VAN FRAASSEN AND THE PRAGMATIC THEORY OF EXPLANATION

FABRIZIO CARIANI

1. Preliminaries

Main Thesis: Explanations are answers to why-questions.
Hence, a theory of explanation is essentially a theory of why-questions.
According to Van Fraassen, formulating a theory of explanation poses two main problems:

(1) Providing an account of legitimate rejections of explanation-requests.
(2) Accounting for the asymmetries of explanation: why, in some cases, are we inclined to say that given two facts one explains the other, but not the converse? (e.g. the tower and the shade).

On VF's theory, why-questions and their answers are individuated only relative to a context. The point of construing a contextual analysis of explanation is to put the concept of explanation on grounds that are respectable from an empiricist standpoint. VF takes the "received view" as characterizing explanation as a relationship between theory and fact, and as ending up being committed to the existence of causal relationships beyond the observable phenomena. VF's attempt on the other hand is to solve problems (1) and (2) without invoking irreducibly counterfactual or modal elements. To this end, he takes explanation as a ternary relationship that holds between theories, facts and contexts.

2. The Formal Pragmatics of Why Questions

2.1. **Identifying why-questions.** Consider the stock example:
(*) Why did Adam eat the apple?
The why-question asked by (*) is determined by:

(1) *A Topic*: a proposition expressing the fact whose explanation we are asking. (e.g., in the case of (*), that Adam ate the apple).
Neither having the same topic nor being expressed by the same sentence constitutes a criterion of identity for why-questions. In other words, the same sentence can express different why-questions. For VF, there are various ways in which this can happen. This leads him to bring in two more contextual factors:

(2) *A Contrast Class*: a set of propositions, including the topic, that determines the range of alternatives against which a why-question is asked. E.g. Why did *Adam* eat the apple (as opposed to Mary), why did Adam *eat* the apple (as opposed to, say throwing it away because it was old), why did Adam eat *the apple* (as opposed to the oranges, the grapes... etc.). .

(3) *A Relevance Relation*: "a respect-in-which a reason is requested".
An example of how (3) discriminates among different why-questions: consider the question "Why was Mr. Jones killed by his wife?". On a natural understanding, it is irrelevant to

answer "Because she sticked a knife in his throat, thereby causing such and such events that lead to Mr. Jones's death".

2.2. **Answers to why-questions.** Given a why-question $Q = \langle P_k, X, R \rangle$ an answer looks like:

(#) P_k in contrast to the rest of X, because A.

The account of answers to why-questions is obtained by analyzing (#). Here, A is called *the core* of answer (#) to why-question Q. According to VF, in such an answer we claim:

(1) P_k is true.
(2) Every other element of the contrast class X is false.
(3) A is true.
(4) A bears R to $\langle P_k, X \rangle$.

On his account of answers, Van Fraassen builds his account of the presuppositions of why-questions. In turn, with the latter account, he claims to have solved the first of the problems of explanation (legitimate rejections): a why-question can be rejected if its topic is false, if any other element of the contrast class is true, or if there is no proposition bearing the relevance relation to the ordered pair formed by topic and contrast class [i.e. $\forall A \neg R(A, \langle P_k, X \rangle)$].

3. SALMON & KITCHER 1

In their paper S&K argue that Van Fraassen's theory of explanation (a) does not impose enough constraints on the relevance relation R. The moral that they want to draw is that (b) the conditions that characterize why-questions and answers to WQ must be supplemented by the further condition that R actually be a relevance relation; however, (c) accepting this condition requires us to make a non-circular distinction between relevance relations and other relations and (d) actually making this distinction amounts to facing very basic problems in the theory of explanation. Conclusion: VF has not offered a satisfactory theory of explanation.

Here we will focus on (a): S&K's argument to show that R is not enough constrained. They argue that this looseness of characterization of R leads VF to accept as admissible why-questions that should, intuitively, be rejected. A stronger claim, which S&K make anyway, is that on the theory sketched above, basically any proposition can count as the core of an answer to a why question.

Let's focus on S&K's main example, before introducing VF's theory of evaluating answers.
P_k=JFK died 11/22/63
X={JFK died 1/1/63, JFK died 1/2/63,..., JFK died 11/22/63, JFK died 12/31/63, JFK survived 1963}
R=Astral Influence.
So $Q = \langle P_k, X, R \rangle$ Let the answer be like our (#) above: P_k in contrast to the rest of X because A, where A is a *true* description of the positions of the stars and planets at the time of JFK's birth.

So $Q = \langle P_k, X, R \rangle$ qualifies as a genuine explanation-request by VF's criteria. Is this what we want to say intuitively about it? S&K say it is not. [1] There are two sorts of issues here:

- one is whether VF's model wrongly classifies as legitimate explanation requests to which no explanatory answer is forthcoming

---

[1]They write: "In the context of twentieth-century science, the appropriate response to the question is rejection. According to our present lights, astral influence is not a relevance relation. We believe that the positions of the stars and planets on JFK's birthday have no effect on the probability of death on any particular day".

- the other, is whether VF's criteria for evaluating answers as good or bad (which I haven't discussed yet) classify as good explanations which in fact are not such.

Q1: Is the JFK-example a real problem for Van Fraassen? Suppose he were committed to a "anything-goes" theory of explanation. Would this conflict with our intuitions about explanation (as long as he keeps apart explanations and *good* explanations)? After all, part of the point of developing the theory of why-questions using formal-pragmatics was to capture, as it were, "formal" conditions on the practice of posing and answering such questions. Once questions of evaluation are left aside for separate treatment, the "anything goes" theory does not create many problems.

The key question is how examples like the JFK-case are treated in VF's theory of evaluating answers. It will be a problem if this theory classifies such cases as good explanations. But does it?

4. EVALUATING ANSWERS

Van Fraassen's remarks

Given a question $Q = \langle B, X, R \rangle$, background knowledge K, and an answer with core A, Van Fraassen gives three criteria that govern our evaluations of answers to Q as good or bad (he does not want here to specify the theory in complete detail: e.g. we are not told precisely how these criteria are to be ranked to provide an ordinal structure of answers to a given Why-Question).

(i) The probability of A (on the basis of K) is high.

(ii) A together with a relevant subset $K(Q)$ of K, favors the topic (i.e. B) over the other members of the contrast class.[2].

(iii) A compares well to other answers to Q. This comes in three sub-criteria : (1) A has higher probability than other answers, (2) A more strongly favors B over the other members of X (than other answers), and (3) A is not made irrelevant by other possible answers.

Although these criteria are not spelled out in detail, we can extract from them a notion of maximally good answer: if A belongs to K and $K(Q)$ plus A implies B and implies the falsity of $C, ..., N$ then A is a maximally good answer to Q.

Branden's proposed reading of the criteria

Let $S = \{A_0, ..., A_n\}$ be a class of answers to $\langle B, X, R \rangle$ against background knowledge K. Let A be a member of S. Let $f(B, X_n, A \mid K)$ be the degree to which A favors B over X_n, given K. Also, let $Z = \{Z_0, ..., Z_n\}$ be a logical partition of explanatorily relevant alternatives. VF's criteria can then be interpreted as:

(1) $Pr(A|K)$ is high and $Pr(A|K) \geqslant Pr(A_i|K)$ for all i, $0 \leqslant i \leqslant j$.

(2) $f(B, X', A|K(Q)) \geqslant 0$ for all $X' \in X$ with strict greater than relation holding for some X' s.t. $X' \neq B$

(3) $f(B, X', A|K(Q)) \geqslant f(B, X', A'|K(Q))$ for all $X' \in X$ s.t. $X' \neq B$ and all $A' \in S$ s.t. $A' \neq A$.

(4) $f(B, X', A|A'\&K(Q)) \neq 0$, for all $X' \neq B$ and all $A' \neq A$.

(5) $\forall Z' \in Z, f(B, X', A|K(Q)\&Z') > 0 \Rightarrow f(B, X', A|K(Q)) > 0$.[3]

---

[2]Requesting that A by itself favor B over the rest of X seems too weak. On the other hand, taking the whole set K trivializes the criterion in the cases where the topic is already part of our background knowledge. That is why VF speaks of a subset of K, which he calls $K(Q)$ and which is contextually determined.

[3]A technical comment: although "officially" VF does not want to pick any particular measure of favoring, his criteria impose quite restrictive conditions on admissible measures. In particular the range is restricted by the property called *Collapsibility* (listed as (5) above); VF's reasons for invoking it have to do with his treatment of Simpson's Paradox. Among the measures that satisfy *Collapsibility*, Branden suggests using the *Relative Risk* measure: $f(B, X', A|K(Q)) = \frac{Pr(B|A\&K(Q))}{Pr(B|\sim A\&K(Q))} - \frac{Pr(X'|A\&K(Q))}{Pr(X'|\sim A\&K(Q))}$.

5. Salmon & Kitcher 2

What is the significance of the JFK example with respect to VF's theory of Evaluating Answers? Salmon says that (1).1 is met because we can perfectly reconstruct the position of the stars, and other celestial bodies at the time of JFK's birth; equally (1).2 follows because, if A is certain, then no proposition could be *more* certain. To show that the other criteria are met, he has to make a further supposition:

(S) "The astrologer can derive from A by means of astrological theory, that JFK was more likely to die on 11/22 rather than any other day of 1963, and also that JFK was more likely to die on 11/22 than to survive 1963".

This assumption yields the conclusion that (2) A favors B on any element of X_n,(3) that A favors B at least as much as any alternative explanation (because in fact $A \Rightarrow B$), and (4) nothing could make A irrelevant.

Q2: Is assumption (S) justified? For Salmon, a reason to assume (S) is the fact that the astrologer is just *accomodating* data: his explanation comes *ex post facto*. This seems like a shift in the issue. Let's say that A T-implies $\langle B, X \rangle$ iff $A \models B$ and $\forall P \in X(P \neq B \Rightarrow A \nvDash P)$. What sort of 'derivation' does the astrologer carry? It seems one thing to say that VF's criteria give highest marks to answers that T-imply $\langle B, X \rangle$, and a quite different thing to say that *any answer* can be construed as T-implying $\langle B, X \rangle$. Intuitively, some factors about the characterization of good vs. bad explanations have to be non-subjective. How exactly can we make this intuition into an argument?

Looking closely, S&K put their fingers on a problem shared by every theory based on statistical relevance: in general, B is going to be a perfect match of statistical-relevance criteria for B itself. This problem is going to affect Salmon's S-R account as well. However, Salmon handles the JFK example by saying that B is not causally relevant to B (causality is irreflexive, after all). Can this problem be fixed in an empiricist framework?

One suggestion (Branden's 2001 suggestion) to rule out (S) is to build into the characterization of explanation a *predictive* criterion. *Predictivism* is the view that pieces of evidence that are already available when a theory is formulated do not confirm the theory as strongly as evidence successfully predicted by the theory does (Maher's coin example). From the slides:

- A scientific answer A to a why-question Q is a *prediction* of P_k (*vs.* P_j) made by (or derived from) a scientific theory T, together with background knowledge K (circumscribed by context).
- (CR) An answer A to Q is "better than" a collection of alternative answers A_i if T *better predicts* P_k (*vs.* P_j) than each T_i does, relative to K.

This is an attractive view, and it easily rules out cases like the JFK/astrology example. However there are some problems:

Q3 Does (CR) bring into the picture a direction of the prediction/explanation symmetry? If so, can we defend such a principle in full generality? Does it follow that there is no scientifically good explanation for unpredictable events?

Q4 Does speaking of scientific answers violate VF's point that scientific explanation is not a *different sort* of explanation?

There is also a more general problem, which becomes more urgent when we add (CR) to our list of evaluative criteria:

Q5 Which evaluative criteria prevail over others? Should we put particular weight on (CR)?