

THE FUNCTIONS OF INTENTIONAL EXPLANATIONS OF ACTIONS

Erik Weber
Robrecht Vanderbeeken
Ghent University (UGent)

ABSTRACT: This paper deals with the *functions of intentional explanations* of actions (IEAs), i.e., explanations that refer to intentional states (beliefs, desires, etc.) of the agent. IEAs can have different formats. We consider these different formats to be instruments that enable the explainer to capture different kinds of information. We pick out two specific formats, i.e. *contrastive* and *descriptive*, which will enable us to discuss the functions of IEAs. In many cases the explanation is contrastive, i.e. it makes use of one or more contrasts between real intentional states and ideal intentional states (ideal from the point of view of the explainer). In many other cases IEAs have a descriptive (covering-law) format. The aim of this paper is to analyze the functions the two kinds of explanations can have. We will show that certain functions are better served by one rather than the other format. This leads to pluralism with respect to formats. We argue that both formats are necessary and that their functions are complementary.

Key words: actions, contrastive explanations, intentional explanations

Actions as we conceive them transform an indeterminate situation into a determinate one. This terminology is taken from John Dewey. An indeterminate situation is a situation in which we know that *something* must be done, but we do not know *what* to do. A determinate situation is a situation in which we know what to do. By limiting ourselves to actions defined in this way, we exclude emotional and habitual behavior from the scope of this paper. However, since no restrictions are posed on the decision procedures by which the indeterminate situation is transformed into a determinate one (e.g., flipping a coin is allowed, as well as maximizing expected utility) our analysis is *not* restricted to actions that result from procedures in which the possible consequences of each action are considered. When the agent flips a coin, his choice is made without considering the outcomes of possible actions. When the agent maximizes expected utility, the possible outcomes of actions do influence decisions. We want to include both cases in which an outcome-neglecting rule (such as flipping a coin) is used and cases in which an outcome-oriented rule (such as expected utility maximizing) is used.

AUTHORS' NOTE: We thank the Fund for Scientific Research – Flanders for their financial support through project nr. G.00015.99. We also thank John Staddon and two anonymous referees of this journal for their comments on the first draft of this paper. Please send all correspondence to Erik Weber or Rob Vanderbeeken, Centre for Logic and Philosophy of Science, Universiteit Gent, Blandijnberg 2, B-9000 Gent, Belgium. Email: Erik.Weber@UGent.be; Robrecht.Vanderbeeken@UGent.be

The literature on intentional explanations of action (i.e., explanations that refer to intentional states—beliefs, desires, etc.—of the agent) can be characterized as intense inquiry into the appropriate format and model of intentional explanations. This inquiry is motivated by the common but problematic assumption that there is one specific format that can do all the work (and that we should look for that format). We take a more pragmatic approach. Models and formats of explanations are instruments and it is their use that legitimizes them. Why should we restrict ourselves to one instrument if it turns out that we can use different instruments that can serve different functions? Hence, the claim we will be arguing for is that intentional explanations of actions (IEAs for short) can have different formats that serve different functions.

We discuss two specific formats and aim to point out the difference in functions these formats of IEAs can serve. In many cases, the explanation makes use of one or more contrasts between the real intentional states and ideal intentional states (ideal from the point of view of the explainer). Explanations of this type will be labeled *contrastive* IEAs. In the first section of this paper we attempt to clarify the *structure* of such explanations. In the second section we investigate the *functions* of contrastive IEAs. The first section has a preparatory status: it is impossible to discuss the functions without some insight in the structure.

In many other cases IEAs are purely descriptive; they aim to give a list of explanatory factors that are involved in the origination of the action to be explained and establish a nomological link between these causal factors and the effect (the action). The *structure* of such explanations will be clarified in the third section. In the fourth section we discuss the *functions* of descriptive explanations of actions. Again, the third section has a preparatory status.

The joint conclusion of these sections will be that contrastive explanations serve some functions better than descriptive ones, and vice versa. Hence, we conclude in the fifth section that we have to be pluralists about the format of IEAs. In order to avoid misunderstanding of what this pluralism means, the following analogy is useful. In court we are supposed to tell the truth, the whole truth, and nothing but the truth. Our pluralism does not imply “anything goes”: explanations must not be false; they should contain “nothing but the truth.” However, explanations do not always contain “the whole truth”; they contain the part of the truth that is relevant in the context. The fact that we do not need the whole truth to answer a *why* question gives rise to the possibility of different formats.

There is a long-standing controversy about the validity of intentional explanations. Many readers of this journal are presumably skeptical about such explanations. Obviously, we do not share this skepticism; we think that there are cases in which intentional explanations are appropriate, even indispensable. However, we certainly do not believe that intentional explanations are the only proper explanations, or superior to other types. We are also pluralist with respect to the *theories* on which explanations are based, not only with respect to their format. This second aspect of our pluralism will be clarified and defended in the last section of this paper.

Contrastive Intentional Explanations of Actions

The aim of this section is to get some insight into the structure of explanations that play on the *discrepancies* between the agent's actual decision-forming process and the ideal pattern we (i.e., the explainers) have in mind. These discrepancies can be spelled out by first determining the ideal pattern (this part of the explanation is called the *normative* part) and then point at the differences (this part is called the *differentiating* part). We start with some examples, then we give a general characterization of the normative and differentiating part and discuss some important properties.

Example 1

Consider a game with the following setup. An urn contains 80 red balls and 20 green ones. On the left of the urn is a green box and on the right is a red one. The game master draws a ball randomly and gives it to the blindfolded candidate who must put it in one of the boxes (the candidate knows the color/position of the target boxes, but does not know the color of the ball). The rules of the game are:

If your ball is green and you put it in the green box, you receive €100.

If your ball is green and you put it in the red box, you receive €10.

If your ball is red and you put it in the green box, you receive €10.

If your ball is red and you put it in the red box, you receive €50.

The ideal pattern we have in mind could be:

Desires: The more money, the better.

Beliefs about Circumstances: The candidate knows that the urn contains 80 red balls and 20 green ones.

Belief about Causal Relations: The candidate knows the rules of the game.

Beliefs about Opportunities: The candidate knows that there are two options (the red box on the right and the green box on the left).

Decision Principle: Always maximize expected utility.

This is the normative part of the explanation. It is easy to calculate that a candidate satisfying this ideal pattern normally chooses the green box (expected utility of €82, as opposed to €18). If the candidate chooses the red box, we can point at the differences between his real state of mind and the ideal pattern (= differentiating part) in order to explain his "deviant behavior." Maybe the candidate did not know the rules, or did not see the opportunities correctly, or did not want to get as much money as possible, or. . .

Example 2

A second example of a normative part is:

Beliefs about Circumstances: s should believe that s' has done a towards s .

Beliefs about Opportunities: The opportunities which s should see are doing a or not doing a towards s' .

Decision Principle: s should make up his mind by means of the rule “If people do a to you and you can take revenge by doing a to them, do it.”

Here we have an example of an outcome-neglecting decision rule (unlike the rule in example 1). According to this rule s should take revenge no matter what his desires and his causal beliefs with respect to the effects of revenge are. For instance, the agent should not bother about a possible escalation of the conflict, which might cause him further harm.

If s does not take revenge we can again explain this “deviant behavior” by pointing at the differences between the ideal pattern and the real state of mind of s . For instance, we might point out that s does not follow the primitive revenge rule described above.

General Characterization

In general, we can characterize the normative part as a specification of:

- (1) the *opportunities* the agent should see
- (2) the *decision principle* he should use
- (3) the beliefs he should have about the *circumstances*

This is the *minimal* content of the normative part, and it is illustrated by example 2. If the decision rule requires beliefs about causal relations and desires as inputs (as in example 1), the normative part must also specify the presumed ideal causal beliefs and desires.

The part of the explanation that specifies the differences between the real state of mind and the ideal pattern put forward in the normative part is called the *differentiating part*. We distinguish between two contrastive formats. The first is:

$$(A) \quad s \text{ does } a_r \text{ rather than } a_n \text{ because of } D_1, \dots, D_m$$

a_r is the real action; a_n is the action that the person who asks the question considers normal. D_1, \dots, D_m are discrepancies between the agent’s mental states and the mental states described in the normative part of the explanation (cf. the differentiating parts in the examples). The second format is:

FUNCTIONS OF INTENTIONAL EXPLANATIONS OF ACTIONS

(B) *s* does a_r rather than a_n (though he intended to do a_n) because of abnormal obstructing factors F_1, \dots, F_m .

Format (A) presupposes that the agent's real action is identical to the action the agent preferred (a_p). a_r and a_p can differ for a number of reasons—an event (e.g., some accident or an action performed by another person) physically inhibiting the agent, an event or person who distracts the agent so that he neglects time, an unforeseen emotional obstruction (e.g., shyness, fear), a cognitive malfunction (e.g., amnesia), bad coordination of bodily movements (e.g., anarchic hand syndrome), etc. In such cases format (B) will be appropriate.

Properties of Contrastive IEAs

Though our characterization of the normative part is liberal with respect to the decision rule that is used (and, consequently, with respect to the types of inputs that are involved), and though we distinguish two contrastive formats for the differentiating part, we do not want to argue that all contrastive IEAs fit our characterization. However, we hope that the reader will share our view that scientists and ordinary people often formulate contrastive IEAs that *do* fit our characterization. This means that the analysis is a good starting point for clarifying the functions of contrastive IEAs; maybe it is not exhaustive, but it characterizes a vast majority of contrastive IEAs.

Contrastive IEAs have two properties that are important to distinguish them from the descriptive ones to be discussed later. First, they are almost always *incomplete* in the sense that they do not refer to all the causes of the action. They pick out one or a few causes because they deviate from what the explainer considers normal. The explanation remains silent about the other causes. For instance, if the candidate in our game does not know the causal relations, while his state of mind fits the ideal pattern in all other respects, the explanations will simply state “He chose the red box because he got the causal relations wrong.” All other causal factors of the action are neglected. In other words, the explanation picks out a part of the truth (cf. the analogy in the introduction). Only in the exceptional case, in which the real state of mind is in all respects different from the ideal pattern, will the contrastive explanation refer to all causal factors.

The second important property is that contrastive IEAs have a normative dimension, in which two closely related aspects can be distinguished. First, they all contain a normative part—but there is always a second normative aspect. The discrepancies mentioned in differentiating parts of type (A) have the form “*s* is in M and should be in M' ,” where M and M' are two mutually exclusive mental states—so what comes after the “because” in format (A) is not purely descriptive. In differentiating parts of type (B) there is an analogous normative aspect: the obstructing factors are abnormal.

The Functions of Contrastive IEAs

We start our analysis of the functions of contrastive IEAs with an example. Consider a scientist who has decided that he and the research group he is directing will cooperate in the development of weapons of mass destruction. An explanation of the decision of the scientist can be useful for several reasons:

- (a) The explanation helps us determine our moral judgment about the decision (i.e., was the scientist's decision morally justified?).
- (b) Assuming that withdrawal is still possible, the explanation might teach us how to make the scientist withdraw.
- (c) The explanation might teach us how we can prevent the same or other scientists volunteering for similar research programs in the future.

We will now discuss these functions in detail. The material for the example we use is taken from Lackey (1994). The aim is to show that contrastive IEAs as we have defined them can serve these different functions.

The First Function

Lackey considers a scientist leading a group of researchers who decides to take funds budgeted through the Strategic Defense Initiative (SDI). According to Lackey, three cases can be distinguished:

- (1) The scientist thinks his research will not serve any military purpose because the goals of SDI cannot be achieved.
- (2) The scientist believes that his research will serve some military purpose, and that this purpose is evil.
- (3) The scientist believes that his research will serve some military purpose, and that this purpose is benign.

In all three cases there is a discrepancy between the actual preference-forming process of the scientist and the presumed ideal pattern. The latter may be characterized as follows¹:

- (1') The scientist should know that research toward unattainable goals is often diverted toward other military purposes (a causal belief).

¹ We are fully aware that not everyone will agree that this is the ideal pattern; however, this is the ideal pattern on which Lackey—the author from whom we took our example—bases his contrastive explanation.

FUNCTIONS OF INTENTIONAL EXPLANATIONS OF ACTIONS

(2') The scientist should be ready to avoid evil (moral principle).

(3') The scientist should know that the use of force by the United States since the Korean War has been mostly motivated by economic and geopolitical self-interest (factual belief).

In case (2), where the discrepancy is that the scientist does not conform to (2'), our judgment is clear: the act of the scientist is immoral. In the other cases, where the scientist has false beliefs so that (1') or (3') is not satisfied, our judgment depends on whether the ignorance is acceptable. These cases are less straightforward; they depend on the minimal effort in information gathering that we require for a moral decision. The discrepancies revealed by the explanations—together with other factors such as what is acceptable ignorance and what is culpable ignorance—determine our moral judgment.

The example shows that contrastive IEAs with a differentiating part of type (A) can have a *judgment-determining* function. In our example the judgment is of a moral order. The question to be decided is whether the action was morally justified or not. In other cases the question may be of a legal order (e.g., is this person guilty of murder or not?), or may relate to status (e.g., should president X resign from his job for what he has done?). The answer to such questions also depends on the discrepancies between the real world and the ideal pattern we (i.e., the explainers) have in mind.

Contrastive IEAs with a differentiating part of type (B) can have the same function. This can be illustrated by a variation on our example. Suppose that our scientist does not work in the SDI program, but for a dictatorial regime. He is convinced that developing arms for this regime is wrong, but he will be deported to a labor camp if he does not cooperate. This coercion is ethically relevant because it may be regarded as an extenuating circumstance.

The Second and Third Function

With respect to the second function (withdrawal), it may be hard to change someone's moral principles (second discrepancy), but changing his factual beliefs may be easier (when they are mistaken). To the extent that such beliefs are typical for scientists working with military funds, explaining the decision of one research group can provide a strategy for avoiding similar actions in comparable groups (cf. third function). According to Lackey, historical ignorance (third discrepancy) is typical for groups working with military funds:

It has not been my impression that many scientists involved in defense work or subsisting on the defense dole have undertaken an examination of the historical evidence and the present world scene. On the contrary, one finds, given the educational attainments of these people, a surprising ignorance of history, and an incredible lack of exposure to alternative interpretations of what happened and why. William Broad's account of the political naiveté of young scientists at the Lawrence Livermore National Laboratory describes not the odd case, but the

typical ignorance of history and the humanities one finds among these researchers. (p. 404)

The study referred to is Broad (1985). The example shows that contrastive IEAs with a differentiating part of type (A) can have a *therapeutic* function (they diagnose what went wrong and thus help us to restore an ideal state) or a *preventive* function (they help us to avoid similar actions by other people on future occasions). This can be extended to contrastive IEAs with a differentiating part of type (B). If we know, for instance, that a scientist is threatened with forced labor, we can only change his behavior by removing this threat. If such threats are common, removing or preventing such threats is an adequate strategy for preventing similar behavior by other scientists.

Descriptive Explanations of Actions

The Structure of Descriptive IEAs

The aim of this section of to get some insight into the structure of descriptive explanations of actions (explanations that give causes and provide a nomological link between the causes and the action). There is a long tradition, starting with Hempel himself, of attempts to fit explanations of actions in the Hempelian covering-law model. Hempel (1965) distinguishes two types of covering-law explanations of particular facts: deductive-nomological (DN) and inductive-statistical (IS). The DN format is the ideal one and looks like this:

$$\begin{array}{ll}
 C_1: & P_1a \\
 C_2: & P_2a \\
 \dots & \\
 C_m: & P_ma \\
 L: & (\forall x)[(P_1x \wedge P_2x \wedge \dots \wedge P_mx) \Rightarrow Qx] \\
 E: & \hline & Qa
 \end{array}$$

The explanandum must follow deductively from the premises, but it should not follow from C_1, \dots, C_m alone.

A classical objection against the DN model applied to intentional explanations is that there are no strict reason-to-action laws (cf. Davidson, 1980). Actions are very complex; all sorts of intentional states and other factors play a positive or negative role in the origination of an action. We have given examples of this in the section on the general characterization of contrastive IEAs. Having an intention is, in general, not sufficient for acting according to this intention because of external obstructions (e.g., someone or something inhibiting the action) or internal obstructions (e.g., insufficient control over our body). This brings us to what we can call *the problem of causal closure of actions*. It is extremely difficult to give a set of causes that is sufficient for an action because this set should include the

FUNCTIONS OF INTENTIONAL EXPLANATIONS OF ACTIONS

absence of all possible obstructing factors. This means that DN explanations of actions are quite impossible.

The problem can be solved by requiring that covering-law explanations of action adopt what we call the *default rule model*. In cases where there are no obstructing factors (cf. the list of possible obstructing factors mentioned in the general characterization of contrastive IEAs), a covering-law explanation must fit the following pattern:

- (I)
- | | |
|------------------|---|
| C ₁ : | P ₁ s |
| C ₂ : | P ₂ s |
| | ... |
| C _m : | P _m s |
| L: | ($\forall x$)[(P ₁ x \wedge P ₂ x \wedge ... \wedge P _m x) \rightarrow x usually prefers a]
 [usually] |
| E: | s prefers a. |
| | |
| C': | s prefers a. |
| L': | ($\forall x$)(x prefers a \rightarrow s usually does a)
 [usually] |
| E': | s does a. |

P₁ . . . P_m describe the decision rule, the beliefs about opportunities and about circumstances, and (if they are relevant) causal beliefs and desires. In order to understand this pattern, four remarks are useful:

(1) Format (I) looks like a sequence of two DN explanations, with one important difference: L and L' are default rules, not universal generalizations as required in the DN model. Default rules (e.g., “birds usually fly”) differ from universal generalizations in that they allow exceptions (e.g., “penguins do not fly”). This means that we can try to be as complete as possible about the representation of the motivation of the agent and at the same time respect the problem of causal closure of actions. We aim at completeness but do not claim that we have succeeded. Note that we do not want to consider here the possibility of *ceteris paribus* laws (hedged laws) due to some classical problems (for a discussion see Vanderbeeken, 2004).

(2) Default rules also differ from probability statements in that they do not specify the relative frequency of exceptions and “normal” cases (“usually” can mean anything fairly close to probability 1). If we would specify the relative frequency we would obtain inductive-statistical (IS) explanations in Hempel’s sense. However, the information required to do this is usually not available. So for explanations of actions, IS explanations are not a viable alternative to DN explanations. Variants of IS explanations that use probabilistic laws with intervals (e.g., “at least 95% of birds fly”) are also impossible in many cases because we cannot determine the lower limit of the interval. The vagueness of the default rules is unavoidable when formulating reason-to-action rules. As a consequence, descriptive explanations of actions fit our default rule model, not the IS model.

(3) The default rules are not analytical truths. On the one hand, the rule

$$(\forall x) (x \text{ prefers } a \rightarrow s \text{ usually does } a)$$

is contingent on the fact that in most cases there are no obstructing factors. If, for instance, from tomorrow on most people have amnesia or constantly forget about time, it will no longer be valid. On the other hand, rules of the form $(\forall x)[(P_1x \wedge P_2x \wedge \dots \wedge P_mx) \rightarrow x \text{ usually prefers } a]$ are only valid if most people apply the decision rule (which is described in one of the P_i s) correctly.

(4) Hempel prohibited the use of accidental generalizations in DN explanations. If we shift from exceptionless generalizations to default rules, the requirement of non-accidentality must be retained. This can be done by requiring that the default rules meet the invariance condition proposed in Woodward (2001) or the spatio-temporal stability condition proposed in Mitchell (2000).

In the second case (in which obstructing factors occur) the general pattern is:

(II)	C ₁ :	P ₁ s
	C ₂ :	P ₂ s
		...
	C _m :	P _m s
	L:	$(\forall x)[(P_1x \wedge P_2x \wedge \dots \wedge P_mx) \Rightarrow x \text{ usually prefers } a]$
		===== [usually]
	E:	$s \text{ prefers } a.$
	C ₁ ':	$s \text{ prefers } a.$
	C ₂ ':	Obstructing factor f occurs.
	L':	$(\forall x)[(x \text{ prefers } a \text{ and obstructing factor } f \text{ occurs}) \rightarrow s \text{ usually does } a'].$
		===== [usually]
	E':	$s \text{ does } a'.$

Again, L and L' are contingent but non-accidental default rules. As in format (I) this entails that explanations are not deductively valid arguments. The conclusion follows with high but unspecified probability (i.e., “usually”) from the premises.

Structural Differences between Contrastive and Descriptive IEAs

As in the first section, we have been very liberal in our characterization; we allow for complex combinations of goals, obstructing factors may occur, and the covering-laws can be default rules. Nevertheless, we aim for—but do not claim—completeness. We hope to have convinced the reader that covering-law explanations as we conceive them (with non-accidental default rules, not with exceptionless deductive laws or probabilistic laws) are possible.

FUNCTIONS OF INTENTIONAL EXPLANATIONS OF ACTIONS

At the end of the first section we mentioned two important properties of contrastive IEAs. The first was their incompleteness. Descriptive IEAs aim at completeness (without necessarily achieving it). This difference implies that a contrastive IEA cannot be transformed into a descriptive one without adding new information. Why is this so? As one of the referees of this journal has pointed out, we can start from the contrastive IEA

$$s \text{ does } a_r \text{ rather than } a_n \text{ because of } D_1. . . D_m$$

and try to recast it into a covering-law structure²:

$$\begin{array}{ll}
 C_1: & D_1 \\
 C_2: & D_2 \\
 & \dots \\
 C_m: & D_m \\
 L: & (\forall x)[(D_1x \wedge D_2x \wedge \dots \wedge D_mx) \rightarrow (x \text{ usually does } a_r \wedge x \text{ usually} \\
 & \text{does not do } a_n) \\
 & \text{===== [usually]} \\
 E: & s \text{ does } a_r \wedge s \text{ does not do } a_n
 \end{array}$$

The discrepancies stated as initial conditions would have the form “*s* is in *M* and should be in *M'*,” where *M* and *M'* are two mutually exclusive mental states. Such transformation will only be possible in exceptional cases in which the mental state of *s* is completely different from the ideal. In all other cases the descriptive explanation requires premises of the form “*s* is in *M* and should be in *M'*,” where *M* and *M'* are *identical* rather than *mutually exclusive*. In the descriptive explanation we also need the points of agreement, not only the discrepancies.

In the first section we also mentioned that contrastive IEAs have a double normative dimension. Because of this dimension, descriptive IEAs cannot be transformed into contrastive ones without new information. This can be clarified by an example that was brought up by one of the referees. According to Aristotle, singular “because” sentences like “Socrates died because he was made to drink hemlock” are elliptical formulations of syllogisms (in modern terminology, elliptical formulations of DN structures). In the Socrates case the syllogism is: “People who drink hemlock die. Socrates was made to drink hemlock. Therefore Socrates died.” It is important to see that even if Aristotle is right this does not entail that contrastive IEAs are elliptical for DN explanations because they cannot be identified with simple “because” sentences. They *contain* a “because” sentence (they also contain a normative part), and this sentences must have a specific format (referring to discrepancies or abnormal obstructing factors) giving it a normative dimension.

To sum up, contrastive IEAs provide some of the causes without attempting to be complete (they tell “part of the truth”), whereas descriptive IEAs aim at causal

² To abbreviate, the intermediate step from preference to action is omitted.

completeness (they try to tell “the whole truth”). Descriptive IEAs give only information about causes (they tell “nothing but the truth”), whereas contrastive IEAs have a double-normative dimension (they tell “more than the truth”).

The Functions of Descriptive IEAs

Descriptive IEAs and the Functions of Contrastive Ones

That descriptive IEAs cannot have a judgment-determining function is obvious; the normative part is essential for that function. We believe that they are also not very well suited for therapy and prevention. To clarify this, we take Robert Axelrod’s analysis of unofficial truces (Axelrod, 1984; discussed in Little, 1991, p. 58) as an example. In World War I there were unofficial truces by units on both sides; each side continued to fire its weapons but without inflicting much damage. Axelrod explains these truces as rational behavior based on a strategy of conditional cooperation in a repeated prisoner’s dilemma situation (this strategy amounts to: start with cooperation and keep on cooperating as long as the opponent cooperates). A relevant circumstance of choice is *trench warfare*, which guarantees a relatively stable, clearly identifiable enemy (units are not replaced overnight) whose reactions can be easily observed. The underlying idea is that in different types of warfare (e.g., blitzkrieg, guerrilla) similar truces are impossible because there is no stable enemy.

In this example a truce is explained as the aggregate of result of the behavior of two units. The behavior of each unit can be explained as follows:

- C₁: Unit *a* adopts a conditional cooperation strategy.
- C₂: Unit *a* considers unit *b* to be its relatively stable enemy.
- C₃: Unit *a* can observe the reactions of unit *b*.
- L: If a unit *x* adopts the conditional cooperation strategy, considers *y* to be its stable enemy, and can observe the reaction of *y*, unit *x* usually prefers to fire its arms at *y* without inflicting any damage.
===== [usually]
- E: Unit *a* prefers to fire its arms at *b* without inflicting any damage.
- C': Unit *a* prefers to fire its arms at *b* without inflicting any damage.
- L': If a unit *x* prefers to fire its arms at *y* without inflicting any damage, the unit usually does this.
===== [usually]
- E': Unit *a* fires its arms at *b* without inflicting any damage.

This explanation fits pattern (I) of the previous section. The example nicely illustrates the limitations of covering-law explanations of actions. They are poor instruments from the point of view of therapy and prevention because they are not specific enough. The reason is that the explanation cites beliefs about causal relations which we want to retain in the ideal situation; we want the units to fight

FUNCTIONS OF INTENTIONAL EXPLANATIONS OF ACTIONS

in the trenches even when they know their enemies well. A contrastive explanation of why the units fire badly rather than try to destroy the enemy's trenches would focus on the conditional cooperation strategy—this strategy is the malefactor. Because covering-law explanations must aim at completeness (leaving out relevant causes can make the derivation relation collapse), this is not an idiosyncratic feature of our example. In covering-law explanations the causes that are relevant from a therapeutic and preventive point of view are always put on the same line with causes that are irrelevant from those points of view. The result is that we cannot see the wood for the trees.

This limitation of covering-law explanations is not unique for explanations of actions. Peter Lipton gives a nice example:

Suppose that my car is belching thick black smoke. Wishing to correct the situation, I naturally ask why this is happening. Now imagine that God (or perhaps an evil genius) presents me with a full Deductive-Nomological explanation of the smoke. This may not be of much help. The problem is that many of the causes of the smoke are also causes of the car's normal operation. Were I to eliminate one of these, I might only succeed in making the engine inoperable. By contrast, an explanation of why the car is smoking rather than running normally is far more likely to meet my diagnostic needs. (Lipton, 1993, p. 53)

Two Functions of Descriptive IEAs

Thus far we have only discussed the function covering-law explanations do not have or serve less well than contrastive ones. It is important to see that they can provide us with information that contrastive explanations do not give us. Axelrod's explanation helps us to decide whether or not to expect similar truces in other contexts, such as a blitzkrieg or a guerrilla war. As already indicated, his explanation entails that similar truces should not be expected in those circumstances. In our view, this is the first function of covering-law explanations of actions: providing information that is relevant for predictions in identical and similar situations in the future.

A second function of covering-law explanations of actions is that they enable us to make sense of the action. Knowing the motivation of the units, we can see them as reasonable actors and imagine that we would act in the same way. Since such explanations aim for completeness, they are good instruments for understanding the particular action at hand.

Summary and General View on Explanations

Summary

The results of the previous sections can be summarized as follows:

(1) Contrastive IEAs can serve at least three functions in a proper way: sometimes they co-determine our moral and legal *judgments*; sometimes they

diagnose what went wrong and thus help us to *restore* an ideal state (therapeutic function); and sometimes they help us *avoid* similar actions by other people on future occasions (prevention).

(2) Descriptive IEAs (covering-law explanations of action) show that the action was the predictable consequence of some intentional states of the agent. In this way they provide information that is relevant for *predictions* in identical and similar situations in the future, and for “making sense” of the action (understanding).

These results show that we should adopt a pluralism of formats of intentional explanations rather than keep on searching for the ideal format that can do all the work.

Explanations of Facts and Explanations of Contrasts

Before turning to our defense of intentional explanations against “behavioristic skepticism” it may be useful to spell out the general view on scientific explanation into which the ideas in this paper fit. We are adherents of an erotetic model³ of explanation, which regards explanations as answers to *why* questions. We can distinguish at least four types of explanatory questions that scientists often ask. First, there are *non-contrastive* questions:

Why does object *a* have property P (at time *t*)?

Examples are:

Why is Belgium a monarchy?

Why did the Confederates open fire on Fort Sumter in Charleston, South Carolina in April 1861?

Contrastive questions, on the other hand, can relate to contrasts in property (P-contrasts), contrasts in object (O-contrasts), or contrasts in time (T-contrasts). The general formats of the questions are:

Why does object *a* have property P (a time *t*) rather than property P'?

Why does object *a* have property P, while object *b* has property P' (both at time *t*)?

Why does object *a* have property P at *t* but property P' at *t'*?

³ The logic of questions is usually called erotetic logic; this is why this approach to explanation was given this label.

FUNCTIONS OF INTENTIONAL EXPLANATIONS OF ACTIONS

P and P' are mutually exclusive properties. Examples (two of each) are:

Why did a large part of the Irish population migrate in the 1840s rather than stay in Ireland?

Why was there polarization in Chilean politics in the 1960s rather than integration?

Why were 75% of Dutch Jews deported and killed in World War II, while in France only 25% were deported and killed?

Why do rice markets in Southeast Asia work with auctions while rubber markets do not?

Why was there a successful revolution in Russia in 1917 but not in 1905?

Why was the joint-family system dominant in rural areas of Taiwan in the period before 1930, while after 1930 nuclear families became dominant?

We think that philosophers of science must start from this variety of explanatory questions and try to determine the motivation behind them. Then they can clarify the structure and function of various types of explanation. This is what we did in this paper for explanations of actions: descriptive IEAs are explanations of facts and contrastive explanations are explanations of contrasts. We have argued that they serve different functions.

Intentional and Other Explanations of Actions

First of all, we want to argue here that there are cases in which intentional explanations are indispensable. Our argument starts from the different motives that scientists or other people may have to ask explanatory questions about actions. We have mentioned such motives in this paper: therapy, prevention, prediction of similar behavior, making sense of actions, and—last but not least—making ethical and other judgments.

Behaviorist explanations are typically dispositional; they explain behavior B of a system x (i) by referring to external causal factors, i.e. a situation of type S including certain stimuli that triggered this behavior, given that x has or had a disposition D to do B in S, or (ii) by referring to a disposition D of x that explains why x behaves as it does in a situation of type S which includes certain stimuli, given that x is in a situation of S. The situation S is the *triggering cause* of the behavior and D is the *structuring cause*. D is taken to be the result of a history of reinforcement. Such explanations may be helpful if our explanation request is motivated by prevention, prediction, and therapy, but they are certainly inappropriate when the motivation behind the question is judgment formation (in that case we need an intentional explanation).

What if the explanatory question is motivated by practical considerations like prevention, therapy, or prediction? One might argue that behavioristic explanations can do the job in such contexts. However, such explanations require stable (deterministic or probabilistic) relations between a system's behavior and its environment. Some systems react completely differently when confronted with the same situation, so the behavioral output depends highly on internal determinants. This is where an internalist approach should take over from an externalist (behavioristic) one⁴.

To sum up: there are contexts in which intentional explanations are the best option and contexts in which behavioristic explanations are the best choice. Neither approach is superior in all possible contexts.

References

- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- Broad, W. (1985). *Star warriors*. New York: Simon & Schuster.
- Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon.
- Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: The Free Press.
- Lackey, D. (1994). Military funds, moral demands: Personal responsibilities of the individual scientist. In E. Erwin et. al. (Eds.), *Ethical issues in scientific research* (pp. 397-409). New York: Garland.
- Lipton, P. (1993). Making a difference. *Philosophica*, 51, 39-54.
- Little, D. (1991). *Varieties of social explanation*. Boulder: Westview Press.
- Mitchell, S. (2000). Dimensions of scientific law. *Philosophy of Science*, 67, 242-265.
- Vanderbeeken, R. (2004). Models of intentional explanation. *Philosophical Explorations*, 7, 233-247.
- Vanderbeeken, R., & Weber, E. (2002). Dispositional explanations of behavior. *Behavior & Philosophy*, 30, 43-59.
- Woodward, J. (2001). Law and explanation in biology: Invariance is the kind of stability that matters. *Philosophy of Science*, 68, 1-20.

⁴ For examples and further discussion see our paper in a previous volume of this journal (Vanderbeeken & Weber, 2002).