

Is an Explanation a Reason?

Paul Rastall
Independent Academic, UK

Abstract

Are reasons and explanations the same things? The paper considers explanations and reasons in a range of contexts and identifies similarities and differences in them. It is suggested that, while the two notions are linked in reasoning, there is a difference of focus, with reasons being concerned with more general principles and explanations being concerned with the application of principles to specific instances. However, in a “model-dependent reality”, there can be explanations, but not ultimate reasons.

Keywords: causality, determinism, explanation, reason, reasoning, why and how questions

Reasons and explanations are closely related. The search for reasons – curiosity – is a key feature of human rationality. We want explanations for everything. Giving an explanation, either for understanding some phenomenon or for a practical course of action, is usually taken to constitute providing a reason, and giving a reason is taken to constitute offering an explanation (which may or may not be accepted), but are reasons and explanations different things, and – if so – how do they differ? If our conceptions of the world are provisional mental (or scientific) models, what is the status of explanations and reasons offered using such models?

Reasons and explanations imply the attempt to understand; since reasons and explanations are central to understanding, their characteristics show how we understand. We want to achieve certainty, or at least a high level of probability (as Locke maintained, 1973 (1689): 415 ff), ultimately leading to belief as an emotional attachment to certainty. Understanding – finding reasons and explanations – is also its own reward, and a pleasure (as Plato pointed out in the *Republic*). What we consider certain or highly probable through reasoning, as opposed to immediate awareness in perception, depends on the nature of our reasoning and on our level of knowledge and general “culture” of thinking at a given point in time. A major part of our reasoning, however, has always been the idea that reasons must be (logically) sufficient (the consequence follows from the premise), and thus we arrive at the principle of sufficient reason: for every fact there is a reason which explains why the fact is as it is; that is, there must be a connection between a general “truth” and a particular instance (Strawson, 1973, pp. 3–4). If we have a sufficient reason, then we have an explanation. But we may ask what constitutes that sufficiency of reasoning, and the nature of the connection between premises and conclusions.

There is no simple answer. Reasons will be sufficient, in part, if they correspond to our common experience of reality, but we know that the plausibility of explanations varies with time and the level of knowledge (the cultural background), and hence of the meaning of terms in the propositions. However, correspondence with “common experience” is insufficient as a criterion. Science constantly challenges our “common-sense” views. The finding that the earth looks red from outer space using an infra-red camera hardly corresponds to common experience or belief about our “blue” planet (vegetation reflects infra-red light). It is better to speak of experience guided by rational principles (Popper’s “enlightened common sense”, 1972a, p. 37). Further, we expect that a reason or explanation is more than a feeling, perception, or matter of belief, but feelings, perceptions and beliefs may be the reasons or explanations for ideas or actions (e.g. John spoke to Mary because he felt she was unhappy; John informed the police because he thought it was the right thing to do; religious and other ideological beliefs have been offered as reasons for atrocities from *auto-da-fes* to the *gulags*). Our reasons and explanations exist within conceptual frameworks (Poston, 2014) and, in those frameworks, there must be some fixed points (statements or ideas accepted as generally true; in John’s case (above) a moral principle is taken as a justification). Our mental models contain reference points, otherwise there is an infinite regress in explanation. Our frameworks are applied “topically” – they are applied to the particular fact to be explained and are valid for the time and level of knowledge of the explanation. However, the conceptual frameworks are our, human, conceptual frameworks. Thus, while beliefs or feelings may vary and may not be shared, one can see that they constitute reasons (reasonable or misguided) for others. They exist in the individual’s conceptual framework or a scientific paradigm (a “model-dependent reality”, Hawking & Mlodinow, 2010). Of course, explanations are interconnected, and a particular reason may be just part of a more general explanation (often a “suppressed” major premise in an argument). For example, I may observe, one day in September (in the UK), that the swallows are no longer around and conclude they have gone south for the winter. Generally

(as a (suppressed) major premise), I know that their migration is connected to the length of daylight hours and the availability of food. In all these cases, there is a role for intuition (Ewing, 1951, 48ff); one must just “see” that bird migration is connected to factors such as food and daylight, or that contacting the police and sense of moral obligation may be connected.

One aspect of the issue we are considering is whether the universe is deterministic, that is, whether there is a chain of causes leading to any event or set of circumstances from an ultimate cause so that, if we knew every causal circumstance then every event or circumstance would be explained as a function of all those causes (which seems to have been the deterministic position of Spinoza (2002)). The migration of the swallows would be caused by decreasing daylight time and food sources acting on the birds’ brains and behavioural dispositions; the length of day would be determined by the changing position of the earth relative to the sun and its inclination; the amount of food would be decreased by the reduced warmth; the tilt of the earth would be determined by physical forces in the solar system, and so on. Or, in an everyday case, John tripped over a brick; the brick had been left by a builder; the builder had been building a wall; a neighbour employed the builder because he wanted a perimeter barrier, and so forth. The chain of causes and effects goes on indefinitely to include gravity on earth (to hold the brick down) and the rotation of earth around the sun (to account for the presence of anything on earth), back to the formation of the universe. (However, we should note that this form of determinism is incompatible with the probabilistic nature of quantum physics.)

We should note that many of the points just mentioned as background understanding were unknown in earlier times and may still be unknown in some societies (or not accepted in some worldviews), so our reasons and explanations are dependent on time, place, and belief systems. The possibility that our understanding or mental modelling may not correspond to reality “as it is” raises the problem of the relation of theory to reality, and the meaning of terms so that we can relate words to experience. If a theory is successful in explanation, can we assume that the theory is a picture of reality, and hence an ultimate “reason”?

But we might also wonder whether there are non-causal explanations. We first consider some possibilities.

Conceptual Frameworks and Applications to Cases

An apparently fit and healthy middle-aged man with a good lifestyle and diet, a non-smoker who takes regular exercise and has no symptoms suddenly suffers a cardiac arrest and dies. It happens (more often to men than women). It is natural to ask—why? The autopsy shows that his coronary arteries were narrowed by plaque. The plaque forming the blockage became unstable and, when a piece broke off, there was bleeding. A blood clot formed and blocked the supply of blood to the heart, which stopped supplying oxygenated blood to the lungs. The lungs could not function, and death ensued in about four seconds.

The autopsy results give a causal explanation of the man’s death. Does that constitute a reason? Clearly, the causal explanation answers the question – how? When we ask for a reason, we normally want an answer to the question – why? In a purely deterministic conception of the world – one of cause and effect –, an answer to the question – how? would be equivalent to an answer to the question – why? That is, an explanation would be a reason (and vice versa). Can all why questions be reduced to how questions?

However, in the case of the man with a cardiac arrest, we might note that the man had had a build-up of plaque over a long period. Why did death occur when it did? A causal explanation would seek to identify a factor such as stress leading to a contraction of the arteries, thus causing the plaque to break off. Of course, the man may have had many periods of stress previously without adverse effects, so the question of why then? remains, and it would be necessary to try to identify all the factors leading up to the instability of the plaque. If all that information were available, we could still have a causal explanation. A persistent questioner might point out that almost everyone has some narrowing of the arteries (it is found in babies and the fittest of soldiers) with no infarction or cardiac arrest, so the question – why that particular man and not others? – remains. A statistical approach showing the percentage of the population who are affected in the way described is of no assistance here (correlation and probability are not causes), so many more details of the man's genetic and physical condition would be needed for a full causal explanation of plaque instability at that precise moment. If we had all that information, would we have a reason? Or is there something about the idea of a reason that escapes causal explanation and the chain of causes and effects? (Some people with religious views might feel that it was the man's "time" to die.) Are there phenomena which are beyond causality or are outside the chain that leads us back to some ultimate cause of everything? There may still be a sense of discomfort (or emotional outcry, however irrational) that no quantity of information really tells us why.

We could model a chain of causes as consisting of <{set of observable phenomena}, {set of connections between the phenomena}; {set of sequenced processes linking the phenomena}>. In the case considered, the phenomena include the fact of the cardiac arrest and the phenomena of the plaque, blood clot, arteries, heart, lungs, and death; the connections include those between the oxygenated blood and the diameter of the arteries as well as the relation of oxygenated blood to heart and lung function, and the relation of (absence of) lung function to death; and the sequence of processes – stress, contraction of the arteries, destabilisation of plaque, breaking off of plaque, blood clot formation, heart failure, lung failure, death. The components of the model serve as components of a causal explanation, but we would also need a set of general principles and logical argument to organise the causal explanation. The components we have described would fit easily into a hypothetico-deductive explanation of the sort given by philosophers of science such as Popper (1972) or Hempel (1965; 1974) with general principles such as, the human body dies without oxygenated blood by the lungs through the cardio-vascular system. The lungs must be supplied with oxygenated blood by the heart. If the heart stops pumping oxygenated blood, then the lungs fail, and death ensues. If the coronary arteries are completely blocked, then the heart stops pumping blood. If these conditions apply to a given individual, then the death of that person is explained (Strawson's general principles applied to the particular case). Of course, the general principles and their logical linkage are not themselves causal; they provide a framework of ideas for connecting the chain of events into an explanation.

However, the acceptance of the explanation involves a number of conditions. First, we must have a general understanding of the cardio-vascular system and its function in the maintenance of life. For people of earlier ages (say, before Harvey's demonstration of the circulation of the blood) or in less technologically advanced societies (or perhaps for people with particular belief systems), such a mental construction is unthinkable. In particular, those who regard life and death as decisions of a deity (or fate) will view the physical-determinist reasoning as insufficient or false. Second, we must accept the logic of the model (logical relations between premises and consequences must ultimately be intuited (as noted above, we must just "see" the logical connections) and apply it to the individual; i.e. we see intuitively that there is a

correspondence between the overall model and the facts relating to the individual – the general must be applicable to the particular case. This means, among other things, that the named phenomena, connections, and processes correspond to observable reality. Connecting the stages of the argument and linking named components to real-world phenomena clearly involves the meaning of terms and relations (a semantic issue, raising the possibility of understanding depending on arbitrary semantic distinctions). Third, we must accept that the idea of a “cause” includes not just a sequence of phenomena (that the cause is antecedent to the effect), but also the idea that phenomena may affect one another in a linkage (unprovable but always presumed, according to Hume, 1968; 1738); that one phenomenon has the “power” to affect (or be affected by) another (as Locke, 1973 (1689): pp. 162–179) put it). In general, the explanation of our man’s demise must be related to an overall mental construction (which may be subject to change in the light of further evidence, although that also requires a wider construction in which theory and evidence are linked). In turn, that implies human rationality – our common experience and common sense. There would be no point in offering our explanations of cardiac arrest to a cow. Explanation and reason, therefore, are dependent on human attributes – our way of understanding, or the search for reasons.

Fourth, then, we accept the most general conditions of an explanatory argument. Our notion of “cause” involves accepting the unidirectional “arrow of time” for recognising antecedence, and the idea of “uniformity of nature” (same causes–same effects). But we also accept some limitation in reasoning to that which is considered relevant to the phenomenon to be explained. No doubt, ultimately everything is connected, but reasoning would be impossible if we had to account for every aspect of a phenomenon. The man’s heart condition might be linked to weather conditions, his digestive tract, or the state of his left big toe, but we are satisfied to leave those issues aside unless there is some pressing reason to include them; similarly, tripping over the brick does not, in everyday practicality, require quantum physics for a reason.

To some extent, a deterministic view is also built into human thinking. On the other hand, it is known that the quantum universe is probabilistic, as noted above. The physicist, Roger Penrose (2010, 13ff), illustrates the notion of entropy with the example of an egg falling off a table and smashing on the floor. The egg goes from a high level of organisation to a highly disorganised state as a mess on the floor. It is conceivable, but overwhelmingly unlikely, that the molecules in the egg could somehow re-form themselves into an organised egg. If it happened, it would look like magic (as Penrose points out). The possibility of such an eventuality is incalculably small, but not absolutely nil. On the other hand, as Penrose also points out, if we had a film of the egg smashing and ran it backwards, no one (or at least no one with an experience of film) would be in the least surprised to see the egg rise from the floor and re-form itself on the table—it is part of the humour of children’s cartoons. As he says, “past teleology” (reversing the film) is common experience, whereas “future teleology” (the self-organisation of the smashed egg) is “just something that we never seem to encounter” (p. 51) – it is inconsistent with the second law of thermodynamics. It would be such “an enormously improbable sequence that we can simply reject it as a realistic possibility” (p. 14). The point here for our discussion is that our understanding of entropy and the egg is in some degree a matter of our way of constructing reality – our common experience. That common experience includes our common sense as a basis for rational thought. Popper (1972a, p. 33) asserts that common sense is the ultimate foundation of scientific thinking.

Naturally, then, we are amazed by animals that can restore lost organs or body parts (such as salamanders or geckos) and by the ability of the human body to self-repair some wounds, but we look for the physical mechanisms which explain such phenomena. Regeneration is, of

course, not reversal. It requires energy and new material as well as genetically based organisation to overcome entropy and is not an exception to the general principle of universal entropy. However, we could imagine worlds in which the self-reorganisation of eggs was the norm (i.e. where entropy is reversed), but it would not be our world even if we see it in cartoons).

Some questions in physics have no ultimate answers. Why we have an entropic universe rather than some self-maintaining organisation, why the cosmological constant is the way it is (along with a number of other key quantum values); these are well-established facts, and they can help to explain our existence (we wouldn't have the universe we know otherwise, and wouldn't exist – the so-called (weak) “anthropic principle”), but we do not know why they are so. There are speculations (e.g. every possibility that can happen does happen in an infinity of time and space) and, no doubt, cosmologists are working on the problems, but what is illustrated is that why questions push us ever further back into more explanations. Why? then, transcends, how? Even if the search seems ultimately pointless, people have always dreamt up explanations of first causes in creation myths and have provided explanations of disease – now known to be false – (humours, miasmas) based on the conceptual frameworks of the time.

However, not all explanations are causal (or perhaps even deterministic), even if some causes may be involved (as is also maintained by proponents of “grounded” explanation, e.g. Dasgupta, 2016). For example, it is known that albinism (say, in blackbirds) is disadvantageous to survival and reproduction (albino birds are more easily predated and mate less frequently), whereas camouflage (e.g. in the tree creeper to merge in with the bark) is advantageous to survival. Most birds have poorly developed smell, but nocturnal birds, such as the kiwi, have developed a strong sense of smell to find prey in the dark. Clearly, our Darwinian explanations of such phenomena involve a similar logical construction to the physical ones above. There are phenomena (albinism, camouflage, sense of smell); connections (colouration to reproductive chances or danger); and processes (development of appearance, sequences of mating or territorial behaviours). We use a similar hypothetico-deductive logic to produce explanations – all animals strive to survive; albinism endangers survival, camouflage assists survival, so albinos survive less well and reproduce less frequently, whereas well-camouflaged birds survive better. The arguments can then be applied to the specific cases of the blackbird, kiwi, or tree creeper. We can add genetic factors (e.g. mutations affecting melanin production or sensitivity to pheromones) to account for the occurrence of, say, white feathers or the olfactory glands.

The Darwinian explanation has similar requirements to the ones listed above, but it is not entirely causal. It requires us to recognise a principle of “survival” which involves feeding, safety from predation, and desire to reproduce. Such a principle is entirely reasonable (at least in a modern world view), but it again arises from our common experience – it is not a cause in the physical sense and requires a specification of its meaning. It is a framework of ideas applicable to particular cases. Even if any biological mechanism is physically determined in genetic terms at a molecular level, a Darwinian explanation appeals to our feeling of “what makes sense” (whereas the idea that the bird god got tired of all blackbirds being black, and decided to create some beautiful white ones, which might be the basis of a nice story, does not make practical sense). But again, various why questions suggest themselves, such as “why should kiwis have become nocturnal? why should tree creepers have adopted their particular way of making a living? Why do they peck upwards while nuthatches (with similar feeding behaviour) peck upside down?” We can only guess that some birds found a niche, exploited it, and passed on behavioural dispositions with a gradual genetic divergence of one population

from another without intermediate forms. (Note Darwin's (1901; 1859, 124ff) explanation of the absence of intermediate forms – they are less well adapted than animals that develop advantages in a particular niche and are not able to survive – we do not find “half-way kiwis”). Why those behaviours and not others? The explanation involves some causal mechanisms at a genetic/molecular level but seems to involve also exploitation of random opportunities in the environment. In other cases, markings involved in species identity and identification seem to have developed as a particular outcome from a range of possibilities through sexual selection. Thus, some bumblebees have red tails, some have white tails, and others have buff tails. While differentiation is advantageous, the particular development of red, white, or buff (species) seems random compared with the selective advantage of camouflage or the development of smell in nocturnal birds. We are dealing with what is, rather than what must be, so – red tails, white tails, and buff tails are equally good for the purpose of species differentiation (one assumes).

Similar remarks might be made about human societies. Rituals, politeness behaviours, and gestures are found in all human societies. Their explanation is connected with human sociality and group cohesion, but the particular forms of rituals and politeness strategies vary across cultures (you might greet by bowing, handshakes, forehead contact, etc.). Our account of them, linked to sociality (Dunbar's “social brain hypothesis”, 1998), involves, then, a non-causal construct which “stands to reason” in our model of common experience, but has the same sort of hypothetico-deductive reasoning. Human success depended in early homo sapiens on social cooperation (and other factors such as inquisitiveness, learning, and cultural transmission), but why human development took the particular path it did (e.g. reduction of sexual dimorphism – unlike in gorillas, for example) is unknown, and similarly is why this society adopted a certain structure (e.g. matriarchal or clan-based), ritual or politeness strategy, while another society adopted another. That is, we can offer a how explanation for the actual state of affairs, but not a why reason. The explanation again requires a construction with phenomena, relations, and processes, but the explanatory theory applied to those phenomena is not entirely causal. It requires the notion of “sociality” and its meaning.

In language phenomena, we find changes which spread memetically (and differentially in different sections of a population), and which seem to have no obvious reason. They seem, as the great Danish linguist Hjelmslev (1969) says, “capricious” (p. 170). For example, English has retained apophony in the “strong verb”, sing-sang-sung, but help has changed from being a strong verb in Old English (helpan-healp-holpen) to being a “weak verb” – help-helped-helped. Other verbs have disappeared from the language “in favour of others” (e.g. belgan “become angry”, frignan “ask”, weorpan “throw”), and others have changed in meaning (winnan “to fight”). Generally, the past participle of go is gone in most varieties of English, but in some varieties (e.g. around Tyneside in England) it is went (He should have went). In Scottish and some other varieties of English, the sound change /u:/ > /au/ is non-existent or incomplete (thus, /du:n/ “down”, /tu:n/ “town”), and also occasionally in Canadian English (/əbu:t/ “about”). In American English, the French loan word “route” is often pronounced with the (medieval) vowel change as /raut/, whereas in British English it is standardly /ru:t/. The exact origins of language changes, and the social forces and preferences which favour one form or another are generally unknown. It is supposed that changes appeared sporadically in prestige groups and spread by imitation (memetically) through sections of the population, but why a given change (rather than no change or a different change) was adopted is usually quite opaque. It may be that the adoption of a language change is a kind of social badge showing adherence to, or membership of, a peer group or association with a prestige group (and is resisted by other groups). Here we have a kind of unconscious motivation for the adaptation of verbal behaviour

(rather than a rational or causal reason) with plausible (but to some extent speculative) explanations involving similar forms of (non-causal) reasoning to the ones above. The changes are not the result of causal factors, and in some ways the social-memetic forces are similar to the gradual differentiation of bumblebee populations by unconscious sexual selection. However, it is also clear that there may be alternative points of view or explanations (some language change may be due to population changes or conflicts between more “progressive” and more “conservative” speaker groups, for example). That is, there are no definitive paradigms, but conflicting points of view – each with its own “semantics”.

Thus, we can explain the difference between the noun, house ending in /s/, and the verb, house ending in /z/, by reference to the Old English contextual phonological “rule”, voiceless fricatives become voiced between vowels (e.g. /s/ > /z/), where the verb goes back to *husen* (/hu:zen/), but the noun goes back to *hus*, /hus/, and the fossilisation of the two forms into Present Day English, but we cannot say why such a rule emerged in early Germanic.

As Hempel (1974) points out, the form of reasoning in historical explanation also follows the same “nomological-deductive” or hypothetico-deductive pattern as in natural science, but the components in historical reasoning involve the analysis of motivating forces. Naturally, historical explanations/interpretations or reasons for actions, events, or circumstances also involve the analytical perspective (paradigm or political viewpoint) of the researcher. They thus imply wider issues of the “philosophy” of the researcher, and hence an appeal to what is considered reasonable. Historical explanations (such as the causes of World War Two or the Russian revolution) may, like linguistic analyses, be many, subject to definitions, and open to debate. That debate further raises the question of whether it is reasonable to speak of “historical necessity”. While it may be reasonable to apply forms of deductive reasoning in historical explanation, the logical necessity of conclusions should not be confused with the necessity of events. The fact that something occurred and can be explained does not imply that events were somehow inevitable consequences of factors in the explanation. As in the language examples, above, even reasonable and well-founded explanations are not necessarily ultimate reasons but may be arguably plausible (and aid understanding).

This is even more obvious in the case of imaginary constructions, such as stories. Thus, we might ask, Why did Anna Karenina fall in love with Vronsky in Tolstoy’s novel? Any phenomena to be explained here are not real-world people or events (even if there are known to have been people on whom Tolstoy modelled his characters to some extent). We have a complex document (and its textual development) as well as any hints we can glean from Tolstoy’s remarks and correspondence. We know something of the social background, and of the commentary, challenges, and ideas Tolstoy was offering to the public. The answer to the question can be taken in different ways. We could be asking what Tolstoy was offering as a reason for Anna’s love, or we might study what the text as presented shows regardless of Tolstoy. The significance of the answer does not lie in the analysis of a real-world event like the cardiac arrest, or of a real-life love affair. It lies in the projection of the novel onto the society of the time and onto universal human relations. It could lie in Tolstoy’s “message” in the novel, or in our interpretation from a given perspective. Whatever interpretation of the question we take (and of course there may be many others), we still form a construct of ideas involving the phenomena of the novel; the connections between characters and circumstances; and the sequential processes leading to Anna’s elopement. There has to be correspondence of our representation with the “facts” of the novel and the connections and processes, and there must be a logic that “stands to reason” as part of our construction of Tolstoy’s novels. We must feel that Anna’s love is clearly motivated, “realistic”. Any explanations we offer (a loveless

marriage of convenience, release of repressed sexuality, rejection of social convention, etc.) will not be causal, but motivating/dispositional in nature; and they are unlikely to be in any sense definitive or such as to channel thinking into the kind of establishment of relevant experiments to test uncertain points of theory, as in physics. In fact, the field of debate is open. Perhaps, we could say that in social and literary studies there is more emphasis on why and on reasons qua motivations, whereas in physical science there is more emphasis on how and physical-causal explanation. One would doubt whether a causal view of Anna's falling in love in the novel starting with Tolstoy's entire life history and showing the deterministic inevitability of the text would tell us very much at all.

In the case of legal reasoning, there may be (at least in many cases) a high degree of clarity. For example, we may have a law that it is illegal to drive a motor vehicle without a valid licence and insurance, and that infringement of the law is punishable in a prescribed manner. The logic in applying "theory" to the particular case is very clear: if someone is found driving without the relevant documentation, then he or she is guilty and subject to punishment. Such a law is a reasonable way of protecting other people, but the principle of such a law, and hence its use as a reason for punishment, depends on a decision by the relevant power (parliament or other body) and its acceptance by the public – some people may not accept the restriction on personal liberty or may have religious views which conflict with the whole idea of insurance. The application of an arbitrary legal requirement is clearly not a causal matter, but a matter of convention (and definition of terms). That becomes even more obvious in the case of more controversial laws (and punishments). For example, the laws of *lèse-majesté* in Thailand or the laws relating to treason and espionage in the old Soviet Union would be regarded as extreme and unacceptable restrictions on freedom of speech and action in many societies. Those freedoms themselves are matters of human conventions and definitions; all societies must find a balance between personal liberty and social responsibility by agreement. That too, is a matter of a non-causal construct.

The search for reasons is part of everyday life, not just a matter of scientific, medical, professional, or academic debate. We can ask why a man might drink Glenfiddich rather than any other whisky, why a woman chose to buy, say, a blue dress rather than a green one, why she married John and not Peter, why the fruit bowl has not been replenished, why there are crumbs on the floor, why one should invest in bonds rather than stocks (or vice versa), and so forth. The answers to such questions involve reasons concerned with personal preference (taste in whisky) or opinion (about her appearance or colour likes/dislikes), priorities (in finding a husband), the set of circumstances leading to the observed state of the fruit bowl or carpet, and attitudes to risk and return in investment. Of course, it may be that the decisions and circumstances, preferences, priorities, and attitudes are somehow determined by a long chain of events and psychological forces (or may be whimsical or inscrutable), but what we present as reasons are mental constructions which we assess as reasonable or otherwise – they belong in a human mental framework and tell us more about how we think than about particular decisions or choices.

In legal and everyday contexts, an explanation may also be offered as a justification (a clear category confusion). For example, a speeding driver may plead the necessity of getting to a school to pick up the children on time or a neighbour might explain/justify loud music as part of a birthday party. One might counter-argue that such considerations do not constitute a reason (= justification) for endangering others or disturbing the neighbourhood. That is, in the special case where an explanation is a proposed justification, it may not be a reason (at least for other people). A failed explanation (such as in the case of the speeding driver or noisy neighbour) is

not a reason, that is, the failure involves the lack of a sufficient reason (due to wider considerations), as is also the case of scientific or academic explanations which are not accepted.

Conclusions

To sum up, explanations and reasons are connected notions in our reasoning. We have seen that some explanations are not causal reasons, but involve social and semantic conventions (e.g. legal explanations); in other cases (e.g. language change) we can offer explanations, but we do not know whether the proposed factors are reasons (i.e. they describe events or outcomes) but we cannot know if they constitute actual real-world motivations for human behaviour; and in yet other cases (literature) explanations may not be unique reasons, but instead may appeal to our sense of reasonableness for the purposes of discussion – or they may just be thought-provoking alternative ideas. It seems that explanations and reasons can be differentiated into types.

Our patterns of reasoning seem to work in similar ways, but in science the pathway seems more clearly defined, and has fewer theory options. Proposed explanations can be tested in relation to the phenomena, relations and processes in question. Progress can be made in explanation and in detail. While scientific explanations are almost exclusively causal and physical, we have seen that they require an appeal to common experience and to conventional definition of terms, where quantification is not a central concern. They also exist in a framework of human rationality and presuppose the conditions of reasoning. In other areas of our understanding, explanations involve concepts which are clearly not causal (e.g. survival, adaptation). There is still an appeal to common experience or reasonableness. Furthermore, in our rationality there is a need for some sense of certainty. There may be competing scientific explanations. For a long time, there was dispute between upholders of the “steady state” view of the cosmos and upholders of the “expanding universe” view (currently the favoured view). At the moment, some unexpected observations about the stability of galaxies and galaxy clusters (e.g. why they do not spin apart) are accounted for by the idea of “dark matter”. That explanation is disputed by those who invoke a modified view of gravity or those who think the computations are wrong (Clegg, 2019). That is, there are different explanations. Without greater knowledge, we cannot be certain about any explanation and, therefore, cannot speak of definitive reasons for the observed phenomena.

In particular, it is important to resist the strong temptation to equate a successful theory with reality (i.e. be an ultimate reason), rather than to see it as a best explanatory method. Newtonian physics had to be replaced by Einsteinian physics. As Levenson (2015) describes, some scientists were convinced of the existence of a planet, Vulcan, between Mercury and the sun because such a body could explain variations in Mercury’s orbit in a Newtonian framework – some even claimed to have seen it. More accurate information rendered the Vulcan hypothesis unnecessary.

Non-quantifiable concepts require definition, and thus involve our notions of meaning. “Semantically based” reasons, as we have seen, include not only common experience or reasonableness, but also preferences, priorities, attitudes, points of view. Reasons include social or political perspectives and correlations which are open to debate. In both (physical) science and non-science, reasons and explanations exist in the framework of human rationality, and in wider constructions, which are not always the same for all, but which are all “model-dependent”.

A significant issue in the classification of reasons is the principle of “same cause – same effect”. While in physical science, many effects can be explained by the same causes, and any different effects raise questions for research, it should be clear that in other disciplines we may rarely find “same causes”, let alone “same effects”. The events and circumstances of history, language changes, literature, the making and application of laws, and everyday events may have similarities, but they each require separate explanations as both the causes (or relevant circumstances) and effects vary.

What we have seen is that reasons and explanations cover a range of possibilities from the strictly causal applying to all cases through those involving a reliance on the definition of terms to those which concern only a single everyday event or set of circumstances. All require mental constructions (and sometimes a “paradigm” of thinking), but the conditions for the acceptance of an explanation differ.

If we can find a difference between reasons and explanations, it may lie ultimately in the emotional attachment to why questions (for reasons), and the desire to reveal “the truth” or to achieve a sort of ultimate understanding, possibly illusory – a desire for belief. Explanations answering how questions may sometimes seem prosaic or unfulfilling by comparison. (An autopsy report may not satisfy someone on questions of human mortality and destiny.) Curiosity demands reasons and more reasons. We want to know, for example, whether or not our existence is “real” or the product of a hugely complex quantum simulation, or – as Berkeley would have said – whether we and all existence are ideas in the mind of God- is there any way of telling?; whether we could tell the difference between a rational, conscious being with free will and a zombie who acted in every respect like a rational, conscious being with free will; did the conditions of the big bang determine everything that has happened in our universe? As we have seen, explanatory reasoning requires both a model of the “explanandum” – the phenomena, their relations, and the sequence of processes – and a logic containing general premises accepted as true or reasonable with a deductive argument to show that the phenomena and their relations are accounted for by general principles, along with a recognition of the applicability of the model to the specific phenomenon under consideration. Reasoning further involves the norms and conditions of human rationality, as well as conventions of meaning, as we have seen.

We might propose, then, that an explanation is an end point. It is the application of reasoning to account for a set of phenomena, events, circumstances, and their relations and processes. A reason is a starting point. The focus in a reason is the general premises and logic of the reasoning from which we start an explanation. That is why failed explanations are not reasons; the failure lies in the general premises or the lack of correspondence between phenomena and general premises. A reason is also the starting point for further why questions, that is, the search for further reasons. That is what underlies the emotional content or connotation of a reason as opposed to an explanation.

Our discussion does not resolve the question of whether the universe is deterministic – in purely physical terms it may be, but those are not the only terms that are relevant to us as humans. An acceptance of physical determinism does not imply the acceptance of purely explanatory determinism. As Hempel (1974) points out, commonality of reasoning does not imply an entirely mechanistic universe. What we have seen is that all our reasoning takes place in a human perspective (the social and psychological universe of our experience), and reasons are what we construct and accept within the limits of our capacities, often from multiple perspectives- our mental and scientific models. Part of our reasoning is semantic – and, hence,

a matter of social convention. However reliable our models may be in giving reasons as general principles and leading to acceptable explanations, they are subject to change or debate, and so do not give ultimate reasons. Reasons and explanations are not the same things.

References

- Berkeley, G. (1910). *A treatise concerning the principles of human knowledge*. Open Court. (Original work published 1710)
- Clegg, B. (2019). *Dark matter and dark energy*. Icon Books.
- Darwin, C. (1901). *The origin of species*. Murray. (Original work published 1859)
- Dasgupta, S. (2016). *Metaphysical rationalism*. *Noûs*, 50(2), 379–418. <https://doi.org/10.1111/nous.12082>
- Dunbar, R. (1998). The Social Brain Hypothesis. *Evolutionary anthropology*, 6, 178–90. [https://doi.org/10.1002/\(SICI\)1520-6505\(1998\)6:5%3C178::AID-EVAN5%3E3.0.CO;2-8](https://doi.org/10.1002/(SICI)1520-6505(1998)6:5%3C178::AID-EVAN5%3E3.0.CO;2-8)
- Ewing, A. (1951). *The fundamental questions of philosophy*. Routledge and Kegan Paul.
- Hawking, S. and Mlodinow, L. (2010). *The grand design*. Bantam.
- Hempel, C. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. Free Press.
- Hempel, C. (1968). Explanation in Science and History. In P. H. Nidditch (Ed.), *The philosophy of science*, (pp. 54–79). Oxford University Press.
- Hjelmlev, L. (1969). *Le langage* (M. Olsen, Trans.). Minuit.
- Hume, D. (1968). *A treatise of human nature*. (Vol. 10). Everyman. (Original work published 1738). <https://doi.org/10.1093/oseo/instance.00046221>
- Levenson, T. (2015). *The hunt for Vulcan*. Random House.
- Locke, J. (1973). *An essay concerning human understanding*. Fontana/Collins. (Original work published 1689). <https://doi.org/10.1093/oseo/instance.00018020>
- Penrose, R. (2010). *Cycles of time*. Bodley Head.
- Plato. (2007). *The republic* (H. Lee, Trans.). Penguin.
- Popper, K. (1972). *Logic of scientific discovery*. Hutchinson.
- Popper, K. (1972a). *Objective knowledge*. Blackwell.
- Poston, T. (2014). *Reason and explanation*. Palgrave. <https://doi.org/10.1057/9781137012265>
- Spinoza, B. (2002). *Complete works* (M. L. Morley, Ed.), (S. Shirley, Trans.). Hackett.
- Strawson, P. (1973). Introduction. In P. Strawson (Ed.), *Philosophical Logic* (pp. 1–16). Oxford University Press.

Corresponding author: Paul Rastall

Email: paul.rastall@googlemail.com