

## EXPLAINING THE ACTIONS OF THE EXPLAINERS\*

Explaining has been seen by many as the paradigm display of rationality on our part, and rationality as our most distinctive attribute. To be an explainer is, along the way, to be one who can generalize, infer, predict, offer justifications, conform to the rules of evidence, and, in order to offer more ambitious explanations, one must be a theorizer, and a theory evaluator. But other things we do are equally complex and equally distinctive. Such capacities as the capacity to compose, perform and interpret music, to stage an opera, to design institutions and conform to their norms, to write comedies and comedies about comedy writing, are not plausibly seen as merely spinoffs from our capacity to become theorizers. The reverse seems a likelier hypothesis, if we feel compelled to reduce some capacities to others. We might well wonder which of the many things done by those capable of offering explanations are the most wonder-provoking, and most interestingly unlike the doings of volcanoes, whirlwinds, minerals, plants, viruses, bacteria, insects, and most animals. Hempel believes that it is our rationality, that property which, among its other workings, enables us to give and improve explanations, and offer theories of explanation as well as theories about other *explananda*, that entitles our actions to a special explanation schema. I shall discuss his version of this rationality, and try to see how much it does and doesn't explain about us, according to his version of explanation.

Among the many things which only explainers do is to argue about the nature of explanation. *Why* questions seem to be answered in a variety of ways in a variety of contexts, and any attempt to reduce the variety to some canonical essential form will provoke resistance among the users and defenders of the forms which are most transformed by the reduction. Hempel's Kantian account of explanation as subsumption under law has met with most resistance from those who have attended to those after-the-event-explanations we give of unlikely and unpredicted happenings. It is not necessarily that these are not subsumable under law, but rather that the laws do not appear to be what explains them. We seem to explain some

happenings by giving a narrative, a sequence of connected events made possible but sometimes only improbable by the laws we know. Explanations of events in human history seem to take this form, and since some of the events in human history are intentional actions, so also do ordinary explanations of some human actions. When Becket changes his lifestyle on being made archbishop by Henry the Second, and surprises him by the austerity and piety he assumes, this unpredicted turn of events and the subsequent human actions leading up to Becket's murder are understandable, although few of them would naturally be described as what any rational agent in those circumstances would have done. The style of life Becket adopted was quite appropriate for an archbishop, although a surprising change of style from the *bon vivant* he had been before. What he did when he excommunicated Eynsford despite the King's protest, and refused to hand over to the civil powers a priest charged with debauchery and murder, insisting on the church's privilege to try its priests, were understandable assertions of the church power by its primate, but surprising defiance of his former friend, companion, and benefactor. But such surprising actions by rational persons are so familiar to us that it is surprising that we retain our capacity for surprise. It is the norm for rational persons to display their rational freedom in unexpected ways. When they do, we seek to reconstruct their reasons, to understand how they came to do the surprising things they did, and sometimes we succeed in getting an explanation which satisfies us. Rarely will it show us that there was nothing else for such a person in such circumstances to do, and we seem content if we can be shown that what, say, Becket did was *a* reasonable thing to have done in the circumstances. Had he cooperated with the King in curbing the church's power, we would have understood that too, given his friendship with the King, and his earlier apparent agreement with him about the desirability of retrenching ecclesiastical privileges. But we know that people do sometimes change their minds, change their allegiances, change their goals and values, change the norms they endorse, so we are not unduly perplexed by such happenings, provided we can discern *some* motivation for them. We accept the limits of our power to predict the paths of golfballs and, both for the same reasons and for the extra reasons brought in by rationality, creative imagination, and freedom, of people's lives. Like anything else subject to merely probabilistic laws, we are sometimes the subjects of improbable happenings, and since we have the power

to revise or reject the rules and policies we try to adhere to, we sometimes do improbable things.

These considerations make it hard to see how we could ever get something like Becket's change of lifestyle on being appointed archbishop explained by Hempel's Schema R, for rational action. That schema is

*A* was in situation of type *C*.

*A* was a rational agent.

In a situation of type *C* any rational agent will do *x*.

Therefore *A* did *x*. (*Aspects of Scientific Explanation*, hereafter *Aspects*, p. 473.)

The general regularity, subsumption under which explains the action, here seems to be a general claim about what rational agents always do, and a derived generalization about what rational agents do in type *C* situations. What is it that rational agents such as Becket always do? Hempel's answer to this comes in various forms. They pursue their "total objectives," where this is usually a matter of achieving some wanted end state without violating norms they have adopted (*Aspects*, p. 465). But Becket seems to have been adopting new norms when he changed his ways. He switched from a worldly and hedonistic style of life to an ascetic other-worldly one. Why? Hume, in trying to understand Becket's life and death, finds pride and ambition to have been the driving forces, albeit "under the disguise of sanctity and zeal for the interests of piety and religion" (David Hume, *History of England*, Ch. VIII, pages on the year 1170). But why this particular disguise? When Henry made him archbishop, he might have satisfied his ambition by retaining rather than renouncing his position as chancellor, and so exercised both ecclesiastical and civil power. Why did he choose to separate himself from and oppose Henry, rather than to continue as his ally, and as his supporter in the attempt to re-confine ecclesiastical privileges within their earlier bounds? Was it his pride which, offended by his debt to Henry for such power as he had, dictated that he take the path of separation and opposition rather than continued friendship and cooperation? It is not clear that he could expect more power by opposing Henry and consolidating his power as archbishop, than by combining his power as chancellor with the power he would have had as archbishop of a church less powerful than he in fact chose to aim to make it. As Hume said (*op. cit.*, writing of the year 1162) his appointment as arch-

bishop "rendered him for life the second person in the kingdom, with some pretensions of aspiring to be the first". He chose to exploit those pretensions, rather than to be second person as well as chancellor and friend of the first. Hume makes his choice intelligible by pointing out how it satisfied his ambition, and by mentioning his pride as well as his ambition, but all that his account does is show us how it was *as* rational for Becket to do what he did as it would have been for him to satisfy his pride and ambition under the combined guises or disguises of archbishop and chancellor.

Hume retells the story of an episode from the days before Becket was made archbishop, one that might be thought to explain why Becket, once archbishop, renounced the chancellorship.

'One day, as the king and chancellor were riding together in the streets of London, they observed a beggar who was shivering with cold. Would it not be very praiseworthy, said the king, to give that poor man a cloak in this severe season? It would surely, replied the chancellor; and you do well, Sir, to think of such good actions. Then he shall have one presently, cried the king, and seizing the skirt of the chancellor's coat, began to pull it violently. The chancellor defended himself for some time, and they had both of them tumbled off their horses to the street when Becket, after a vehement struggle, let go of his coat; which the king bestowed on the beggar...' (ibid.).

This public playful display of the way Henry saw Becket's duties as a member of the clergy to combine with his privileges as chancellor may have been taken by Becket as a warning as well as a public humiliation. Prudence as well as hurt pride may have dictated his refusal to try to combine the powers of archbishop with those of chancellor, after this demonstration of the precariousness of his dignity as chancellor. Still, we scarcely get a tight argument of the form that any rational person who is proud and ambitious, and whose pride and ambition had been both furthered and mocked in the way Becket's had by Henry, would choose separation rather than alliance, given Becket's situation and opportunities. We do understand Becket's motivation better, when we hear of this episode, but not because we see that it made his actual choice the only one a rational person could make. If his norms could change to require mortification of bodily appetites, to give his pride and ambition better scope, then they might also have changed to require mortification of his pride, to give his ambition *wider* scope. Even if we knew more of the secrets of Becket's heart than we are ever likely to know, we would not necessarily come to know that what he did was, for him, *more* rational than what he

could have done. Nor do we need to know this to understand his action by coming to be acquainted with more details of his situation, his personality, and his past.

Does Hempel think that, to explain as a rational action Becket's adoption of a hostile attitude, we must show it to be *the* action any rational agent would take, given Becket's situation, goals, and norms? No, for as Hempel points out there may be no one course of action which is *the* rational one, in his sense, under conditions of uncertainty (*Aspects*, pp. 468-70). Becket's choice was such a decision, but we may feel that even had Becket been a firm adherent of maximax, or of maximin, or of some other rule, this still would not settle whether he should indulge his pride to the extent he did, on the interpretation adopted. It is not attitude to risk or uncertainty which seems unfixed in Becket's case, as much as goals and norms. Can we only be rational in Hempel's sense provided we achieve stability in our goals and norms? I find some instability in Hempel's position here. On the one hand he says "I will not impose the requirement that there be "good reasons" for adopting the given end or norms: rationality of an action will be understood in a strictly relative sense, as its suitability, judged by the given information, for achieving the specified objective." (*Aspects*, p. 465). This limits the use of Schema R to actions which do not display a currently occurring change of mind about objectives and norms, and it rules out explaining any such changes by schema R. On the other hand Hempel characterizes rationality in terms of what presumably is Ryle's concept of *higher order dispositions* (*Aspects* footnote 4), and it is hard for me to see how Rylean higher order dispositions can be prevented from leading to the possibility of alteration or revision of objectives and norms as well as change or revision of beliefs and perhaps of attitudes to risk and uncertainty. Hempel does not *say* that his concept of rationality as the capacity for higher order response is Rylean, but he had earlier (*Aspects*, p. 458) referred us to Ryle for details of what he, Hempel, means by a disposition, and when he calls rationality a "broadly dispositional trait" in the section where higher order dispositions are spoken of, he again refers us to Ryle (*Aspects*, footnote 17). Ryle, like Hempel, was trying to give an account of what distinctive features the human mind displays. Ryle's aims were more descriptive than explanatory, nor did he choose the term "rationality" for the cluster of higher level dispositions he found in us. Nevertheless I think we should look to see what he meant by "higher

order", to try to see if what Hempel calls rationality can both be higher order, in Ryle's sense, and also relativised to goals and norms taken as fixed.

Hempel says that the rationality of agents whose behaviour is to be explained by Schema R is a "broadly dispositional trait" (*Aspects*, p. 472) and that the relevant dispositions include higher order ones, "for the beliefs and ends in view in response to which, as it were, a rational agent acts in a characteristic way are not manifest external stimuli, but rather, in turn, broadly dispositional features of the agent" (*Aspects*, p. 473). Presumably this means that for merely intelligent behaviour, like a dog's or an ape's, which manifests beliefs and ends in view rather than manifesting a response to the agent's beliefs and ends in view, Schema R will be inappropriate. To explain the dog's behaviour in digging we will use, perhaps, some weaker relative of R, let us call it "Schema I":

*A* was in circumstances *C* (craving a bone, "remembering" where he buried one).

*A* is an intelligent agent.

In circumstances *C* any intelligent agent digs where he remembers having buried a bone.

There may be a whole array of explanation types for simpler animals – spiders may need "Schema N", for instinctive behaviour, and perhaps heating systems will fall under a parallel "Schema S" for state maintaining systems. The *form* of the explanation is in no way special to rational action, nor even is the fact that the circumstances will always mention beliefs and desires, since they will be needed also in explanations of intelligent animal behaviour, and some primitive analogues of them may be needed for instinctive behaviour. If we want to find out what is special about the explanation of rational behaviour we will have to look at the particular *sorts* of circumstances which need to be listed, given the force of "rational" instead of "intelligent," "instinctive" or "state maintaining" in the premisses. The circumstances will need to include the higher order beliefs, values, adopted strategies, and not just the knowledge and desires which we would cite to explain the dog's behaviour.

How does Hempel's version of rationality, in terms of broad and higher order dispositions, relate to Ryle's account of the sort of behaviour which displays mind? Ryle distinguished between abilities such as the ability to

cycle, and the abilities to comment on, approve, criticize, teach, or analytically describe, cycling. The latter are all higher order abilities since they are directed upon a lower order ability. Ryle devoted few words to reason and rationality, contenting himself with intellect and intellectuality, and it has been left for others, such as Hempel, to give a more or less Rylean account of what a *rational* animal must be. Ryle distinguished the typically intellectual activities of theorizing about and teaching other possibly non-intellectual displays of intelligence both from higher order yet non-intellectual activities such as "heeding" how one was cycling, when the terrain called for special care, and from the base level displays of intelligence, such as learning how to cycle when given some training.

It seems clear that animals as well as human persons show intelligence, in Ryle's sense, and they may show non-intellectual higher order capacities too – the capacity to imitate, and to monitor their own and their offspring's attempts at various tasks. They also seem to display the ability to train their young. Are they rational in the sense Hempel thinks that we are, if to be rational is to respond intelligently not just to intelligent performances but to *one's own beliefs and desires*? The mother cat when she suckles her kitten after being importuned by it can be seen as responding to the recognized desire of the kitten and to its belief that it has come to the right place to gratify that desire. We can treat cats as what Dennett calls "intentional systems," attributing to them beliefs and desires, and we *can* also attribute to them some sorts of responses to beliefs and desires. But then again we might equally well describe the mother cat's response as a direct response to the kitten's *behaviour*, rather than to the kitten's expressed *desire*. Hempel contrasts response to beliefs and ends in view with response to environmental conditions and external stimuli. Do the expressive and purposive moves of one's fellows count as merely external stimuli? The cat seems capable of responding to the intelligence-displaying and purpose-displaying behaviour of other cats, just as we respond to the expressed but unspoken purposes of our fellows. The cat displays nothing that we would have no choice but to take as the manifestation of a response to its own ends in view, and it is hard to see how any animal without linguistic expression of beliefs and ends in view could conclusively demonstrate such self-directed higher level attitudes. (Such activities as burying nuts for later consumption, or preparing a nest, which might be seen as directed on continuing and future desires, we dismiss as instinctive.) But if to be ra-

tional and to behave in a way calling for Schema R is to display not merely higher order dispositions, but self-directed ones and perhaps the special case of those which are intellectual self-directed higher-order dispositions, then most of our actions, as well as all the cat's, will not qualify as *explananda* requiring explanation by Schema R.

Ryle's themes, in *The Concept of Mind*, were that intelligence can be displayed in non-intellectual tasks as well as in intellectual tasks, and that self-directed higher order dispositions (such as that displayed in thinking "*sum res cogitans*") are special cases of attention to intelligent performances, whose-ever they are. "At a certain stage the child discovers the trick of directing higher order acts upon his own lower order acts. Having been separately victim and author of jokes, catechisms, criticisms and mimeries in interpersonal dealings between others and himself, he finds out how to play both roles at once." (*Concept of Mind*, (hereafter *C.M.*), p. 192). "*Cogito ergo sum*" happens to be both an intellectual and a self directed higher order performance, and Ryle's aim was to enable us to see that it no more requires a ghostly agent than do all the other intelligent but non-intellectual performances or other higher order and self-directed but non-intellectual acts, such as self-ridicule or self-mimicry. "A person can, indeed, and must act sometimes as reporter upon his own doings, and sometimes as a prefect regulating his own conduct, but these higher order self-dealings are only two out of innumerable brands, just as the corresponding interpersonal dealings are only two out of innumerable brands" (*C.M.*, p. 194). Ryle wanted to get our *studied* moves properly related to our *unstudied* moves, and our *studied* didactic and theorizing moves properly related to non-intellectual moves, as well as to relate higher order to lower order acts, and self-directed to other-directed acts. He gives us no definition of rationality, but merely a series of spectra of more or less intelligent, more or less studied, more or less intellectual, and of lower and higher order acts. Indeed he derides the epistemologist's and moralist's concept of Reason, and of that Conscience which is "just Reason talking in its sabbatical tone of voice" (*C.M.*, p. 315). "These internal lecturers are supposed already to know, since they are competent to teach, the things which their audience do not yet know. My Reason is, what I myself am not yet, perfectly rational, and my Conscience is, what I am not yet, perfectly conscientious. They have not anything to learn. And if we asked 'Who taught my Reason and who taught my Conscience the things that they have



learned and not forgotten?" we should perhaps be told of corresponding instructors lodged inside their bosom" (*ibid.*). Reason is a mythical omniscient lecturer and theorizer in one's breast, and rationality is its wisdom.

Those of us who are unwilling to cede the terms "reason" and "rationality" to the rationalists, against whom Ryle's devastating attack is mounted, would probably want to say that to have the capacity for these Rylean higher order didactic self-directed activities one must be rational, while to have merely the capacity to benefit from training, as the kitten has, or even the capacity to do the training, as the adult cat has, but to be without the capacity for studied higher order acts, let alone intellectual ones, and to have no capacity to "discover the trick" of turning higher order capacities upon oneself, is not yet to be rational. But what of all the intermediate possibilities? And what is the relationship between rationality and the capacity for higher order acts? Do we use Schema R to explain what the ape does when it imitates the successful food washing moves of its more inventive fellow? Do we use it for the human child reared by wolves who can laugh at himself as well as at others but as yet has no language, nor any chance to study? And, in our own case, where we believe the full range of capacities to be present (if anything counts as the "full" range, given that ridicule, criticism, endorsement, can go at indefinitely many levels) but not constantly displayed, do we use Schema R for those of our actions that do not display anything studied, higher order, or self-directed?

Hempel is well aware that much of our behaviour that we explain with reason explanations exhibits something much less than a clear case of response to our own recognized beliefs and desires. He discusses not merely the ideal or limit case of someone like Bismarck responding to his own clearly perceived various options – to make the unedited Ems telegram public, not to make it public, to edit it before publicizing it, but also cases where the agent is not fully aware of the desires he is indulging, censoring or sublimating. Bismarck's all things considered endorsement of his own preference for the public opinion-manipulating wily move to provoke France to declare war on Prussia is a clear case of higher order response to clearly seen aims and beliefs about effective strategies (although, had he decided against editing or publicizing it, we would not need to invoke different aims or norms to explain that alternative action – Bismarck might have judged his editing move too risky). A different sort of reason, however, would be given by a Freudian for the less successful person's choice

of a tricky subterfuge which gained her nothing that she could avow as a wanted end, but rather led to her own exposure as a deceiver. The analyst might cite, as an explaining reason, the woman's need to feel that she controls what others believe, especially about herself, and a belief that such control is better demonstrated by deceit than by openness, by self-concealment than by self-revelation. This would-be wily person responds to her own unconscious or not fully conscious ends by recognizing opportunities to pursue them. Schema R is to be used for explaining her behaviour, as well as Bismarck's. So to be a rational agent, in Hempel's sense, is to respond to beliefs and ends one may or may not be able to avow as one's own.

This considerably widens the field where a response counts as the response any rational agent would make. A particularly obtuse person may never get to the point of acknowledging the aims and companion beliefs which Freudians would cite to explain her behaviour, so may never give any evidence of capacity to respond to these particular beliefs and ends, *qua* her own beliefs and ends. The "real reasons" for her behaviour will not coincide with her reasons as she sees them. Real reasons, whether or not they are avowed or even avowable by the agent, are what we will try to cite when Schema R is used. Bismarck's avowed reasons are presumably taken by Hempel as his real reasons – whether or not some deep unacknowledged personal desire to control the fate of nations lay behind his desire to preserve and enhance Prussia's national honor by provoking a war from which he expected Prussia to emerge victorious, he really did want that war – it was his end-in-view, even if not his ultimate end, so the reasons he gave for editing the telegram were real reasons, really explaining his action. He acts in response to his acknowledged beliefs and ends in view, whereas the less selfknowing agent acts in response to her unacknowledged beliefs and to ends not yet in clear view to her. The question I raised was just what sort of higher order response to beliefs and ends Hempel thinks necessary to demonstrate rationality. It cannot be conscious guidance by those ends and beliefs, implying as that does an ability to recognize them as one's ends and beliefs. The person whose avowed reasons are discrepant from her real reasons shows an ability to avow, but not to recognize or acknowledge true reasons, and if she were clever enough at selfdeception and unfortunate enough to have a motive to employ it, she might never acknowledge any of her real reasons. Should we

use Schema R for her, or is she not a rational agent? Should we use it to explain her behaviour as long as she has shown *some* disposition to acknowledge real reasons, although on this occasion the real reasons are ones she never shows any ability to acknowledge? It is not yet clear how extensive a disposition to what sort of response to her real beliefs and real ends she must demonstrate to count as rational. Let us assume that it is enough if she has displayed the ability to acknowledge *some* of her real beliefs and ends, whether or not she shows that ability in relation to the act to be explained by Schema R. Then among circumstances *C* we should specify not only her real ends and beliefs, but also specify whether or not she is aware of them. For what a person does when consciously guided by her beliefs and ends will differ in many ways from what she does when she needs to disguise them and their force. Reason shows one sort of cunning when reasons are unacknowledged, a different sort when they are clearly in view.

A question now arises about the scope of Schema R. If Freudians can use it to explain action inappropriate relative to the agent's avowed objectives that is, however, appropriate relative to some postulated unconscious objectives, then can it be invoked to explain *all* action which is less than fully rational relative to avowed objectives? Can the postulation of unavowed objectives serve to transform all apparently less than rational action into rational action? Do we use Schema R to explain irrational and weakwilled action, and action displaying incompetence or carelessness in deliberation or execution, as well as to explain fully rational action? Of course a necessary condition for doing so is that the resulting explanation have some merits as an explanation – that the attribution to the agent of beliefs and desires not avowed by that agent really would have some explanatory power. Freudians, when they attribute unconscious objectives, usually have independent evidence, from dream contents and the like, for such attribution, so that the acceptability of their claims about a person's unconscious goals do not rest solely on the explanatory gains of merely postulating such goals. If we have no such independent evidence for unavowed goals, but attribute them merely to “rationalize” the agent's actual behaviour over a stretch of time, then the acceptability of such attributions will rest solely on their explanatory power and on the lack of better confirmed explanations. Our only reason to accept the statement attributing such beliefs and desires to the agent will be its explanatory power. Other-

wise we would simply have a groundless version of that part of the *explanans* which takes the form "A was in situation of type C". But is it sufficient for use of Schema R that there is sufficient reason to attribute to the agent a particular set of unavowed beliefs, goals and norms?

This will depend on the force of that other part of the *explanans* asserting *A* to be a rational agent. It must mean more than that guiding beliefs and desires are plausibly attributed to the agent, otherwise all the higher animals count as rational agents. It must mean less than that what guides *A*'s behaviour are consciously recognized beliefs and desires, otherwise the schema cannot, as Hempel wants it to, cover Freudian explanations. Do we or don't we invoke it to explain all less than rational behaviour on our part? If what it means is "*A* aspires to conform to the demands of rationality," then this would indeed limit it to those agents capable of such higher level aspiration, and it would have a role to play in explaining their failures as well as their successes, but not always or often by displaying what they did as the thing *to do* given these norms of rationality. If we are to show that what *A* did, when he acted less than rationally, was what any agent who aspired to be rational *would do* in circumstances *C*, then circumstances *C* must contain those details of the limits of *A*'s competence or of his will to be rational which explain the partial *failure* of his attempt at rationality. Freudian explanations purport to do this in one special way – by citing aims *competing* with the avowed aims, norms competing with the rational norms. There could be other reasons for our failure to be as rational as we consciously want to be. We may sometimes be singleminded but feebleminded, rather than, as the Freudians portray us, strongminded but not singleminded. Inner conflict need not always be the explanation of failure to conform to endorsed norms, be they the formal norms of rationality or more substantive norms. We surely are imperfectly rational beings and whatever schema fits our behaviour has to be a schema to explain attempts at more than we usually succeed in doing.

Hempel, however, does not seem to want the force of "*A* is a rational agent" to be "*A* accepts and tries to conform to the norms of rationality". For when he explains why the *explanans* must contain the claim that *A* is rational, he says "Now the information that agent *A* was in a situation of kind *C* and that in this situation the rational thing to do was *x*, affords reason for believing that *it would have been rational for A to do x*, but no grounds for believing that *A* in fact did *x*. To justify this latter belief we

clearly need a further explanatory assumption namely that – at least at the time in question – *A* was a *rational agent* and thus was *disposed* to do whatever was rational in the circumstances” (*Aspects*, p. 471). Here “disposed to” must mean more than “thinks he should, and aspires to”, otherwise we still would have, from the *explanans*, no way of deriving more than that *x* was what *A* thought he should do, and made some efforts to do. What Hempel later says about the “model of a consciously rational agent”, comparing its usefulness to other ideal concepts, such as that of ideal gas behaviour, makes it quite clear that the norms of rationality cited in Schema R are taken as norms which are in fact conformed to, on the occasion in question, not ones making a difference by being on that occasion unsuccessfully aspired to. “Very broadly speaking, the explanatory model concept of the consciously rational action will be applicable in those cases where the decision problem the agent seeks to solve is clearly structured and permits of a relatively simple solution, where the agent is sufficiently intelligent to find the solution, and where the circumstances permit careful deliberation free from disturbing influences” (*Aspects*, p. 482). This means, I think, that Schema R can rarely be used of actual human action, although it may explain some parts of human *planning*. Bismarck may have acted on a rational plan, in this ideal sense, but Becket’s decision problem was less clearly structured – or was it that Becket’s pride distracted him from seeing the rational solution to the clearly enough structured problem of how to maximize satisfaction of his ambition? Or did he in fact solve it perfectly, achieving by martyrdom greater fame and glory than Henry? As Hempel says, (*Aspects*, p. 478) in real life the notion of optimal action relative to one’s total set of objectives and beliefs is impossibly obscure.

To be disposed to be rational, in Hempel’s sense, seems to mean to be about to succeed in doing what any perfectly rational agent in those circumstances would do, unless prevented by disturbing factors external to the agent’s competence, beliefs, and total objectives (conscious or, alternatively, unconscious). So this Hempelian rationality comes and goes, and, when relative to conscious beliefs and objectives, is gone more often than not. Thus Hempel can speak of a person being “a consciously rational agent (*at a certain time*)” (*Aspects*, p. 479, my emphasis). Rationality comes and goes, as we succeed or fail to conform to its norms. But, as we usually speak, a rational agent is one with the *capacity* both to endorse

the norms of rationality and to achieve some measure of success in living up to them. The capacity *may* come and go (when one is mentally ill, it may go) but it will not be gone whenever we act less than perfectly according to the norms of rationality, (whatever they are – I shall later raise a question about Hempel's version of them). There seem, then, at least two reasons to question whether to be a rational agent is to do what is the rational thing to do. First, there may be nothing which is *the* rational thing to do, merely several things all counting as *a* rational thing to do, if rationality is the capacity of higher order response of a Rylean sort. Secondly, even when there is a unique answer to the question of what it is rational for her to do, a rational agent may fail to see that answer, or having seen it, fail to act on it, without thereby showing that she has lost her rational capacity. What transpires will be different in nature from what would transpire in a being who had no pretensions to doing what is seen to be rational. The failed cunning of reason shows itself differently in rational animals than in non-rational animals.

The trouble, I think, lies in the way Hempel moves from rationality as a normative concept to rationality as an explanatory concept. His move is to let the explanatory concept apply only to those in whom, or those times when, the norms of rationality are conformed to perfectly. He postulates the ideally rational agent, and lets Schema R apply only to such an agent. But agents as we know them are not ephemeral beings but lasting beings who have acquired and learned skills, which they display with varying success, and whose objectives make what sense they do only because they *are* lasting beings, who know themselves as variable in performance, capable of progress and regress when judged by the norms they endorse, and capable of progressive or regressive change in such endorsements. We cannot be agents *at a certain time*, only over time. If rationality is a properly characteristic of agents, and only derivatively of their actions, then it too cannot be displayed in isolated episodes, but only in a pattern of activities of forward and backward looking agents. Our actions vary in the extent to which they display our rational capacities, so are more or less imperfectly rational actions, but that does not mean that the higher level capacities in which rationality consists come and go. We often need to refer to our norms, including the norms of rationality, to explain our failures as well as our occasional successes. Schema R, however, does not help us to explain our less than fully correct moves. The truth of that part of

the *explanans* which asserts *A* to be a rational agent is not independently testable, if all it means is that *A* is on this occasion the locus of a visitation by the spirit of perfect rationality, enabling *A* to do the rational thing relative to the aims and beliefs attributed to *A*. It becomes trivially true whenever the beliefs and aims attributed do make the action done the rational thing to do. But it is (trivially) *true* only if “rational agent” means “ordinary lasting agent having a passing moment of perfect rationality,” while being false if we take rationality as a capacity for reliable displays of such perfect rationality. The only capacity actual human agents have is the capacity to make more or less rational plans, and to try to carry them out in action. Their actual actions are seldom completely according to plan, and their plans seldom perfectly rational. Schema R, then, seems rarely to get application to our intentional consciously purposive actions. It will apply more often to our prior planning, and it may more often apply to our unconsciously purposive doings. But it fails to apply to our only more or less successful attempts at consciously rational action.

If we cannot use Schema R for our actual actions, only for some of our plans and for some of our unconsciously motivated failures to live according to our avowed plans, norms and objectives, then we must use some other schema to understand ourselves as we usually display ourselves in intentional action. Schema R turns out to be more suitable for robots than for us. It turns out also to be minimally different from my hypothetical Schema N for successfully evolved insects, say, cockroaches. Varying the previously quoted statement of Hempel’s (*Aspects*, p.482) we could say “The explanatory model of the instinctive action will be applicable in those cases where the decision problem solved by the insect’s instinct is clearly structured and permits of a simple solution, and where the insect is well enough adapted to have that solution wired in, and where circumstances allow instinct to operate undisturbed.” Such an ideal concept will get application to insect actions more often than that of ideal rationality will get application to human action. To explain our actions to ourselves, we will do better with a schema for imperfectly self-conscious, imperfectly rational, imperfectly singleminded animals, capable of revising their beliefs, goals, and strategies, and capable of surprising one another and themselves. This describes the higher animals as much as it describes us, and I think that is what we need – a schema for ourselves which will enable us to see how ones who act as we do could have come to do so, given what it is rational to believe about our animal ancestry.

A question I deferred earlier was that of the content of the norms of rationality, those norms to which we aspire with varying degrees of success. I shall give no positive answer to this hard question, but merely point out a tension in Hempel's account. He speaks of rationality as requiring the adoption of means appropriate to our total objectives or ends in view, constrained by whatever other norms we espouse. He also speaks of maximizing expected utility. This instrumental conception of rationality seems to me to fit uneasily with his version of scientific and so presumably rational explanation. To give a Hempelian explanation is to subsume under law, not to suit means to ends. If, as rational explainers, we must see events as conforming to law, if we must try to conform to the imperative "seek for laws from which occurrent events can be deduced as instances," then why, as rational agents intent on *other* activities than explanation, are we free to seek merely for efficient means to our ends? There is a tension between Hempel's Kantian account of rational explanation, and his means-end relativist account of our *other* rational activities. If to be rational explainers we must ourselves conform to the laws or rules of the explaining game, themselves dictating that we try to see *everything* as law governed, then would we not expect many or all of our *other* rational activities would also display such respect for laws and rules, rather than mere unconstrained goal pursuit or mere maximization?

Two alternative reconciling moves are open to the Hempelian here. One might try to show the concern with laws as a special case of means-end rationality, or one might try to show means-end rationality as a special case of Kantian law centered rationality. Hempel does the latter in as far as he encourages us to subsume our own instrumental rationality under descriptive law to understand it. But what of the practical norms we are following if we do try to use Schema R, or some other Hempelian explanation schema, to understand ourselves? Do such self-explaining moves display Kantian or instrumental practical reason? Are we to have a teleological or a deontological account of the explaining game? Hempel, like so many others, seems to want to accept a Kantian law-centered account of theoretical reason, but to combine that with a means-end relativist account of practical reason. Explaining and its norms, which dictate law discernment, would then have to be shown as suitable means to the explainers' total objectives. The Hempelian explainers' preoccupation with laws will then derive from his special objectives, not from his rationality



as such, so Kantian reason will be absorbed into the instrumental reason of those with a taste for law discernment.

Our self-approved rational actions are a subset of our actions as intelligent and creative animals capable of speech and so some degree of self-consciousness, and capable of other norm-constrained, inventive, and higher order acts. Among our rational activities is explaining. Any acceptable explanatory schema for our actions ought to explain our explanatory moves. Schema R may be got to do so, if we take rationality as adopting efficient means to achieve our objectives, and good covering law explanations as among our objectives. To be a rational explainer will simply be to make the moves which can be expected to yield the sort of explanations which one aims to get. But suppose we ask why we avow and endorse this aim? Why do we, once we reflect on the matter, *endorse* the objectives of improving our explanations and our understanding of their structure and role? Is it rational, given our *total* objectives, to care about explaining in the way we do? If these objectives include the protection of redwoods, the continued preservation, creation and enjoyment of good music, good novels, good jokes, good wine, as well as good explanations, good science and nonlethal technology, then we would hope to see some sort of coherence among our multiple goals, and some explicable place for explaining and theorizing among them. But without its companion human activities, explaining becomes a mysterious, inexplicable activity. The rather rationalist conception of rationality which Hempel adopts, stressing as it does the formal demands of consistency and maximization, and of rational explanation as subsumption under law, is one whose ancestors are to be found in Descartes and Leibniz, ones which see reason as a divine faculty, by which we approximate to a god's eye view of things. But if one thing seems quite certain about divine activity it is that gods need do no explaining whatever. Omnipotent creators and legislators understand their creation and its laws without having to explain it to themselves – they *intend* it, so know it without either observation or explanation. If we are to understand ourselves as explainers, we had better not model our version of the rationality which shows in our better explanations on any theology-based version of our reason. Hempel shows no wish to base his concept of rationality on any theological notions, nor is he concerned with the intellectual ancestry of the concepts he relies on. I think that our usual conceptions of deduction and of law *are*, however, concepts with a theo-

logical ancestry, and our usual version of consistency and coherence in a comprehensive system can be seen as an intellectual descendent of the theologian's version of divine unity in an encompassing variety of attributes. This ancestry can be ignored, but it tends to reassert itself.

It shows itself in a dualism between what is seen as our less than rational animal habits, desires, means-end strategies, social propensities, customs, artistic and musical tastes, and our more "divine" capacities for mathematics, explicit inference and reasoning, for law discernment, for scientific theory, decision theory and other abstract systematizing activities. This dualistic version of our nature has difficulty doing justice to the linkages between, say, music and mathematics, or constitutional law and scientific law, between inference and associative thinking, explaining and explaining away. I cannot here try to defend a more naturalistic Rylean conception of our capacities against the rationalist conception, traces of which I find in Hempel's concepts of explanation and of rationality, in tension there with the more naturalistic and Wittgensteinian elements taken from Ryle. The rationalist logic of logical empiricism never did combine very well with the empiricism. To see ourselves as combining a quasi-divine or ex-divine reason with other capacities of obvious animal origin is to see ourselves as demi-gods in beasts' bodies, worse monsters than ghosts in machines. We need some account of human rationality which makes it neither mere animal means-end intelligence, nor divine contemplation of an orderly creation, nor an uneasy alternation between these.

Hempel's account of explanation, and of its varieties, has and will continue to provoke reflective responses. My own response to it, in this essay, has been less "rational" than the responses it more naturally invites, but I hope not the less reflective nor the less appreciative for that. I have tried to express some worries I have with the Hempelian account, worries which may show more about the limitations of my rationality than the limitations of the account whose implications I have tried to explore. At the end I attempted to subject that account of the explanation of rational action to the test of reflexivity – to ask if it can, in its own terms, explain itself. This is to ask of it whether, like divine and rationalist mind, it can include itself in its survey. But, of course, if I am rejecting such a theological conception of mind in favor of a more social and naturalistic one, why should this be a reasonable demand to make of an account of explanation? It may not be a reasonable demand. Still, if we are to exercise our capacity for higher

order response (whatever its ancestry), we seem bound to want to try to turn accounts on themselves. It would, perhaps, be more accurate to say that it is understandable that those of us who have no account of our own to offer, but who are fascinated by and curious about the accounts of the creative thinkers such as Hempel, should indulge in such parasitic reflections.

## NOTE

\* I am grateful for helpful comments on earlier drafts of this paper from Kurt Baier, Nicholas Rescher, Arthur Ripstein and Patrick Maher.

Manuscript Received 14 May 1984

University of Pittsburgh  
Dept. of Philosophy  
Pittsburgh, PA 15260  
U.S.A.