# LOGICAL POLYGRAPH

Yaël COHEN

The first publication of Nelson Goodman was not a piece which would normally be considered to be of special philosophical interest. It was a logical puzzle, a 'daily brain-teaser' about liars and truth tellers which appeared anonymously on the front page of the *Boston Post* in June 1931[1]. In his book *Problems and Projects*, Goodman mentiones a variation of this puzzle, which, in its different forms, is quite familiar by now:

> There is a tribe whose population is divided into inveterate liars and unexceptionable truth tellers. A stranger approaching a fork in the road wants to ascertain from a native, who may be either a liar or a truth teller, which way leads to the capital. How can he ask one question, to be answered 'yes' or 'no', so that the answer will tell him which way to go?

This puzzle presents no special problem. The answer which comes to mind is likely to be of the following type:

(1) if I *were* to ask you if the right-hand road leads to the capital, *would* you say 'yes'?

No matter whether the questioned man happenes to be a liar or a truth teller he has to answer 'yes' if the right-hand road leads to the capital, otherwise he has to answer 'no'.

This solution is fine as long as we do not have scruples about counterfactuals. But since counterfactuals are known to be problematic, Goodman wondered "whether there is a truth-functional solution to this problem – a non-subjunctive, non-contrary-to-fact question the traveller can use"[2]. He also indicated that there is such a solution.

What Goodman probably envisaged was the following type of question:

(2) Are you a liar if and only if the right-hand road leads to the capital?

Both, the liar and the truth teller will have to answer 'no' if the right-hand road leads to the capital; otherwise they both have to say 'yes'. Again, there seems to be no special problem involved here.

However, we can think of a more sophisticated version of this puzzle which is also much more intriguing. Imagine that the son of our traveller is also a traveller. He learns from his father about the tribe of liars and truth tellers and decides to go there. Having also learned about genetics, he now expects the tribesmen to be of three kinds: obstinate liars, unshakeable truth tellers and capricious liars, who sometimes tell the truth and sometimes lie at will. Could our young traveller find his way to the capital by asking a single yes-or-no question the first tribesmen he happens to meet?

Here it may seem to us that no question which would do this job is feasible. Yet, however surprising it might be, such a question can be formulated and it comes out in a form which is quite similar to question (2):

> Are you going to lie to this question if and only if the right-hand road leads to the capital?

or more patently:

(3) Are you going to lie to question (3) if and only if the right-hand road leads to the capital?

This formulation enables the traveller to derive the needed information since, in relation to question (3), there are not three types of potential informants, but only two. One group consists of those who are going to lie to the question (no matter whether they are persistent liars or whether they just chose to lie at this particular occasion), while the other group consists of those who would answer truthfully (again irrespectively of whether they are incorrigible truth tellers or only occassional ones). No matter whether the tribesman is deceitful or truthful in replying to question (3), his answer will be 'no' ('yes') if

the right-hand (left-hand) road leads to the capital.

Now again, the solution appears to be fairly straightforward and seems to have no startling implications. However, when we come to think about it, we find out that we have actually discovered how to elicit the truth from *anybody* (or more accurately – the truth according to anyone's best knowledge). We should note that this tricky question for obtaining truthful answers is not peculiar to our outlandish tribe, but that it applies just as well to the free-willed members of our permissive society which does not outlaw inveracities. This may be somewhat disquieting. For should you have any doubts about your wife's infedelity, all you have to do is to ask her: "Are you going to lie to this question if and only if you are being unfaithful to me?". If she is unfaithful she has to answer 'no', whether or not she decides to be frank about it; while if she is loyal, she has to say 'yes', again irrespectively of whether or not she chooses to be truthful. It thus seems that we have discovered a reliable and relatively simple way how to get rid of certain potentially disturbing uncertainties. Moreover, our formula can easily be generalized so that any information whatsoever can be extracted from its possessor even if he chooses to lie. The generalized question has the following form:

(4) Are you going to lie to (4) if and only if P?

Since we can easily think of more incriminating questions to which our magic formula would provide truthful answers, we may indeed wonder what do we need such sophisticated instruments as polygraph for. Is it really possible that by force of logic alone we can force people to tell the truth?

Before we turn to this strange almost paradoxical consequence of our innocently looking puzzle, we should first analyze our question by transforming it into a declarative statement. Both the truthful and the lying person who would answer 'yes' to question (4) should be ready to assert the following statement:

(5) I am lying by asserting (5) if and only if P.

By the same token, both the liar and the truth teller who would answer 'no' to question (4) should be ready to assert a kind of negation of (5):

(6) I am lying by asserting (6) if and only if it is not the case that *P*.

Let us examine the truth conditions of these statements, taking into account: (a) the analysis of "if and only if" by means of material equivalence, (b) the dependence of "I am lying by asserting (5)" (to be called $Q$ thereafter) on the truth value of (5) as a whole, i.e., if (5) is true $Q$ must be false and if (5) is false $Q$ must be true. Now we may consider the following truth table:

| (5) | $Q$ | $P$ | (5) = $(Q \equiv P)$ |
|---|---|---|---|
| T | F | T | F |
| | | F | T |
| F | T | T | T |
| | | F | F |
| (I) | (II) | (III) | (IV) |

Note that the transition from column (I) to column (II) follows from (b) and the transition from columns (II) and (III) to (IV) follows from (a). The truth table shows that when $P$ is true than if (5) is true – (5) is false; and if (5) is false – (5) is true, which reminds us of the famous Liar's Paradox (LP: The statement LP is false). When $P$ is false then if (5) is true – (5) is true, and if (5) is false – (5) is false, which reminds us of the well-known counterpart of the Liar's Paradox (TP: The statement TP is true). *Nevertheless, these results do not compel us to conclude that we cannot assert (5).* What *does* follow from this analysis is that whoever asserts (5), whether truthfully or deceitfully, pragmatically implicates that he believes that $P$ is false. Analogically, it follows that whoever asserts (6), whether truthfully or deceitfully, implicates pragmatically that he believes that $P$ is true. The presupposition of asserting (5) – the falsity of $P$ – is a pragmatical presupposition of the speaker.

Objections can be raised as to whether it is at all possible to assign a truth value to $Q$ in order to construct the above truth table. It may be

argued that we cannot assign a truth value to $Q$ because $Q$ cannot be asserted on its own, i.e., without first asserting (5). However, although the truth value of $Q$ is logically dependent on the truth value of (5), the determination of the truth value of $Q$ can precede (in temporal order) the determination of (5). Prior to asserting (5) the speaker can decide whether or not he is going to lie. Moreover, the truth value assignment to $Q$ is only instrumental for the truth value assignment to (5), that is to say: $Q$ is not asserted at all.

Another objection might be the following: if (5) is accepted as a legitimate statement then it follows that a person at a given time with a given information can lie by asserting (5) just as he can tell the truth by asserting the very same statement. If one could show that accepting (5) really implies this wild conclusion, the objection would indeed be fatal, however, the objection is too hasty. When we are lying by asserting (5) and when we are asserting (5) truthfully, we are *not asserting the same statement*. For when a person lies, the pronoun "I" in $Q$ refers to the speaker *qua* a liar, while when he tells the truth, the "I" in $Q$ refers to the speaker *qua* truth teller (at the time of the utterance).

We have already noted that (5), given certain empirical information (when $P$ is true), is paradoxical, while in another state of affairs (When $P$ is false) it is non-informative (in the same way that TP is). It seems that it is the self-referential character of (5) which is responsible for these inconvenient results. Moreover, this kind of self-reference appears to be the same as the vicious self-reference of the statements involving semantic terms which leads to semantic paradoxes like the Liar's Paradox and its counterpart. Therefore those philosophers, who claimed that the kind of self-reference involved in LP and TP is responsible for the failure of these sentences to function as statements, should for the same reason object to (5). Thus, if this kind of self-reference is to be outlawed, (5) should also be banned. Or more precisely: if the self-reference of (5) and LP (or TP) are of the same kind, and if it is this kind of self-reference *alone* which is responsible for the unassertability of LP and TP, then (5) too should be rejected as unassertable. Now, I do not want to deny that the self-reference involved in (5) and LP (or TP) is the same, because it indeed is. But because there seems to be no reason to say that (5) is not assertable, the paradoxical nature of LP and TP cannot be confined to its

self-referentiality. If this is so, then all attempts to pin down the blame for generating semantical paradoxes squarely on self-reference are unsatisfactory.

Let me take up a thorough and comprehensive analysis of semantically vicious circles which was recently elaborated by Kripke. In his "Outline of a Theory of Truth" [3] Kripke constructs a formal model for the notion of truth which also pertains to the way in which the predicate 'is true' is learned. The key notion of his theory is the concept of 'ungroundedness'. The basic intuition behind this concept, which is formulated in his formal theory, is the following:

> "If a sentence... asserts that all (some, most, etc.,) of the sentences of a certain class C are true, its truth value can be ascertained if the truth values of the sentences in the class C are ascertained. If some of these sentences themselves involve the notion of truth, their truth values in turn must be ascertained by looking at *other* sentences, and so on. If ultimately this process terminates in sentences not mentioning the concept of truth, so that the truth value of the original sentence can be ascertained, we call the original sentence grounded; otherwise, ungrounded. ... whether a sentence is grounded is not in general an intrinsic (syntactic or semantic) property of a sentence, but useally depends on the empirical facts." [4]

Kripke directs our attention to two important points:

a) "there can be no syntactic or semantic theory sieve that winnow out the 'bad' cases while preserving the 'good' ones" [5];
b) "we make utterances which we hope will turn out to be grounded" [6] in virtue of some empirical facts. (I take it that Kripke means that we can make an assertion by uttering a sentence only if there are actual or possible states of affairs which make the sentence grounded.)

Let us now return to our formula (5) ("I am lying by stating (5) *iff* $P$.") and see how it scores on Kripke's theory. It turns out that (5) is ungrounded given *any* empirical facts. If $P$ is true (5) comes out ungrounded and paradoxical, and if $P$ is false (5) comes out also ungrounded but not paradoxical. So (5) is not assertable, according to Kripke, because there are no circumstances that could make (5)

grounded. However, let us look at (5). Is it unassertable? Certainly not, for a meaningful information can be communicated by (5).

Kripke might be right that we learn the notion of truth through grounded sentences. But after we have mastered this notion we adopt pragmatic principles by means of which we can use and understand even those sentences that Kripke dubbs ungrounded. One of these pragmatic rules that we adopt is a weak kind of the principle of charity: *Assign a truth value to an uttered sentence and its constituents so that the whole sentence will not become paradoxical.*

The assumption which underlines this principle is that in normal communication people do not utter paradoxical sentences. The crux of the matter is that it is sometimes possible to assign a truth value to the compound sentences not through a prior value-assignment to its constituents. We can often assume that the compound sentence has a truth value and from this infer the truth values of its constituents. Usually, when we cannot assign a *specific* truth value to the compound sentence, we cannot infer the specific truth values of the components. However, in this respect (5) is unique. No matter which truth value (T or F) we assign to the whole sentence, we commit ourselves to a specific truth value of one of its constituents – namely, the falsity of *P*.

The above mentioned principle of charity is not so peculiar. An analogous procedure is quite common when we come to understand or to learn a meaning of a term. We can often understand a meaning of a whole sentence without a prior understanding of all the terms in it. Assumptions about the intentions of the speaker carry us through the gaps of semantics.

So far I have managed, I think, to defend (5) as a legitimate statement. But did we really manage to outwit the bluffer? Did we really show that the polygraph is superfluous? Can we really always elicit the truth from anybody?

With all the trouble we have given the deceiver he still has a way out. Lying is not the only way of deceiving. We can also deceive by telling the truth if we know that we are, for some reason, expected to lie. The person who asks question (4) ("Are you going to lie to (4) *iff* *P*?") presupposes that you are either going to lie or to tell the truth. But one can refuse to play the game of truth-telling/lying and simply choose the 'yes' or 'no' answer haphazardly.

To elucidate the distinction between lying and deceiving we can imagine two kinds of question-and-answer games: We can picture our capricous liar in the truth-telling/lying game as one who before answering question (4) consults a randomizing device whose directions are either "Tell the truth" or "Tell a lie". In the deceiving game the player is pictured as one who consults a randomizing device with the directions "Answer 'yes'" or "Answer 'no'". If the information we want to elicit from the player is the truth value of (5) as a whole, there is no winning strategy against him, no matter which randomizing device he consults. However, if the information we want to evince from the player is whether $P$ is true, there is a difference between the two games. While there is no winning stratedy against the player-deceiver, who consults the "Answer 'yes'"/ "Answer 'no'" device, there is a winning strategy against the player who consults the "Tell the truth"/ "Tell a lie" machine.

What really happens is that when the lier decides to lie in response to question (4) he lies with respect to the implied statement (5) as a whole. But by this he is giving away the information concerning $P$. In this case, even though he is lying, he, so to speak, stands by his word. But of course you can choose not to give away the information concerning $P$. Then, all you have to do is to reverse the answer which you would give in the previous case. Thus you would be neither lying nor telling the truth in relation to question (4) as a whole – you would be deceiving.

I hope that this allows me to sum up that even the counter-intuitive results of accepting (5) as a legitimate statement do not amount to a serious objection. It therefore seems that (5) is a counter-example to semantic theories (including that of Kripke) which explain the illegitimacy of the Liar's Paradox (and its counterpart) as a statement. These theories suggest semantic rules by which the Liar's Paradox and its counterpart are outlawed. But these rules outlaw (5) as well.

Accepting (5) does not only enable us to solve puzzles with amusing results. Self-referential sentences which involve semantic notions may serve as a convenient way for representing the semantic relation between assertions and their presuppositions. In particular they may elucidate the *rationale* of *ad hominem* arguments against the sceptic. But these issues would require a detailed examination of their own.

*The Hebrew University of Jerusalem* Yaël COHEN

## REFERENCES

[1] Goodman N. "The Truth-tellers and the Liars", *Boston Post*, June 8, 1931: p. 1, cols. 1 and 2.
[2] Goodman N. *Problems and Projects*, Bobbs-Merrill, Indianapolis and New-York, 1972; p. 450.
[3] Kripke S. "Outline of a Theory of Truth", the *Journal of Philosophy*, LXXII, No. 19, Nov. 1975.
[4] *op. cit.*, pp. 693-4.
[5] *ibid.*, p. 692.
[6] *ibid.*, p. 694.