

Some results on the limits of thought

Andrew Bacon and Gabriel Uzquiano*

January 11, 2018

Consider the principle that for a given agent S , and any proposition p , it is metaphysically possible that S is thinking p , and p alone, at time t . According to philosophical folklore, this principle cannot be true, despite its initial appeal, because there are more propositions than possible worlds: the principle would require a different possible world to witness the thinking of each proposition, and there simply aren't enough possible worlds to go around.¹ Some theorists have taken comfort in the thought that, when taken in conjunction with facts about human psychology, the principle was not on particularly firm footing to begin with: most propositions are far too complicated for any human to grasp, much less think uniquely.²

However, in 1961 Arthur N. Prior made a striking observation that undermines this diagnosis.³ Prior constructively provides an example of a proposition which cannot be thought by S at t , and thought alone, and, moreover, it is one that is not particularly complex or difficult to grasp.⁴ Indeed, we have every reason to believe that agents other than S can think it at t , and that S herself can think this proposition at times other than t . Prior's proposition is simply the proposition that everything that S is thinking at t is false. The fact that S cannot think this proposition uniquely at t follows from a general theorem of quantified propositional logic supplemented with a single unary operator Q . The language of quantified propositional logic results from that of propositional logic when we add a stock of propositional variables that take sentence position, a quantifier that binds them:⁵

*Thanks to Jeremy Goodman for feedback on an earlier version of the paper, and thanks to audiences at the University of St. Andrews.

¹This argument, attributed to Kaplan, is discussed by Martin Davies in Davies (1981) appendix 9, and by many others subsequently (Kaplan's own discussion didn't appear until Kaplan (1995)). Kaplan's puzzle begins with the observation that if the set of all worlds has cardinality κ , then the set of all propositions, determined by sets of worlds, must have cardinality 2^κ . This is incompatible with the further assumption that for each proposition p , and a given agent and time, there is a world in which the agent thinks p , and only p , at the given time. Otherwise, we would have a map f from a set of worlds *onto* the set of all propositions, i.e., the power set of the set of all worlds, which is ruled out by Cantor's theorem.

²See Lewis's discussion in §2.3 Lewis (1986).

³Prior's observation predates the literature on Kaplan's paradox, but its significance to that discussion has for the most part been overlooked. See Prior (1961).

⁴Indeed the assumption that Prior's proposition is thought uniquely is not only metaphysically impossible but logically incoherent.

⁵The resources needed to prove Prior's theorem are surprisingly minimal: you only need

Theorem. (Prior’s Theorem) $Q\forall p(Qp \rightarrow \neg p) \rightarrow \exists p(Qp \wedge p) \wedge \exists p(Qp \wedge \neg p)$

This result is perfectly general. In the above, Q may be read as any sentential operator such as “it is not the case”, “it is necessary that”, etc. However, on its most puzzling interpretations, it is read in terms of some propositional attitude such as “it is thought that”, “it is said that”, “it is feared that” or “it is hoped that”. In particular, Prior’s observation tells us that an agent cannot, at a given time, think the Prior proposition $\forall p(Qp \rightarrow \neg p)$, which we abbreviate λ , without thinking a true and a false proposition at the relevant time. So, one cannot be thinking uniquely that everything one is thinking is false.⁶

1 From Russell to Prior

Those familiar with the semantic paradoxes will have noticed a striking resemblance with the liar paradox.⁷ Less obvious is the connection between Prior’s paradox and Russell’s. The latter turns on a simple theorem of the predicate calculus:

$$\neg \exists x \forall y (Ryx \leftrightarrow \neg Ryy)$$

As with Prior’s result, this theorem is perfectly general: R could be interpreted to mean any binary relation. Suppose an object a bears R to all and only those objects that do not bear R to themselves. Then a bears R to itself if and only if it doesn’t, which is inconsistent. For example, taking R to mean “shaves”, it follows that there is no barber who shaves all and only those who do not shave themselves. Russell’s paradox arises when we let R mean: “is a member of”. On this interpretation, the above theorem tells us that no set can have as members all and only non-self-membered sets.

This theorem also has a straightforward connection to Prior’s paradox. To fix ideas, let us assume a simple model of propositions on which they are sets

classical propositional logic and universal instantiation for the propositional quantifiers. Weakening universal instantiation allows one to block the theorem, but as argued in Bacon et al. (2016) this move leaves us exposed to other problems.

⁶Prior’s proposition is related to a familiar diagonal argument for Cantor’s theorem that there is no function from the set of worlds W onto the set of sets of worlds $\mathcal{P}W$: suppose $f : W \rightarrow \mathcal{P}W$ maps W onto $\mathcal{P}W$ and consider $\Lambda = \{w \in W : w \notin f(w)\}$. If w is such that $f(w) = \Lambda$, then $w \in \Lambda$ iff $w \notin \Lambda$. If we interpret f as mapping w to the set of worlds at which belong to some proposition thought at w , then Λ just becomes the set of worlds at which Prior’s proposition is true. It follows that at no world is one thinking propositions whose disjunction is necessarily equivalent to Prior’s proposition. Thus, in particular, it follows that Prior’s proposition is not thought uniquely at any world. A similar connection is explored in Moore (1984) (above we exploited a proof that no function from W to $\mathcal{P}(W)$ is onto, whereas Moore exploits a proof that no function from $\mathcal{P}(W)$ to W is one-one. Since there many worlds at which any given proposition is thought uniquely, Moore’s analysis requires the axiom of choice, whereas ours does not).

⁷Unlike the liar, however, Prior’s theorem is provable in quantified propositional logic without appeal to any controversial disquotational premises.

of possible worlds.⁸ Let us introduce a binary relation, Rxy , on W , which we shall pronounce as ‘ y is accessible from x ’:⁹

Accessibility Rxy if and only if some proposition thought at x is true at y .

If x is not accessible to itself, then no proposition thought at x is true at x . When we interpret the Prior sentence $\forall p(Qp \rightarrow \neg p)$ (henceforth, λ) in this setting, it expresses the set of worlds where no proposition thought at that world is true at that world. Thus Prior’s sentence just expresses the set of worlds that are not accessible to themselves. Russell’s theorem immediately tells us that there is no world which is accessible from all and only the worlds that are not accessible to themselves, i.e., the worlds at which Prior’s proposition is true. It follows that there is no world at which we think λ and nothing else. For the worlds accessible to such a world x would be exactly those worlds at which λ is true, i.e., worlds that are not accessible to themselves, contradicting the theorem above.

In fact, Russell’s theorem suggests a strengthening of Prior’s result. After we learn that some propositions cannot be thought uniquely, one is led to wonder whether, for $n > 1$, there are n propositions p_1, \dots, p_n such that one cannot think all of them without thinking any further propositions. The answer is yes. Take n propositions p_1, \dots, p_n whose disjunction is necessarily equivalent to λ . Then if I think p_1, \dots, p_n , and nothing else, at a world x , then the worlds accessible to x are exactly the worlds at which at least one of p_1, \dots, p_n is true. But those are exactly the worlds at which λ is true, which by the above argument is impossible.

This style of argument can be turned into a proof in quantified propositional logic. Suppose the disjunction of p_1, \dots, p_n is materially equivalent to λ . Now, assume I’m thinking all of p_1, \dots, p_n . Then, if λ is true, all of p_1, \dots, p_n are false and their disjunction cannot be materially equivalent to λ . Thus some proposition q I’m thinking is true. However, q cannot be one of p_1, \dots, p_n , since otherwise their disjunction would be true and thus not materially equivalent to λ . It follows that I can only think all of p_1, \dots, p_n if they are all false and I’m thinking some further true proposition q .¹⁰

⁸This model is not uncontroversial — indeed Deutsch Deutsch (2013) has questioned whether a possible world is the sort of thing that can be a member of a set or class. However, as we shall see later, nothing hangs on this modeling assumption. This result can be formulated, as Prior’s is, in the language of modal quantified propositional logic.

⁹In epistemic logic we say that a world y is epistemically accessible to x iff every proposition known at x is true at y . Parallel arguments can be made in the present context with an analogous universal definition of accessibility, but the existential definition makes some proofs more straightforward.

¹⁰More formally, in quantified propositional logic, we can argue as follows:

1. $((p_1 \vee \dots \vee p_n) \leftrightarrow \forall p(Qp \rightarrow \neg p)) \rightarrow ((Qp_1 \wedge \dots \wedge Qp_n) \rightarrow \neg(p_1 \vee \dots \vee p_n) \wedge \neg \forall p(Qp \rightarrow \neg p))$
2. $((p_1 \vee \dots \vee p_n) \leftrightarrow \forall p(Qp \rightarrow \neg p)) \rightarrow ((Qp_1 \wedge \dots \wedge Qp_n) \rightarrow (\neg p_1 \wedge \dots \wedge \neg p_n) \wedge \exists p(Qp \wedge p))$
3. $((p_1 \vee \dots \vee p_n) \leftrightarrow \forall p(Qp \rightarrow \neg p)) \rightarrow ((Qp_1 \wedge \dots \wedge Qp_n) \rightarrow \exists p(Qp \wedge (p \neq p_1 \wedge \dots \wedge p \neq p_n)))$

2 From Mirimanoff to Prior

These observations naturally lead us to wonder whether other set-theoretic paradoxes can be turned into corresponding paradoxes in quantified propositional logic. We have in mind a paradox of naive set theory due to Mirimanoff according to which there is no set of all well-founded sets.¹¹ A set is well-founded if it heads no infinite descending membership chain. A set x that belongs to itself is not well-founded because it heads the infinite descending chain $x \ni x \ni x \dots$. Thus Russell's paradox shows that there is no set of all sets that are non-well-founded in this particular way – i.e., there is no set of all non-self-membered sets. However, there are many other ways in which a set could be non-well-founded. Take a pair of non-self-membered sets x and y that are members of each other. Then each of them heads an infinite descending chain, i.e., $x \ni y \ni x \ni \dots$ and $y \ni x \ni y \ni \dots$. Or one could have an infinite descending chain of pairwise distinct sets. And so on.

Mirimanoff argued roughly as follows: suppose W is a set of all well-founded sets. Then, on the one hand (i) every set of well-founded sets is itself well-founded. On the other hand, (ii) no self-membered set is well-founded. To establish (i), note that a set x_0 can only head an infinite descending chain $x_0 \ni x_1 \ni x_2 \ni \dots$ if at least one of its members, namely x_1 itself heads an infinite descending chain. So, no set containing only well-founded sets can fail to be well-founded. For (ii), clearly any self-membered set x heads an infinite descending chain, namely, $x \ni x \ni x \dots$. If W is a set of all well-founded sets, then, by (i), W is well-founded and thus a member of itself contradicting (ii).

To motivate our paradox in quantified propositional logic, let us return to our simple possible worlds model. As before, say that a world x is accessible to a world y , written Rxy , if and only if some proposition thought at x is true at y . A world x_0 is well-founded if and only if it heads no infinite chain of the form $x_0Rx_1Rx_2Rx_3\dots$. Call the set of all such worlds W . Is there any world, w , at which the proposition corresponding to W , and that proposition alone, is thought? An argument reminiscent of Mirimanoff's suggests that the answer is no. On the one hand, (i) only well-founded worlds are accessible to a world w in which W alone is being thought, which means that w must be well-founded. On the other hand, (ii) no world belonging to some proposition being thought at that world is well-founded. To establish (i), note that if W is uniquely thought at w , then, by definition of accessibility, every world accessible to w is a member of W . So, w cannot head an infinite chain $wRx_1Rx_2Rx_3\dots$ for otherwise, x_1 would not be well-founded and thus not accessible to w . For (ii), clearly if W is true at w , then w is accessible to itself and thus non-well-founded. Since, by (i), w is well-founded, $w \in W$, which means that w belongs to a proposition being thought at w and is thus non-well-founded by (ii).

¹¹See Mirimanoff (1917).

3 Another result on the limits of thought

This type of reasoning was thoroughly model-theoretic. However, it turns out that there is a sentence of the limited language of modal quantified propositional logic that is true at exactly the well-founded worlds, and thus one can state in this language that this and only this proposition is being thought.

The language of modal quantified propositional logic is the language that results from quantified propositional logic when we add a modal operator \Box , which we interpret as “it is metaphysically necessary that.” As before, we also include a unary propositional operator Q . The standard model theory for this language takes a model to consist of a set of possible worlds or indexes I , a relation of relative possibility on I for the interpretation of \Box , and a function that assigns to each index in I a set of sets of worlds in I . This tells us what sets of worlds lie in the interpretation of Q at the world. Incidentally, it is important not to confuse the relation of relative possibility used for the interpretation of \Box with R , the relation we officially called “accessibility” at the outset. In this setting, R relates w and w' just in case there is some subset J of I such that J belongs to the extension of Q at w and w' is in J .

Consider the following sentence:

$$\exists p(p \wedge \Box(p \rightarrow \exists q(Qq \wedge \Diamond(p \wedge q))))$$

One can show that, in a possible worlds model where the relation of relative possibility is universal, this sentence is true at a world w if and only if w heads an infinite chain $wRw_1Rw_2\dots$ of members of I . Informally, the sentence says that the world of evaluation, w , belongs to a set of worlds, X , such that, given any world x in X , there is at least one set Y corresponding to a proposition thought at x , which is true at some member of X . In other words, w belongs to a set of worlds, X , such that every member of X is related by R to some other member of X . Clearly this can only happen if w belongs to an infinite accessibility chain of the sort described above.¹²

It follows that the following sentence expresses well-foundedness:

$$(\gamma) \neg \exists p(p \wedge \Box(p \rightarrow \exists q(Qq \wedge \Diamond(p \wedge q))))$$

Given any model of the sort we considered before, the sentence γ is true at a world if and only if w is well-founded.

Given our observations from the previous section, it follows that no world in any such model can have γ as the only set in the extension of Q at that world. For γ expresses the proposition corresponding to W and we know that no world is accessible to all and only members of W .

Nothing rests on the details of the model theory we have chosen. We can turn the reasoning from last section into a purely syntactic proof that γ cannot

¹²The above reasoning assumes that every world is possible relative to every other world, but the formal derivation of the paradox will only require the principles of S4. To show that the displayed sentence expresses the set of non-well-founded, without the assumption that every world is possible relative to every other world, a more subtle argument is required.

be Q uniquely. In particular, the following is a theorem of modal quantified propositional logic with S4.

Proposition 1. $\vdash \neg\forall p(Qp \leftrightarrow p = \gamma)$

The proof proceeds by analogy with the steps (i) and (ii) in Mirimanoff's paradox. We firstly wish to prove (i): that if p corresponds to a set of well-founded worlds – which we may formalise as $\Box(p \rightarrow \gamma)$ – then any world at which p is Q uniquely is itself well-founded – formalised $\Box(\forall q(Qq \rightarrow q = p) \rightarrow \gamma)$:

Lemma 2. $\vdash \Box(p \rightarrow \gamma) \rightarrow \Box(\forall q(Qq \rightarrow q = p) \rightarrow \gamma)$.

We rely on instances of schema T ($\phi \rightarrow \Diamond\phi$) and schema 4 ($\Diamond\Diamond\phi \rightarrow \Diamond\phi$). The proof may be found in the appendix.

We then wish to show (ii): that a world w is well-founded only if it is not accessible to itself. In other words, a world at which γ is true is a world w which does not belong to anything thought at w . That is, we wish to prove:

Lemma 3. $\vdash \gamma \rightarrow \forall p(Qp \rightarrow \neg p)$.

This lemma makes use of schema T . See appendix for proof.

The proof of Proposition 1 is now immediate. On the one hand we can infer from lemma 3 that $\gamma \rightarrow (Q\gamma \rightarrow \neg\gamma)$, and thus that $(Q\gamma \rightarrow \neg\gamma)$ by propositional logic. On the other hand from $\Box(\gamma \rightarrow \gamma)$ and T , we can infer from lemma 2, that $\forall p(Qp \rightarrow p = \gamma) \rightarrow \gamma$. So, we have $\forall p(Qp \rightarrow p = \gamma) \rightarrow \gamma$, and by universal instantiation that $\forall p(p = \gamma \rightarrow Qp) \rightarrow Q\gamma$. But since we have $(Q\gamma \rightarrow \neg\gamma)$, we cannot have both directions in the biconditional, and we conclude that $\neg\forall p(Qp \leftrightarrow p = \gamma)$ by propositional logic.¹³

Much like Prior's proposition, λ , given an agent, we have identified a proposition γ , which is not thought uniquely by the agent on pain of contradiction. But how exactly are the two propositions related? There would be little point to our observation if they proved to be equivalent to each other. However, they are not equivalent: in S4, λ is provable from γ , but the converse fails.

Proposition 4. $\gamma \vdash \forall p(Qp \rightarrow \neg p)$

This is an immediate consequence of lemma 3. However, it is not difficult to check that the converse fails:

Proposition 5. $\forall p(Qp \rightarrow \neg p) \not\vdash \gamma$

¹³In symbols:

- | | |
|---------------------------------------------------------------|-----------------------------|
| 1. $Q\gamma \rightarrow \neg\gamma$ | from lemma 3 by logic |
| 2. $\forall p(Qp \rightarrow p = \gamma) \rightarrow \gamma$ | from lemma 2 by logic and T |
| 3. $\forall p(p = \gamma \rightarrow Qp) \rightarrow Q\gamma$ | logic |
| 4. $\neg\forall p(Qp \leftrightarrow p = \gamma)$ | from 1, 2, and 3 by logic |

Recall that λ is true at worlds that are not accessible to themselves, and thus rules out a very special sort of non-well-foundedness, whilst γ rules out all kinds of non-well-foundedness. If one can find a non-well-founded world that is not of the special, accessible-to-itself, kind then there are worlds at which λ is true but not γ .

It is thus a simple matter to find possible worlds models of the relevant sort where λ but not γ is validated. Take a possible worlds model in which I contains at least two worlds a and b where every world is possible relative to every other world. We can now make each world accessible in the sense of R from the other but not from itself. One way to guarantee this is by stipulating that the extension of Q at a is $\{\{b\}\}$ and the extension of Q at b be $\{\{a\}\}$. γ will be false at each a and b , since there is an infinite accessibility chain $aRbRaRb\dots$, but neither of these worlds are accessible to themselves so λ is true at both.

Even if γ and λ are not provably equivalent, one may reasonably wonder whether, given the actual structure of modal reality, they are nevertheless necessarily equivalent. The above model suggests a sufficient condition on modal reality for the non-equivalence of λ and γ : that there are a pair of propositions, p_1 and p_2 , such that it is possible to think p_1 uniquely whilst p_2 is true and p_1 false, and possible to think p_2 uniquely whilst p_1 is true and p_2 false, then it's possible that λ is true and γ false. Indeed, in S5 one can prove the following:

Proposition 6. $\diamond(\forall q(Qq \leftrightarrow q = p_1) \wedge p_2 \wedge \neg p_1) \wedge \diamond(\forall q(Qq \leftrightarrow q = p_2) \wedge p_1 \wedge \neg p_2) \rightarrow \neg \Box(\lambda \rightarrow \gamma)$

This suggests that γ and λ are genuinely different and not necessarily equivalent propositions.¹⁴ So, our adaptation of Mirimanoff's paradox to the language of modal quantified propositional logic results in further information on the limits of thought beyond what we learnt from Prior's original observation.

4 Appendix

Here we sketch the outline of a proof of lemmas 2 and 3 in modal quantified propositional logic with S4.

We begin with lemma 2, which reads:

$$\Box(p \rightarrow \gamma) \rightarrow \Box(\forall q(Qq \rightarrow q = p) \rightarrow \gamma).$$

Steps that can be filled in using ordinary quantificational reasoning are omitted. To show our claim it suffices to prove:

$$(\forall q(Qq \leftrightarrow p = q) \wedge \neg \gamma) \rightarrow \diamond(p \wedge \neg \gamma)$$

For if this conditional is provable then it is necessary: $\Box((\forall q(Qq \leftrightarrow p = q) \wedge \neg \gamma) \rightarrow \diamond(p \wedge \neg \gamma))$, and so one can conclude $\diamond(\forall q(Qq \leftrightarrow p = q) \wedge \neg \gamma) \rightarrow$

¹⁴It is interesting to note that in some sense, λ is just the demodalisation of γ . If one deletes the modal operators appearing in γ one gets the formula $\neg \exists p(p \wedge (p \rightarrow \exists q(Qq \wedge (p \wedge q))))$, which by inspection is logically equivalent to $\forall p(Qp \rightarrow \neg p)$. This suggests that in the special case where there is only one possible world, the two paradoxes coincide.

$\diamond\diamond(p \wedge \neg\gamma)$, which implies $\diamond(\forall q(Qq \leftrightarrow p = q) \wedge \neg\gamma) \rightarrow \diamond(p \wedge \neg\gamma)$ in S4. The lemma follows by contraposition and applying duality and the de Morgan laws.

The antecedent amounts to the following two claims:

1. $\forall q(Qq \leftrightarrow p = q)$
2. $\exists r(r \wedge \Box(r \rightarrow \exists q(Qq \wedge \diamond(r \wedge q))))$

Claim 2 says that some truth, r , necessitates the claim that something Q is compossible with r . It follows that in fact, something Q is compossible with r . Since, by 1, p is the only Q proposition, it follows that p in particular is compossible with r .

3. $\diamond(p \wedge r)$

Moreover, given that r necessitates the claim that something Q is compossible with r , it's necessary that it necessitates this claim, by S4:

4. $\Box\Box(r \rightarrow \exists q(Qq \wedge \diamond(r \wedge q)))$

Given 3 and 4 we can infer:

5. $\diamond(p \wedge r \wedge \Box(r \rightarrow \exists q(Qq \wedge \diamond(r \wedge q))))$

using the inference from $\diamond A$ and $\Box B$ to $\diamond(A \wedge B)$. Now 5 entails

6. $\diamond(p \wedge \exists r(r \wedge \Box(r \rightarrow \exists q(Qq \wedge \diamond(r \wedge q)))))$

using existential generalization, and the fact that we can apply logic inside the scope of \diamond . 6 is just the required $\diamond(p \wedge \neg\gamma)$.

On to now a sketch of a proof of lemma 2:

$$\gamma \rightarrow \forall p(Qp \rightarrow \neg p).$$

Firstly note that one can prove, using the T axiom:

1. $\Box((Qp \wedge p) \rightarrow (Qp \wedge \diamond((Qp \wedge p) \wedge p)))$

Applying existential generalization to p in the consequent gives us:

2. $\Box((Qp \wedge p) \rightarrow \exists q(Qq \wedge \diamond((Qp \wedge p) \wedge q)))$

So by propositional logic:

3. $(Qp \wedge p) \rightarrow ((Qp \wedge p) \wedge \Box((Qp \wedge p) \rightarrow \exists q(Qq \wedge \diamond((Qp \wedge p) \wedge q))))$

Applying existential generalization in the consequent to $(Qp \wedge p)$ gives us:

4. $(Qp \wedge p) \rightarrow \exists r(r \wedge \Box(r \rightarrow \exists q(Qq \wedge \diamond(r \wedge q))))$

Or, equivalently, $(Qp \wedge p) \rightarrow \neg\gamma$. Contraposing and generalizing in p we get 5 from which 6 and 7 follow:

5. $\forall p(\gamma \rightarrow \neg(Qp \wedge p))$
6. $\gamma \rightarrow \forall p\neg(Qp \wedge p)$
7. $\gamma \rightarrow \forall p(Qp \rightarrow \neg p)$

References

- Bacon, A., J. Hawthorne, and G. Uzquiano (2016). Higher-Order Free Logic and the Prior-Kaplan Paradox. *Canadian Journal of Philosophy* 46(4-5), 493–541.
- Davies, M. (1981). *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. Routledge.
- Deutsch, H. (2013). Resolution of some paradoxes of propositions. *Analysis* 74(1), 26–34.
- Kaplan, D. (1995). A problem in possible-world semantics. In W. Sinnott-Armstrong, D. Raffman, and N. Asher (Eds.), *Modality, Morality and Belief*, pp. 41–52. Cambridge University Press.
- Lewis, D. K. (1986). *On the Plurality of Worlds*. Blackwell Publishers.
- Mirimanoff, D. (1917). Les antinomies de Russell et de Burali-Forti: et le problème fondamental de la théorie des ensembles.
- Moore, A. W. (1984). Possible worlds and diagonalization. *Analysis* 44(1), 21–22.
- Prior, A. (1961). On a family of paradoxes. *Notre Dame Journal of Formal Logic* 2(1), 16–32.