Counting the minds of split-brain patients

Gert-Jan C. LOKHORST

1996

Logique & Analyse, Nouvelle série, 39e année, No 155–156 (Septembre-Décembre 1996), pp. 315–324.

Abstract Using Fagin's and Halpern's local reasoning models and an epistemic variant of Jennings's and Schotch's semantics of weakly aggregative modal logic, we argue that the hypothesis that split-brain patients have two coherent minds is preferable to the hypothesis that they have one incoherent mind.

1 Introduction

The human brain consists of two hemispheres which are connected by the massive *corpus callosum* and several other, less important bundles of nerve fibres. In order to prevent the spreading of epileptic seizures from one cerebral hemisphere to the other these interhemispheric commissures have been transected in several dozens of patients. This surgical procedure was apparently successful. The resulting so-called "split-brain" patients have been studied in a series of experiments which was crowned by the 1981 Nobel prize in medicine and physiology.[1]

Long before the first split-brain operations on humans were actually carried out, philosophers, psychologists and physiologists wondered whether this operation would double the number of minds.[2] One might expect that the experimental data which are nowadays available would enable us to answer this question, but the controversy continues. Most people think that normal persons have one mind and split-brain patients two.[3] Some (Kathleen Wilkes, for example) maintain that the patients have only one mind.[4] Roland Puccetti has become famous for claiming that the experiments reveal that *all* people who have two intact cerebral hemispheres (separated from each other or not) have two minds.[5] On the other hand, Thomas Nagel has asserted that "there is no whole number of individual minds that these patients can be said to have" and that "these very unusual cases should cause us to be sceptical about the concept of a single subject of consciousness as it applies to ourselves".[6]

[1] There is a huge literature about the split-brain syndrome. Lockwood (1989), Lokhorst (1980), Marks (1980), Nagel (1971) and Puccetti (1973) give some references to good introductions.

[2] The first operations on humans were performed in the 1930s. Fechner raised the philosophical issue as early as 1860 (Zangwill 1974). The first split-brain operations on animals were carried out in the eighteenth century. They were intended to refute Lancisi's widely accepted thesis that the *corpus callosum* is the seat of the soul (Neuburger 1981).

[3] See Marks (1980) and Lokhorst (1980) for references.

[4] Wilkes (1978).

[5] Puccetti (1973).

[6] Nagel (1971), quotations from p. 409 and p. 410.

In this paper, we will examine this issue from the point of view of epistemic logic, the logic of knowledge and belief. We will first describe some more or less typical split-brain experiments. We will be particularly interested in what the patients believe and do not believe in these experiments. This provides us with various sentences describing the patients' beliefs and non-beliefs (§2). We will then examine these sentences in the light of several accounts of the semantics of belief-sentences. We will first show that the standard account conflicts with the experimental data (§3). We will then examine two alternative approaches which do not have this defect. The first approach is inspired by Fagin's and Halpern's local reasoning models. If we follow this approach, the patients turn out to have at least two minds in the experimental situations we will describe (§4). The other approach is inspired by Jennings's and Schotch's semantics for weakly aggregative modal logic. According to this alternative approach, the patients have only one mind. They are, however, unable to combine their mental contents into an integrated whole (§5). Thus, we will end up with two analyses, both in accord with the data but different with respect to the number of minds. Fortunately, there is a very simple principle which we may use to show that the local reasoning approach is superior. Thus we will conclude that the patients have two minds in the situations which we will describe (§6). Alternative analyses and alternative sets of data might force us to revise this conclusion. It is not so much the outcome of our deliberations as our way of proceeding which may be interesting to philosophers discussing the split-brain syndrome: we will try to demonstrate by example that even this field, which has thus far been dominated by purely intuitive considerations, may profit from modern philosophical logic.

2 Some Experimental Data

A typical split-brain experiment goes as follows. A picture of an object A, say an apple, is projected in the right visual half field of a split-brain patient. (The techniques by which this is accomplished do not matter for our purposes.) This input goes to the left hemisphere, in which the speech centre is located. A picture of another object B, say a banana, is projected in the other visual half field. This information goes to the non-verbal right hemisphere.

Upon interrogation, the patient will say that A was shown to him. This verbal output is due to the left hemisphere. The patient will, however, deny that B was shown to him or that A and B were shown to him: the left hemisphere has no access to the contents of the left visual half field.

In tests in which only the non-verbal right hemisphere has the opportunity to express itself (for example, tests in which the patient is instructed to identify the displayed object with the left hand, which is controlled by the right hemisphere), it will appear that the patient saw B after all. In such tests, he will, however, show no sign of having seen A, let alone A and B: the contents of the right visual half field are inaccessible to the right hemisphere.

In sum, the patient believes that A was shown, he also believes that B was shown, but he does not believe that both A and B were shown. Nor is he able to infer that both A and B were shown by reflecting upon his beliefs.

The patient's beliefs can accordingly be described by the following formula:

$$Bp \wedge Bq \wedge \neg B(p \wedge q) \tag{1}$$

Or, to be a little bit more complete, they could also be described as follows:

$$B(p \wedge \neg q) \wedge B(\neg p \wedge q) \wedge \neg B(p \wedge q) \wedge \neg B(\neg p \wedge \neg q) \tag{2}$$

In these formulas, B stands for "the agent implicitly believes that" (we say that an agent implicitly believes that p iff he believes that p or is able to deduce p from what he believes), p for "object A was shown" and q for "object B was shown". Both formulas are, of course, no more than partial descriptions of the epistemic situation.

Apart from the genuine experiment we have just described, we will also discuss two imaginary situations suggested by two thoughts experiments proposed by Michael Lockwood.[7] In these cases, the patient's beliefs are respectively described by the following two formulas:

$$Bp \wedge Bq \wedge (B(p \wedge q) \leftrightarrow \neg B(p \wedge q)) \tag{3}$$

$$(3) \wedge Br \wedge B(p \wedge r) \wedge B(q \wedge r) \tag{4}$$

A formula of the form $p \leftrightarrow \neg p$ says that p is "as true as" $\neg p$, namely half-true.

These, then, are our data. They are stated in folk-psychological terms (beliefs), so they are just the kind of data one needs when one wants to count such folk-psychological entities as minds. Let us now try to make sense of these data in logical terms.

3 Standard Epistemic Models

The standard account of the semantics of belief-sentences (originally due to Hintikka) is as follows.

A standard epistemic model is a structure

$$M = \langle W, w_0, R, V \rangle,$$

where W is a set of "possible worlds", w_0 (the actual world) is a member of W, R is a binary relation on W, and V is a function from $AT \times W$ into $\{0,1\}$, where AT is the set of atomic sentences. Given a standard epistemic model, an interpretation function I with domain $WFF \times W$, where WFF is the set of well-formed formulas, is defined as follows:

- $I(p,w) = V(p,w)$ if p is atomic
- $I(T,w) = 1$
- $I(p,w) = 1 - I(p,w)$

[7] Lockwood (1989), pp. 94–100, 290–293. Lockwood's discussion cannot be summarised in a few words, so we refer the reader to his interesting book. Lockwood's terminology is different from ours. We translate his statements as follows. (i) "The agent has experience A" $\Rightarrow$ "the agent experiences that A is the case". (ii) "Experiences A and B are co-conscious" $\Rightarrow$ "the agent experiences that A and B are the case". (iii) "Experiences A and B are quasi-co-conscious" $\Rightarrow$ "the agent experiences that A and B are the case iff he does not experience that A and B are the case". (This means that the sentences "the agent experiences that A and B are the case" and "the agent does not experience that A and B are the case" have the same truth-value, i.e., that both are half-true.) On top of this, we replace "the agent experiences that" by "the agent believes that". The latter move is harmless because experiencing is widely regarded as a propositional attitude ascription with the same formal properties as believing. Lockwood's discussion is not as clear as it could have been. For example, his "phenomenal perspectives" are nothing but similarity circles (in the sense of Carnap) with respect to co-consciousness, but he was not aware of this fact. As he confided to me, "I wish I'd known the term "similarity set" when I wrote my book" (Lockwood 1993).

- $I(p \wedge q,w) = \min\{I(p,w),I(q,w)\}$
- $I(p \vee q,w) = \max\{I(p,w),I(q,w)\}$
- $I(p \rightarrow q,w) = \min\{1,I(\neg p,w)+I(q,w)\}$
- $I(p \leftrightarrow q,w) = I((p \rightarrow q) \wedge (q \rightarrow p),w)$
- $I(Bp,w) = \min\{I(p,x) : wRx\}$, where $\min\varnothing = 1$.

If $I(p,w) = 1$, we say that p is true at w. If $I(p,w) = 0$, we say that p is false at w. If $I(p,w_0) = 1$ we say that p is true in the model we are considering. p is valid in a class of models iff p is true in all models in that class. These definitions of truth in a model and validity apply to all models described in this paper.

We read wRx as "x is an epistemic alternative of w". Thus Bp is true at w iff p is true at all epistemic alternatives of w.

A three-valued standard epistemic model is exactly the same as a two-valued standard model except that V is defined as a function from $AT \times W$ into $\{0,\frac{1}{2},1\}$. The definition of I is exactly the same as before. If $I(p,w) = \frac{1}{2}$, we say that p is half-true at w. The notions of truth in a model and validity are defined as before.[8]

It will be clear that there are no two-valued or three-valued standard epistemic models in which (1), (2), (3) or (4) are true. The standard approach is therefore too strong in view of the empirical evidence. Different analyses are called for.

4 Local Reasoning Models

The first alternative we will consider is inspired by Fagin's and Halpern's local reasoning models.[9] In this approach, an agent is viewed as a "society of minds" rather than a single mind. The beliefs of each of these minds are modelled in the same way as epistemic agents' beliefs are treated in the standard account. An agent believes that p iff at least one of his minds believes (in the standard sense) that p. Fagin and Halpern seem to leave it open whether a mind can have more than one set of epistemic alternatives, but we do not want to leave this open, so we will give a slightly different (but equivalent) account.[10]

We define a two-valued local reasoning model as a structure

$$M = \langle W,w_0,\Psi,S,R,V\rangle,$$

where W, w_0 and V are as above, Ψ is a set of "minds", S is function from W into the set of non-empty subsets of Ψ, and R is a function from Ψ into $W \times W$. Instead of $xR(i)y$ we write xR_iy. A three-valued local reasoning model is defined in the same way, except that V is a function from $AT \times W$ into $\{0,\frac{1}{2},1\}$. Given a two-valued or three-valued local

[8] This notion of validity has been axiomatised by Schotch, Jensen, Larsen & MacLellan (1978) (system **Ł$_3$M$_3$**).

[9] Fagin & Halpern (1988); Fagin, Halpern, Moses & Vardi (1995), chapter 9.6.

[10] Fagin and Halpern define a local reasoning model as a structure $M = \langle W,w_0,C,V\rangle$, where C is a function from W into the set of non-empty sets of subsets of W and the rest is as above. Furthermore, $I(Bp,w) = \max\{\min\{I(p,x) : x \in X\} : X \in C(w)\}$. They only consider the two-valued case. Each two-valued model in our sense has an equivalent model in Fagin's and Halpern's sense: simply define C by $C(w) = \{\{x : wR_ix\} : i \in S(w)\}$. The converse holds as well: define Ψ as the power-set of W, let $R = \{\langle X,\langle w,x\rangle\rangle : X \subseteq W \,\&\, w \in W \,\&\, x \in X\}$, and let $S = C$.

reasoning model, an interpretation function I with domain $WFF \times W$ is defined in the same way as before, except that

$$I(Bp,w) = \max\{\min\{I(p,x) : wR_i x\} : i \in S(w)\}$$

We read wR_ix as "x is an epistemic alternative of w according to i" and $S(w)$ as "the agent's society of minds at w". Thus Bp is true at w iff the agent has at least one mind at w such that p is true at all worlds which are epistemically alternative to w according to that mind. In other words, Bp is true at w iff at least one of the agent's minds at w believes—in the standard, Hintikka sense of the word—that p.

We say that a model is of rank n iff $♣S(w_0)♣ = n$, where $♣X♣$ denotes the cardinality of X. Note that the following formula is valid in the class of local reasoning models of rank n:

$$\left(\bigwedge_{0 \leq i \leq n} Bp_i\right) \rightarrow \left(\bigvee_{0 \leq j < k \leq n} B(p_j \wedge p_k)\right) \tag{5}$$

We may apply this semantical account as follows to the cases we have described. One may easily check that there are no models of rank 1 in which (1), (2), (3) or (4) are true. There are, however, models of rank 2 which have this property. (1) and (2) are true in all models such that $W = \{w_0,x,y\}$, $\Psi = S(w_0) = \{a,b\}$, $R(a) = \{\langle w_0,x\rangle\}$, $R(b) = \{\langle w_0,y\rangle\}$, $V(p,x) = 1$, $V(q,x) = 0$, $V(p,y) = 0$ and $V(q,y) = 1$. (3) and (4) are true in all models such that $W = \{w_0,x,y\}$, $\Psi = S(w_0) = \{a,b\}$, $R(a) = \{\langle w_0,x\rangle\}$, $R(b) = \{\langle w_0,y\rangle\}$, $V(p,x) = 1$, $V(q,x) = \frac{1}{2}$, $V(r,x) = 1$, $V(p,y) = 0$, $V(q,y) = 1$ and $V(r,y) = 0$. Thus we need at least two minds to account for the data in all four cases. There is no reason to assume that there are more than two minds.

It has never been shown that normal people's implicit belief sets are not closed under conjunction. If they believe that p is the case and also believe that q is the case, they are in principle—given enough time, energy and interest in the matter—able to infer that both p and q are the case. *Pace* Puccetti there is therefore no reason to assume that normal people have more than one mind.

5 Non-Aggregative Models

The second approach we will consider is loosely inspired by Jennings's and Schotch's semantics for weakly aggregative modal logic.[11] According to our epistemic version of these semantics, epistemic agents are basically simple. They have only a single mind. Their mental contents (i.e., that which they believe) may nevertheless lack the coherency which both the standard and the local reasoning approaches attribute to the beliefs of single minds.

We model this conception by treating an agent's epistemic alternatives as collections of worlds rather than single worlds. Thus, we retain the assumption of standard epistemic logic that Bp is true at w iff p is true at all epistemic alternatives of w. However, we do not identify these epistemic alternatives with single worlds but with collections of worlds. p is true at such a collection of worlds iff p is true at some of its members.

[11] Jennings & Schotch (1980); Schotch & Jennings (1980 *a*); Schotch & Jennings (1980 *b*); Apostoli & Brown (1995).

More precisely, a (two-valued or three-valued) non-aggregative model is a structure

$$M = \langle W, w_0, R, V \rangle,$$

where W, w_0 and V are as above and R is a binary relation between worlds and non-empty sets of worlds, i.e., $R \subseteq W \times \{X \subseteq W : X \neq \emptyset\}$. I is defined as before except that

$$I(Bp,w) = \min\{\max\{I(p,x) : x \in X\} : wRX\}$$

Thus Bp is true at w iff for all epistemic alternatives X of w, p is true at one or more members of X.

We say that a model is of rank n iff $\max\{\clubsuit X \clubsuit : w_0 RX\} = n$. Note that (5) is not valid in the class of all non-aggregative models of rank n. The following formula is valid in this class and in the class of local reasoning models of rank n as well.

$$\left(\bigwedge_{0 \leq i \leq n} Bp_i \right) \rightarrow B \left(\bigvee_{0 \leq j < k \leq n} (p_j \wedge p_k) \right) \tag{6}$$

One may easily check that there are no (two-valued or three-valued) non-aggregative models of rank 1 in which (1), (2), (3) or (4) are true. There are, however, models of rank 2 which have this property. (1) and (2) are true in all models such that $W = \{w_0, x, y\}$, $R = \{\langle w_0, \{x, y\} \rangle\}$, $V(p,x) = 1$, $V(q,x) = 0$, $V(p,y) = 0$ and $V(q,y) = 1$. (3) and (4) are true in all models such that $W = \{w_0, x, y\}$, $R = \{\langle w_0, \{x, y\} \rangle\}$, $V(p,x) = 1$, $V(q,x) = \frac{1}{2}$, $V(r,x) = 1$, $V(p,y) = 0$, $V(q,y) = 1$ and $V(r,y) = 0$. Thus we have to assume that the patient's epistemic alternatives have at least two members in all four cases. There is no reason to assume that they have more than two members. Nor is there any reason to assume that the patient has more than one mind.

Normal people's beliefs do not seem to be non-aggregative, so these considerations are irrelevant in their case.

6 Which Approach is Preferable?

At this point, we have two different approaches, both in accord with the empirical data, but producing different "mind-counts" in the cases we have described. Which approach are we to prefer?

There is a very simple principle which we may appeal to in order to solve this problem, namely: *choose the strongest notion of validity*[12] *which is compatible with the data*. It is not difficult to motivate this principle: a stronger notion of validity is obviously more useful than a weaker one because the former allows us to draw more conclusions from a given set of premises. Since validity in the class of all two-valued non-aggregative models of rank n implies validity in the class of all two-valued local reasoning models of rank n whereas the converse does not hold (formula (5) is not valid in the former class), we prefer the former notion in the two-valued case. A similar argument makes us prefer the three-valued local reasoning approach in the three-valued case. Thus we conclude that there are two coherent minds rather than one incoherent mind in all (real and imaginary) cases we have described. It is not *necessary* to assume this, but it is the boldest hypothesis we can advance without being contradicted by Nature, and therefore the most attractive one.

[12] Note added in 1998: we should perhaps have used the term "concept of belief".

Viewed in this light, Puccetti was too cautious rather than too bold in advancing the claim that normal people have two minds. Wilkes was too cautious as well, at least when her claim that split-brain people have one mind is to be understood along the lines of the non-aggregative approach. Finally, Nagel's pessimism turns out to be unfounded: as we have shown, there definitely do exist ways of counting minds in whole numbers which differentiate split-brain persons from normal people. Thus the *communis opinio* that split-brain people have two minds in the situations we have described emerges as the most reasonable view.

To end with, we want to point out that we may conceive of cases in which the principle we have mentioned is useless. Suppose there were a situation in which a person's beliefs are to be described by the following formula:

$$Bp_1 \wedge Bp_2 \wedge Bp_3 \wedge \neg B(p_1 \wedge p_2) \wedge \neg B(p_1 \wedge p_3) \wedge \neg B(p_2 \wedge p_3) \qquad (7)$$

There are local reasoning model of rank 3 in which this formula is true. There are no local reasoning models of rank 2 in which it is true. But there do exist non-aggregative models of rank 2 in which (7) is true. For example, let $M = \langle W, w_0, R, V \rangle$, where

- $W = \{w_0, \ldots, w_6\}$
- $R = \{\langle w_0, \{w_1, w_2\}\rangle, \langle w_0, \{w_3, w_4\}\rangle, \langle w_0, \{w_5, w_6\}\rangle\}$
- $\langle\langle V(p_i, w_j)_{1\leq i\leq 3}\rangle_{1\leq j\leq 6} =$ `110 001 101 010 011 100`

Since the corresponding notions of validity (validity in the class of local reasoning models of rank 3 *vis-à-vis* validity in the class of non-aggregative models of rank 2) are not comparable (neither is stronger than the other because (5), with n=3, is valid in the former class but not in the latter, and (6), with n=2, is valid the latter class but not in the former), we cannot apply our principle. It is not clear what one should say in such a case.

Fortunately, there seem to be no cases in which a split-brain person's beliefs are to be described by (7). Local reasoning models of rank 2 are accordingly quite adequate in view of the evidence.

7 Conclusion

Many people have claimed that split-brain patients have two minds in situations such as those we have described. But they have never made it very clear why they make this assertion. We hope to have given a clear argument for their case. Our analysis may well be too simple. But we think that it is safe to say that considerations such as those we have presented will also play a role in more sophisticated formal analyses.

Department of Philosophy
Erasmus University Rotterdam
E-mail: lokhorst@fwb.eur.nl

References

Apostoli, P. & Brown, B. (1995), 'A solution to the completeness problem for weakly aggregative modal logic', *Journal of Symbolic Logic* **60**, 832–842.

Fagin, R. & Halpern, J. Y. (1988), ‘Belief, awareness, and limited reasoning’, *Artificial Intelligence* **34**, 39–76.

Fagin, R., Halpern, J. Y., Moses, Y. & Vardi, M. Y. (1995), *Reasoning about Knowledge*, MIT Press, Cambridge (Mass.).

Jennings, P. M. & Schotch, P. K. (1980), ‘Some remarks on (weakly) weak modal logics’, *Notre Dame Journal of Formal Logic* **22**, 309–314.

Lockwood, M. (1989), *Mind, Brain and the Quantum: The Compound ‘I’*, Basil Blackwell, Oxford.

Lockwood, M. (1993), ‘Letter to G. J. C. Lokhorst (19 may 1993)’.

Lokhorst, G.-J. C. (1980), Homo duplex, Technical report, Department of Philosophy, Erasmus University Rotterdam, Rotterdam.

Marks, C. E. (1980), *Commissurotomy, Consciousness, and Unity of Mind*, Bradford Books, Montgomery (Vermont).

Nagel, T. (1971), ‘Brain bisection and the unity of consciousness’, *Synthese* **22**, 396–413.

Neuburger, M. (1981), *The Historical Development of Experimental Brain and Spinal Cord Physiology before Flourens (1897), translated by E. Clarke*, Johns Hopkins University Press, Baltimore.

Puccetti, R. (1973), ‘Brain bisection and personal identity’, *British Journal for the Philosophy of Science* **24**, 339–355.

Schotch, P. K. & Jennings, R. E. (1980 *a*), ‘Inference and necessity’, *Journal of Philosophical Logic* **9**, 327–340.

Schotch, P. K. & Jennings, R. E. (1980 *b*), ‘Modal logic and the theory of modal aggregation’, *Philosophia* **9**, 265–278.

Schotch, P. K., Jensen, J. B., Larsen, P. F. & MacLellan, E. J. (1978), ‘A note on three-valued modal logic’, *Notre Dame Journal of Formal Logic* **19**, 63–68.

Wilkes, K. V. (1978), ‘Consciousness and commissurotomy’, *Philosophy* **53**, 185–199.

Zangwill, O. L. (1974), Consciousness and the cerebral hemispheres, *in* S. J. Dimond & J. G. Beaumont, eds, ‘Hemisphere Function in the Human Brain’, Paul Elek & John Wiley, London & New York.

December 5, 1997