Suppose that Kripke points to the figure in the mirror and, seeing that the person is deep in thought but not realizing it is himself, gratuitously entertains the following concept:

> *The number n such that n = 1 if he is thinking descriptively of exactly one number and that number is not 1, and n = 0 otherwise.*

(Suppose for the present purpose that Kripke thinks of 0 and 1 therewith nondescriptively.) As with the semantic paradoxes (the liar, Grelling, Berry), the source of Kripke's paradox appears to be the concept-of relation.

NATHAN SALMON

University of California, Santa Barbara

[14] See note 10.