Levels of explanation

At a large vision meeting I was repeatedly struck by the seemingly unbridgeable chasms between people’s different understandings of what counts as an ‘explanation’ to them. This is a most curious phenomenon when you come to think of it, because almost everyone will concede that gaining (some form of) ‘understanding’ is for a large part what our endeavours in science are about in the first place! Here is an example. A friend remarked to me: “now colour vision has been finally solved!” He was referring to our recently much increased understanding of the nitty gritty details of the genetics of photopigments in primates. No doubt, he was quite sincere. However, such a remark strikes me as nothing short of hilarious. Indeed, I would be hard put to state in a few words what would count as a solution to the problem of colour vision to me. But certainly (again, for me!) the details of photoreceptor genetics would play only a minor role. When I start to think of problems touching upon colour vision, I find it hard to round up the list. New problems keep popping up in my mind all the time as I muse on the matter. For me, most of these tend to be of an experiential nature. I’m sure my friends in biology, for whom all animals seem to come in stopped bottles nowadays, will similarly come up with ranges of problems from the molecular domain, my—more trendy—friends in psychology will think first of fMRI-related problems and my artist friends will spontaneously think of yet different problems altogether.

One aspect of this is that ‘explanations’ are much like infinite sets of Chinese boxes or Russian dolls. I mean that you approach the world at many different levels simultaneously, and that explanations (have to) look different at all these levels. There are levels in various dimensions: just think of spatial or temporal detail, combinatorial complexity, or physiological detail, for the moment. A scientific discipline where this is quite striking is physics. Physics is a great source of examples of the Chinese-boxes-nature of explanation. Strange enough, most physicists completely misunderstand the nature of this. The reason is that the various levels of explanation tend to be identified with historical developments. This has led to the common notion that explanations at the more detailed levels (these tend to be the latest) are somehow more ‘fundamental’ and have actually replaced explanations at the more global levels, thus—of course, much to the physicists’ satisfaction!—illustrating our advance of scientific understanding of the Universe. That this is little else but make-belief becomes obvious when you study the way physicists and engineers approach actual problems. I mean problems that occur in settings other than those at the ‘frontier’ where experts narrowly focus upon the most detailed level and try to eliminate the other levels wherever possible—a near impossible task that is part of the make-belief. Take almost any problem of an optical nature for instance. The engineer will freely use the tools from any appropriate level, ‘changing gears’ all the time without so much as thinking twice about it. It would verge upon madness to religiously stick to the latest theory “that contains all previous ones as simple corollaries”, because then the job would never get done. Thus one uses the (ancient) geometrical ‘ray’ theory (Euclid’s optics with some 16thc. additions) to construct the image of a lens, Huygens’s (17th c.) and Young’s wave theory (19th c.) at the diffraction limit, Maxwell’s electromagnetic theory (19th c.) to deal with the modes in fibers or photoreceptors, quantum theory (early 20th c.) to deal with the interaction with matter (rhodopsin say), Lambert’s statistical ‘ray counting’ (late 18th c.) to handle the radiometric problems, and so forth. Not only is this the
common way such problems are handled, for all practical purposes it is the only way to handle them. So what explains the problems the practitioner in luminous matters, such as the photographer, deals with (say)? Definitely not quantum electrodynamics—nor do quarks or gluons play any role in the photographer’s understanding. The photographer may look like the prototypical cave man to the (theoretical!) physicist, but he will have the sympathy of the ‘hands on’ physicist or the engineer that deals with real problems, though. I know, I’m a physicist myself.

What actually happens in physics is that the various theories apply to different strata of spatiotemporal detail or combinatorial complexity. The various theories are most useful in the strata which they ‘fit’ naturally. In the limit of the finest resolution they don’t apply that well, and the theory at the next level does the better job. Yet it would be a mistake to hold that this “proves the theory wrong”. The theoretician tries to show that the more detailed theory actually accounts for the coarser theory and thus may replace it. Cases where this actually applies are much rarer than commonly thought, though. Typically this may be true in a shallow transition stratum, the finer theory being too unwieldy a tool to wield throughout the coarser domain. Rather the opposite is the case. When one has a problem in a wider domain, one tries to forge coarser theories that may be applied effectively, the regular theory being like a microscope to view an elephant. In order to ‘understand’ a phenomenon, the relation to the explaining theory—that is the bridge to the ‘understood’ concepts on the next finer level—has to be sufficiently short to hold up before one’s mind’s eye in one piece. Otherwise, one merely has a ‘computer simulation’ (say) that shows that the theory may account for the phenomena in some essentially not understood way. Indeed, the modern possibilities of computer simulation have blurred our very understanding of ‘understanding’. One often witnesses simulators (people—I mean—not machines) developing rules of thumb to somehow understand (intuitively) the fruits of their endeavours. Such are actually higher-level theories that are indeed necessary if one wants to avoid having to simulate at random, in any blind fashion. To me, it seems best to be perfectly honest about what goes on here. These higher-level theories are way more important and interesting than the results of the number-crunchings themselves. Thus what counts as ‘understanding’ has much to do with the very limits of the human mind; one has to be able to perceive the connection between the phenomena and the theory. Understanding is perception (or—perhaps—a ‘vision’). After all, the Universe itself is the most reliable (and perhaps most relevant) simulation available to mankind. It is continually run for us for free. If understanding the results of the simulation (nature itself in this case!) is the goal, we should come up with the right rules of thumb at our level, rather than try to go for the action of the wheels at the innermost parts. That is a completely different kind of game.

One often distinguishes between ‘explanation’ (pretty much in my sense) and ‘ultimate explanation’. What counts as an ultimate explanation depends on the era. It used to be something like a proof that this is the best of all possible worlds. Nowadays it is more like a proof that He has no choice in the design of the Universe, thus implicitly eliminating Him as a causal factor. In all cases this means getting at the action of the innermost wheels. Some people, like my colleague Gerard ’t Hooft, who recently received the Nobel prize in physics and is in one of the best positions to know, hold the view that physics will be a closed section within a few years, and sciences like ours (perception, again) had better take it easy till then, because everything will have been solved as a consequence. After all, all the sciences are simply applications (easy or perhaps mildly complicated corollaries) of physics. Personally, I don’t buy the argument, because I believe that many phenomena that interest me don’t have an explanation at the level of the innermost wheels at all. For instance, the insects hold quite a few visual competences in common with us (people), but the actual implementation at the neural
level has to be different, simply because their brains are so differently configured from ours. And what if our neurons were replaced with silicon substitutes? I feel it wouldn’t make me any different if it were done expertly, like a hi-tech wooden leg might go unnoticed by the user. Thus my explanations of the phenomena of perception cannot depend on the structure of the stratum below that of the neuron (molecular, or certainly atomic level). The quarks and gluons are irrelevant to my mental makeup.

In cases like the simulations mentioned above, one often talks of ‘emergent properties’ when surprises arise. Such properties are said to be (miraculously!) ‘predicted’ (or accounted for) by the theory. The point is, though, that they cannot actually be predicted from the theory at all—they are ‘discovered’. Often this fact goes unnoticed, because one actually starts with observations at the coarser level and ‘predicts’ the (indeed) surprising properties in retrospect. Starting from the latest theories of physics, such minor manifestations of matter, as for instance the oceans (I’m hinting at the fluid state), would probably escape attention, unless some random calculation happened to display such an unexpected behaviour and someone would actually be struck by that. The reader will notice that I poke fun at reductionism here, one of the generally acclaimed pillars of the ‘scientific method’. Emergent properties are like hat tricks: the actual relation to the theory goes unperceived, as when a magician pulls the rabbit out of the (in our present understanding) empty hat.

Our field (I mean perception, of course) is so interesting because we have to deal with levels of explanation that the physicist (perhaps wisely?) chooses to ignore. I mean the relation between the psychical and the physical, of course. Here is a broad source of differences in the implied meaning of understanding. Some psychophysicists understand perceptual phenomena in terms of other psychophysical phenomena. For instance, they may say that this or that phenomenon is “due to assimilation” and by ‘assimilation’ mean a group of perceptual phenomena commonly referred to as assimilation phenomena. ‘Inhibition’, ‘adaptation’, and so forth, are used in similar ways. I am now downplaying this (one might suspect a circulus vitiosus here), for an intuitive grasp of perceptual phenomena is no mean thing. Others may have general neurophysiological phenomena in mind when they use the term ‘assimilation’—some kind of lateral interaction or ‘spread’ at some abstract neural level, say. Others still may have a very specific neural substrate in mind, say the retina or the primary visual cortex. Yet others will only agree to call this an ‘explanation’ if the actual electrochemistry at the synapse level has been mapped out. And so forth. One may also find many colleagues who look for explanations in terms of ‘meaning’, either on the mental level (it seems obvious to them that leaf-shaped gray blotches will look more greenish than donkey-shaped ones), or on the level of ecological optics (they think it obvious that blotches of smaller angular extent must look more distant than larger ones). What makes discussions at large vision meetings so appealing to me is that people of all these inclinations and convictions are freely mixed up and engage in heated discussions without even bothering so much as to identify their position—after all, everyone is wearing a similar badge around the neck. This must be exceedingly confusing to newcomers in the field, especially when they (as is indeed likely with fresh students) have never thought about levels of explanations to begin with. When you have grown gray hairs and essentially know all the players from previous meetings (like myself) this is a great source of amusement though. One person’s ‘explanation’ is another person’s bogus. Such convictions, of course, also show up in the experimental approaches taken, and indeed the very problems taken up, by the various people.

One type of defence that works is to jump on some bandwagon. Then you can be sure of any number of people who will nod in appreciation of your arguments and will do what they can to make you feel you’ve hit on an important problem and well on your way to further our understanding of it. The downside is that the same crowd
is constantly recalibrating the established pecking order, and as a newcomer you will be directed to the tail of the queue by all the standard means of our scientific community. The virtue of this is that everybody knows (or at least acknowledges) what will count for an 'explanation'.

Since this is perhaps too sad a key to end an editorial, let me change gears at this point. So what may count as an ‘explanation’ in (our) science? I hold the view that any specific answer is thereby wrong, and indeed detrimental to the cause of science, because it (by implication) debunks alternatives. I feel that any explanation can only be enriched by explanations at other levels. Explanations are right when they work, and wrong when they don’t. They are more or less useful to the extent that they allow me to have a clearer view of the causal nexus underlying the phenomena. A clearer view will be more fit to let me conceive of other applications, analogies, and consequences. This goes parallel to other measures of the ‘truth’ of scientific concepts. For instance, a theory that is more formally elegant or more economical in its description is more likely to evoke a clearer image in me. Notice that this view does admit the possibility of having different theories of the same phenomena that are equally ‘true’. Such cases occur in physics. For instance, there are two different versions of the theory of interaction of photons with matter. In one theory, photons are emitted by a source and—by sheer accident—happen to become absorbed by a target. In the other view, the transmitter (an electron in the Sun’s photosphere, say) engages in a transaction with the target (a rhodopsin molecule in your eye, say), then the emission and consequent absorption happen. No accident here. Since all consequences of the theories have been proven identical, you are quite free to take your pick. It seems to me that there is much room for such alternatives in our field. The main thing is to keep the various levels of explanation apart so we can at least interact in productive ways with our friends. When my colleague thinks colour vision has been finally solved that’s fine with me because I know what he or she means. It certainly won’t keep me from attacking what I believe to be the outstanding problems in that field. I expect my colleague to listen when I’m ready to report my results, of course.

Jan J Koenderink
Physics of Man, Helmholtz Instituut, Princetonplein 5, NL 3584 CC Utrecht, The Netherlands;
e-mail: j.j.koenderink@phys.uu.nl

© 2002 a Pion publication