Explanation and Causality: a List of Issues

Alberto Peruzzi[†]
alberto.peruzzi@unifi.it

ABSTRACT

After a concise description of issues concerning the causal and the deductive-nomological models of explanation, the flaws in the alternative view centred on relevance-to-context are examined. The paper argues for the need of a wider spectrum of options which takes into account both the Local/Global and the Internal/External aspects in order to determine the sense and the adequacy of any explanation. As a test for this argument, some specific problems are considered about the range of causal bonds, the admission of top-down causation, the appeal to emergence, the shift from explanation to explainability, the equivalence classes referred to as "cause" and "effect". Finally, the paper deals with the comparison between inequivalent explanations and lists three remaining issues to complete the picture.

Keywords: causality, explanation, deductive-nomological model, emergence, relevance, ILGE-semantics

1. Pasting Chessboards

Intuitively, we can say that an explanation is an answer to a question of the form *Why ...?* and, no less intuitively, we can say that any such answer specifies the cause of ... But this specification can be of different sorts. Aristotle distinguished four possible senses of a why-question and thus four kinds of corresponding answers, each providing a species of causes: formal, material, efficient and final. These four species cover our practice of explanation in everyday life and could even be extended to cover the behaviour of entities such as computer programs, flow diagrams, models, concepts, propositions, street signals and, for what concerns the fourth species, caution would sug-

[†] Università di Firenze.

gest the "final" is limited to the teleonomic behaviour of living systems.

The phenomenology of all such uses of the notion of cause is indeed a remarkable endeavour. A phenomenological taxonomy of causes is, however, inadequate for understanding how they are connected to one another within the common-sense world. But explanations also extend beyond the bounds of what we refer to in our ordinary experience of the macroworld. With the Scientific Revolution of the modern age, the tensions between any such taxonomy and the progress of research became heightened. Two issues were progressively focused upon: one concerning the "modal force", so to say, of the link between cause and effect, and one concerning final causes.

Modern empiricists deny the supposed necessity of the link between cause and effect, for it can't be justified by any finite amount of evidence, and the mechanical paradigm of modern science denies the need to refer to final causes, which had its motivation in thinking of nature as an organic entity. If form is reducible to matter and the only substance is material, chemistry and physics together explain how any piece of "substance" empirically accessible is composed according to laws; in the end we are left with efficient causes alone, intended as localisable sources for the combined action of the elementary constituents in any given system present in nature. It is only this combined action which is responsible for any effect. Organisms were later recognised as being no exception, since evolution based on natural selection only requires paying attention to statistical distribution, and in the light of the "synthesis" with molecular biology, we should say that efficient causes rule the world.

Even so restricted to one species, the notion of cause suffers from the previous inconvenience, since the necessity of the link between cause and effect is absent from evidence, and another link has to be added: cause is also absent from the laws of physics, which are equations with no preferred time-order. If physics is the model of natural science, this is more than a side issue, unless one is prepared to say that reference to causes is legitimate in scientific areas where no laws have been found so far or where the laws (any guesses?) are non-equational and time-dependent.

It has always been clear that precedence in time is insufficient for identifying a causal link, but with relativity theory time-order is absolute no longer. In relativistic spacetime it is not possible to separate time from space and "simultaneity" is a notion relative to a local framework. However, this is helpful in further specifying the idea that cause precedes effect, as now there is a boundary to the range of possible lines of causation by excluding the space-

like ones, although the past light-cone of any event-point still contains a lot of possible causal lines, with a backward ramification of other light cones.

The indeterminacy which entered physics through quantum mechanics provides a third inconvenience and non-locality adds a fourth one.

If our appeal to the notion of cause arises from the need to explain, could the job be done without the notion, thereby bypassing the inconveniences of causal talk? The deductive-nomological (DN) model, see Hempel (1965), presents the explanation of an event *E* as a purely logical derivation of *E* from the conjunction of a finite set *C* of specific factual conditions *CI, ..., Cm* and a finite set *L* of laws *LI, ..., Ln*. Rather than describing a single event, *E* might be a quantified sentence and thus correspond to an empirical generalisation or even a low-level law about a particular kinds of events covered by *L*.

According to this model, explanation is a game unifying explanation and prediction. In this game, three slots *E*, *C* and *L* are to be filled. If no slot is filled, the game does not start: you have nothing to explain or to predict. Otherwise, the game commences and, in order to win, you have to fill any empty slot.

Explanation corresponds to when the *E*-slot is filled at the start, prediction to when it is empty at the start. Suppose two slots are filled. Then there are three possible cases: 1) if *X*, classified as of type τ, fills the *E*-slot and you know the laws to be applied to events of type τ, you have only to fill the *C*-slot, by identifying the specific conditions for *X*; 2) if you know what is in the *C*-slot in addition to *X*, you have to discover the laws; 3) if you know what fills the C-slot and the *L*-slot, the game is a predictive one: it does not matter whether the "prediction" is made forwards or backwards in time (i.e., if the event has yet to occur or it occurred in the past – in such a case the term "retrodiction" is also used). Since the nature of a prediction depends on the (closed or open) status of the system under examination and the form, probabilistic or not, of the prediction, it is already clear that some qualifications are needed – more on this below. Yet, no induction is required: to win the game, you have just to find what makes the case for a deduction.

Moreover, the model does not suppose the premises are (known to be) true, it only matters that the conclusion can be correctly deduced from them. So it sharply separates explanation from truth and in fact there can be more than one way to fill two slots given the third one already filled. When the *E*-slot is filled at the start, the selection of which way is "better" than another will rely on further (supposedly independent) information about each candidate for filling the *C*-slot and the *L*-slot, which is part of another slot-filling game of the same sort.

In Peruzzi (2009) I argued that, so conceived, the game is compatible with some alternative models that have been proposed, by suitably transforming them into extensions of the DN model, each adding a constraint on how the empty slots should be filled. Indeed, the model seems to be inadequate, unless you are an adherent of pluralism who doubts that there is a chance of determining how *any* slot should be filled and is pleased to emphasise that it is merely a matter of conventional stipulations, and your pluralism extends to the form of the laws: be they equational or not, first-order or second-order, deterministic or not, local or global, context-parametrical or context-free, inclusive of the observer or not. If you are not such a pluralist, the issue is how to decide which extension is correct. Moreover, the determination of which convention is "better" than another could in principle become an objective issue if the way the "explainer" works is brought into the picture, although it must be admitted that "explaining the explainer" would be a complex game.

Since the DN model eliminates any reference to causes, if there are explanations in science and they can all be covered by the model, the first two inconveniences no longer produce philosophical trouble. As for the others, however, an adaptation of the model is called for in order to deal with inferences employing statistics and *a fortiori* probability theory (not only quantum indeterminacy but also uncertainty as in social sciences) as well as with non-localisable conditions.

In fact, the model is complemented by a statistical-inductive format but the result makes an explanation semantically non-homogeneous, since some sentences within one and the same inference remain two-valued while other sentences are assigned a probability value within the interval [0,1]. This would demand a logical analysis which seems not to have been carried out so far. If such a mixed format is actually used in scientific practice, the adapted model could also be considered adequate in practice, but, to put it euphemistically, the model would fall well short of providing a paradigm of rationality. Semantic homogeneity requires that the slots are filled either by three probabilistic statements or by three non-probabilistic statements (the latter case does not prevent the use of probability within each slot).

Unfortunately for the empiricists, the model, thus adapted or not, still needs to distinguish a law, intended as a universal assertion of a nomological character (thus modal and, specifically, counterfactual) from a merely accidental generalisation; and for *logical* empiricists the only safe notion of necessity should be inherent in what belongs to the logico-linguistic framework.

Despite the fact that an empiristically palatable way can be found to avoid

having to saddle the model with a commitment to an obscure more-than-empirical necessity, counterexamples have been formulated which show that reference to causes cannot be entirely eliminated: asymmetries associated with the order of time indicate that the indifference in taking a deduction as an explanation of a known fact or a prediction of an unknown fact (be it past or future) cannot be maintained, see Salmon (1998). The counterexamples do not require commitment to a thermodynamic arrow of time but are related to local causal asymmetries. If this objection is acknowledged, one can take vantage from the window of causal lines in relativity but quantum indeterminacy and non-locality remain to be matched. Henceforth, in order to show that there are enough issues concerning explanations independently of this matching, I shall omit reference (apart from some quick remarks) to such a gap.

2. Relevance and Context

There is also another, independent, objection to the DN model. This is based on the requirement that all premises must be *relevant*, whereas the monotonicity condition (if A implies C then the conjunction of A with any B implies C) turns a given explanation into a redundant one, thus violating the requirement and allowing arguments which we would reasonably take as incorrect explanations. This objection, however, might simply suggest that the DN model requires a logic different from the classical one, while remaining no less rigorous than are inferences in any relevant logic.

There is a price to pay for this reply: if the task of replacing so many well-established causal explanations with DN-inferences is already difficult, that of reformatting scientific reasoning from top to bottom within relevant logic, without loss, is no less difficult: the logic implicitly used in the mathematical theories which provide conceptual frameworks and models for much of science is not relevant logic. Moreover, there are many non-equivalent relevant logics. It looks suspiciously *ad hoc* to end up with exactly as many relevant logics as different scientific areas matching different inferential practices with no intrinsic motivation. This is neither an indispensability argument against *any* change of logic nor a preparation for a "cumulative" view of science: it only indicates the cost of *that* option. First, the target of the objection is wider than the DN model; second, in consideration of the price to be paid, one could try to deal with empirical relevance not so much in terms of logic but rather by a different kind of relevance, to be added to the model.

Note that the objection is stronger in the case of a causal inference, where *A* describes the cause and *C* the effect, as *B* might describe the action of another cause which interferes negatively with the cause described by *A*, so that the effect described by *C* is not preserved. As is well-known, monotonicity is not valid for counterfactuals; therefore, also in the absence of any reference to causes, it is possible to extract a different charge to the DN model: the explicit logic of slot-filling is, by default, classical while the implicit logic of counterfactuals, needed to define the nomological character of what fills the L-slot, is non-classical. So, once again, the lesson could be not that DN model is wrong, but only that is too loose and logically non-homogeneous. As for this latter notion, one might respond by noting that if modalities are introduced on top of a non-modal system as it happens in ordinary modal logic, it would suffice to use different symbols for modal and non-modal implications. But even so, we should agree the model needs to be enriched with other principles which suitably constrain the range of inferences.

Now, if the fault of irrelevance is not given to logic, to what else? It could be something concerning content rather than form. But content has two faces, one subjective and one objective, so there are two options: either relevance is subject-dependent and context-laden or relevance is a result of an objectively adequate model, correctly "insulated" (so to say), of the intended system to which our explanatory inferences refer.

The first option brings us back to the link between scientific (formalised) language and ordinary language and to the need for understanding the way they communicate, in particular the concept of explanation, rather than taking the distinction as that dichotomy emphasised by analytic philosophy in the past. As already noted, the translation's task involved in analysing current explanatory talk of causes in terms of cause-free sentences is far from being simple. This is because it cannot be confined to the outcome but needs to go through a step-by-step process of justification. But how wide has this process to be? Even if we are concerned with saving the phenomena, i.e., if the task is intended to furnish an analysis which makes people's answers to why-questions as "rational" independently of their matching what we take as correct scientific formulation, our analysis could be indifferent as to whether they share our "charity" criterion, whether they understand our need of a reformulation , and, more to the point, whether our explanation of why they take an argument as an explanation is accompanied by their taking what we take as an explanation as an explanation for them too.

But if we start being concerned with this, there are consequences which affect any model of explanation we may have adopted, for care should consistently apply to any case in which there is a contraposition of explanatory perspectives. The first lesson of such unlimited "concern" is: abandon the idea that explanations are objective and accommodate only a subjective, audience-oriented, history-sensitive, notion of explanation, which is necessarily relative to a context. To be consistent, this pragmatic lesson applies to any explanation, no matter what its subject matter is: from physics to neurosciences, from strategy games to common sense psychology. The bonus is immediate: the demarcation between ordinary and scientific reasoning turns into a grey area which can be traversed without cognitive shock, apart from the time taken to adapt to new contexts. Someone could dare to claim that the lesson applies to philosophy too, however borderline "philosophical explanations" may be. Thus the above concern self-applies, which is possibly an *objective cause* (!) of embarrassment, namely, the embarrassment felt by that unfortunately self-concerned barber who shaves (≈ provides a trans-contextual explanation for assertions made by) all those who don't shave themselves (≈ provide only context-sensitive explanations).

No wonder philosophers who like the lesson of "concern" do not apply it to their own arguments: there is a price to pay for the bonus and consistency suggests that it must be assessed as a high price. Whoever fails to be convinced that such a lesson is final, must hit the nail on the head and search for an objective account of explanation, now inclusive of reference to the rationality of agents, with their own set of beliefs, and also capable of incorporating our ordinary talk about causality in every-day language. There are already some devices that can be used to achieve this: e.g., the cognitive development of the notion of cause as investigated by psychologists, subjective probability as mathematically formulated, rational decision theory as applied in economics, epistemic logic as a probe for revealing the consequences of combined (iterated, in particular) operators; cognitive neuroscience and evolutionary biology are further resources. *Wir wollen wissen, wir werden wissen*, Hilbert claimed. Yet a satisfactory account of explanation may require less than the completion of such a huge puzzle; in fact it must require less, for such an account is more than a further device: it may interfere with other devices in various ways. Perhaps there is something we have not considered so far.

One unexplored issue concerns the explanation of what happens when an emergent structure has or seems to have a causal role in a top-down dynamics. This issue is also linked with the meaning to be ascribed to common talk about

the effects of mental states on physical states and yet the issue is more general. Rather than dealing with a particular and arguably the most demanding instance of a difficulty, it is reasonable to face the difficulty in a wider setting, starting from its simple occurrences (already demanding) and passing to the less simple ones.

In the following sections, I shall briefly list some issues concerning explanation and causality. Each is the source of problems and, before proposing or endorsing a solution, it is useful to consider the options for solving them. Provided we seek a unified picture and are not content with a mere collage, the task is to find solutions which are jointly consistent. Facing this task has an impact on each issue, but contrary to what this might suggest, I am not implicitly endorsing holism.

3. The ILGE Framework

Independently of allowing, or rejecting, reference to causes, any inference intended as an explanation is composed of sentences supposed to have an intended meaning relative to an intended domain, say *D*. This supposition seems to suffice to eliminate semantic doubts which may shift the focus of attention to the very notion of explanation. Nevertheless, this supposition still allows the meaning of any sentence (or set of sentences) to be dependent on two semantic parameters, namely its being of local or global character and its being internal or external to *D*.

We say that a sentence is of local character if its truth-value can be determined by considering in *D* a suitable neighbourhood *U* of a possibly finite subset of constituents, where *U* inherits the structure of *D* (if the sentence is universally quantified, the corresponding instances in that finite subset have to be generic). A sentence is internal if its value is established only by means of informational resources (about *D*) to which only an observer ("explainer") belonging to *D* has access. It is essential to take into account whether each sentence in an explanatory argument is claimed by an observer internal or external to *D*. Thus one and the same sentence can have a different meaning for an internal or an external observer; and this affects the explanations which make use of the sentence. Global (G) is opposite to local (L) and external (E) is opposite to internal (I).

The pairs I/E and L/G, compose with each other, providing a spectrum of perspectives for viewing any explanation. If one requires that the structure of

an explanation must be homogeneous, then the premises and the conclusion must be considered as being relative to the same perspective. If each operator is idempotent, there are only four kinds of homogeneous explanations: IL, IG, EL, EG. It is a useful exercise in epistemology to describe the phenomenology and review the different meaning of these four perspectives. I started this exercise in Peruzzi (2002), while also providing a few examples of the four combinations in a general semantic context.

Most if not all discussions on explanation and causality omit the consideration of this spectrum of perspectives. Usually, one of them is implicitly used and it's no wonder that, if someone else uses a different one (no less implicitly), an endless philosophical debate will ensue. On each occasion of this sort, the implicit can be made explicit, but as long as a systematic analysis of the ILGE-patterns is lacking, no epistemological argument about explanation and causality can be said to be right or wrong.

4. Causal Bounds

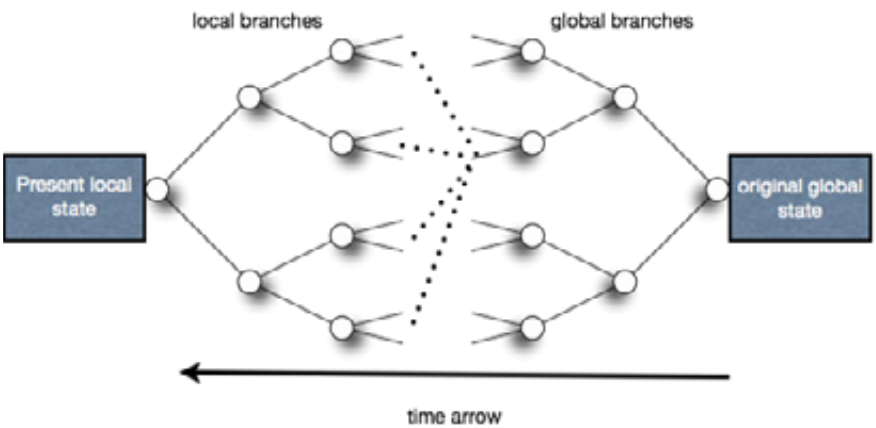
Suppose we admit causal lines between events in explanations. If there are long-range quantum correlations, the set of such lines is different from the set admissible in terms of only local action, "by contact" as Cartesians would say. But what is the trace of this difference on a macro, or cosmic, scale? The issue is clearly of importance for physics (and also for the way coherence appears in the brain). The bounds of independent causal lines, however, are not only at issue in relation to locality vs non-locality. So, let's take a look at this wider landscape.

For each event E in spacetime, one can go backwards along the branches of the tree of events within the light-cone corresponding to the E-past, and all points on any simultaneity-sheet in the past of E which have a causal line to some points on any intermediate sheet having, in turn, a causal line to E, can be considered causes of E. In fact, the relation "x causes y" is commonly taken as transitive. And yet, physics apart, in the ordinary use of explanations relative to an intended domain D of reference, composed of objects and processes at a given scale, we do not invoke transitivity when, in order to answering a why-question, the path from a cause to the cause of the cause of... extends beyond D's scale. E.g. if you ask "What caused my toothache?", I could answer "sugar" or "lack of brushing", but going further I can answer "bacteria" (feeded by sugar) and then, going backward in a transitive causal chain, I could also answer "descendants of the earliest organisms on Earth" and since the existence of their ancestors also

has its set of causes, and the same can be repeated for this set, through a systematic scale-transition we reach the formation of our planet, and so on.

Of course, we usually stop long before this. Relevance is demanding but it also saves time. The unlimited expansion of causal explanation would explain events quite different from the ones in which we are actually interested. Our mental scissors cuts the context from the background and we soon bring the explanatory process to a halt, although we are usually ready to admit a plurality of concurrent causes. We also tend to focus on one of them as "the" cause. When, among, say, 3 factors concurring an event we choose factor 2 as "the" cause, or the "main" cause, we assume factors 1 and 3 to be independent of 2 (so that factors 1, 3 and their conjunction can occur, and possibly we have evidence they already occurred, separately; and in such cases *E* does (did) not occur, or the frequency with which it occurs is much less than the frequency it has in presence of 2).

But there could be a context in which no such halting is at hand, so that we ask recursively what are the concurrent causes of each concurrent cause. Cosmology tells us that, in order to explain the present state of the universe, we have to go much further back than to the first bacteria on Earth, reaching a state of matter/energy which is about 15 billion years in our past. But there we find a common cause which links together 1, 2 and 3, whatever they could be; so, by going sufficiently far back, there is no triple of independent factors (the picture below is even more simplified: from threefold to twofold branching).



This picture can be scaled down, in space and time, from a cosmological background to a smaller one. There is a path linking any present local state to a global state in the far past. So, every backward causal branching has to meet every other in a sufficiently remote global state and this implies that any two (locally) independent causes have a common cause (globally). But, of the two opposite kinds of oriented branching, none would suffice to complete the path: we need both. Looking at a global state in the far past, because of indeterminacy, we cannot determine a path which ends precisely with *this* local state, and vice versa. Thus causal transitivity seems to *damage* our explanatory practice.

Once again, one might reply by saying that any explanation is relative to a context of epistemic interests, and this holds for cosmological interests too... were it not that the range of our interests, be they purely epistemic or not, as well as the range of what can be a "context" and any other arrow for the subjectivist/pragmatic bow, can in principle be explained by cognitive science. For, unless pragmatics is the result of a miracle and the "subject" an extra-cosmic intruder, cognitive science and its object are in turn made possible by evolution. So we are again back to square one. But since explanations provided/accepted in science (cognitive science included) are no less in need of contextual specification, the back and forth game (a metatheoretical fugue, so to speak) continues recursively with no winner – or, if you prefer, with the winner you like, once you fix a limit for the number of moves in the game and choose the limit most in accordance with your favourite (objectivist or subjectivist) standpoint. Standard philosophical debate usually identifies this limit with the number two, corresponding to the step from a-contextual to contextual or to the step from contextual to context-invariant explanations.

Putting the reply to one side, let's pause to consider the *damage*. Suppose we skip the selection of one cause among others as a sort of figure/background psychological bias, and take the whole set of E-concurrent events (on each sheet) as *the* cause. Then, since by going sufficiently far back we can find a causal line leading to whatever occurs 'presently' in, say, our galaxy, the only answer to a why-question about any particular present event E in our galaxy or, more consistently, about the simultaneity-sheet of E, should be "the local state of the universe at a previous time t" (in the past of E). This seems pointless. In addition, t could be any and, to cover indeterministic transitions, we could collect the sequence of states into one global foliation and answer "the global state of the universe at all previous times"! This looks like overkill. Note that the same answer should work for any other present $E' \neq E$, but this would fail to

acknowledge the efforts of researchers in explaining specific kinds of phenomena, as well as ignoring the success achieved by those efforts *without appealing to so large a body of information which nobody possesses.* Here, as indeed anywhere else, unqualified holism is the most comfortable open sesame, for it avoids any omission, and... the most uncomfortable, as it requires that we know more than what we doubt we are capable of knowing.

At best, the last answer might be taken as analogous to Quine's answer to the question "What is there?", namely "Everything". But the issue then reduces to a choice between the undoubtedly complete but pointless and the undoubtedly incomplete but useful. If we want to avoid this choice, we have to change something in what led to the choice, from neglecting ILGE-phenomenology to neglecting the fact that explanations are provided through language by explicators made possible by the emergence of local order. Thus we have to combine a twofold dialectics, one horizontal (ILGE) and one vertical (top-down and bottom-up).

Suppose the architecture of a system is organised into a set of levels, each endowed with a specific kind of interactions so that the state space of the system is different from one level to another (the degrees of freedom vary). The phenomenological differences between different levels call for an explanation. Are there general principles which constrain the hierarchy? In relation to the stratified ontology proposed by Nicolai Hartmann, an axiomatisation of such constraints was presented in Peruzzi (2001). "Horizontal" causation refers to interactions between the components at each level. "Vertical" causation concerns components at different levels. This takes two forms. The first refers to the effects (the state of) lower-level components have on (the state of) higher-level ones. This is a *bottom-up* action; since the Scientific Revolution in the XVII century it has been a prime guiding principle of scientific research. The second form of vertical causation involves effects in the opposite direction: this is a *top-down* action. Whether such action has to be admitted in science, whether it contributes to or is even responsible for the emergence of new structure (e.g., in adaptive selection) and whether there are common patterns for top-down causality across different kinds of systems (such as neural networks, organisms and ecosystems), are questions of longstanding controversy, see Craver, Bechtel (2007). Here, top-down action is simply assumed.

As regards the explanation of mental states and processes, a commitment to the existence of top-down causality is often confused with a denial of the

adequacy of neuroscientific explanation, even though the latter could also admit top-down action between levels of self-organisation in the brain. Moreover, independently of this confusion, the mind-brain debate seems not to acknowledge that the phenomenology of vertical action has a much wider scope in natural science, see Ellis (2012). So the shift from language to mind in the philosophical literature may be misleading to the extent that it aims at general philosophical conclusions.

Thus we have to combine a twofold dialectics, one horizontal (ILGE) and one vertical (top-down and bottom-up).

5. Emergence

In a relatively stable massive system composed of massive relatively stable subsystems, composed of ..., the structure of the system tolerates perturbations at each layer without losing its overall "identity", which in turn is associated with the possibility of top-down effects. Such stability is required in particular for a system within the temperature window which makes living organisms possible as well as the kinds of entities whose shape is sufficiently stable and perceptually recognisable, so to allow for nouns, making our experience of the macro-world "pointed" by kinds of objects and kinds of actions sharply separable from each other. The degree of sharpness must be such that the states of an object can even be confused with the object itself, as witnessed by the way both Aristotle and Kant talked about an object as being the cause of something ("the rock cracked the pot"). We can talk of causes and effects (and not only claim the existence of a "reason" why the state of a system undergoes a change) thanks to a categorisation of what we meet in the macroworld and this categorisation has a sense for sufficiently stable entities: objects and processes become frozen into patterns.

This is no longer the case in a system which is fully interactive: no barriers between any two components, no figure-background threshold, no sharp separation between objects and processes. In addition, we cannot predict *which* quantum fluctuations will occur (or have occurred). We (now as locally external observers) can only predict that they will occur (or have occurred). But finally we realise that unpredictability affects even deterministic systems. Then it becomes difficult to preserve a more than a generic notion of cause, unless one resorts to a Quine-style answer. If there were a top-down action, it would be just "horizontal", i.e. as an effect of the global on the local. But a wider set-

ting is needed to understand the dialectics of bottom-up and top-down causal lines together with an ILGE-enriched model of explanation.

We can think of nature as a many-layered hierarchy of emergent systems based on self-organisation principles which channel the dynamics of constituents of lower layers in certain directions compatible with the boundary conditions. In the light of different layers in complexity and cohesion, it is customary to distinguish two accounts of causality and explanation: bottom-up and top-down. (In the light of previous considerations, one could also think of the two ways as two patterns of cause-free explanation, but this option seems not to have been much explored, so let's stay with the usual reading.)

The first is the classical approach to the dynamics of a system formulated solely in terms of its elementary constituents and the composition of local actions – those each constituent exerts on any other possibly in contact. The second is not only standard in ordinary discourse about events involving agents with an intentional purpose but is also considered an irreducible feature of any explanation of how macro-events can act on micro-events more than additively. Both can work independently of any emergence, i.e. "horizontally", any time the structure of a definite whole plays a specific selective role, through global constraints, on possible state transitions. The existence of a "vertical" structure can also depend on local constraints only and compositionality can have different aspects at each layer, the recognition of which compositionality to be retained even if top-down causality is admitted. Accordingly the range of options is wide and, until we focus on a special kind of emergent (sub-)system within a special kind of system, it is reasonable to maintain this width.

In emphasising the importance of constraints, one must not forget that their selective role is in turn constrained by the nature of constituents. Since the "compositional" character of a system inevitably involves the type of "individual entities" taken as its constituents, we may ask what is their scale, why precisely that scale matters, what patterns of composition are characteristic of that scale and into which "parts" the system can be decomposed at each scale. It may be that, in this identification of constituents and parts, an internal description makes a difference with respect to an external description.

For what concerns the "horizontal dialectics", there are two extreme cases: one arises when the structure of the whole (at the top scale) fully determines the parts, all the way down to the most elementary constituents of the system, which in consequence are implicitly defined and have no objective identity apart from the relational one they acquire through their participation in the

whole. The other, symmetric, extreme case arises when there is no structure (at the top scale) apart from the result of the combined local action of constituents on each other (perhaps in a self-similar way, as with fractals). If we avoid the familiar opposition of reductionism vs holism, we can investigate the full width of the dialectics of bottom-up and top-down causality in systems which lie between these two extremes.

When an emergent system is present, and is self-sustaining (so that it does not decay instantly), the two directions, bottom-up and top-down, are respectively manifested in relation to the effect non-emergent components have on the state of the emergent system and in relation to the inverse effect. Teleonomic behaviour of living beings is an instance of a top-down action on the environment (distributively), whereas the action of the environment (collectively) on living beings is rather a local effect of a global state (unless we personify nature). Instances of inverse action are relative to which resource can be chosen to attain a given goal and which collective results of actions can have global effects (atmospheric pollution is a case).

Since emergent systems are frequently associated with complexity, it should be pointed that simplicity is equally associated with them. (Compare the high-level description of a die as a cube with a number on each face with its low-level description in terms of molecular bonds and the distribution of atoms in each molecule.) But the existence of any specific emergent system, or emergent subsystem of an emergent system, is also the result of a bottom-up selective pressure, relative to the state of a certain environment, which in turn can keep track of previous actions by other emergent systems. The amount of such "memory" distinguishes classes of systems and affects causal talk about them.

Yet again there is a full spectrum of kinds of "vertical" dynamics, no less dialectical. Consideration of this spectrum should be the primary interest of any philosophy free from of reductionist or anti-reductionist bias. Analogously to the ILGE patterns, such spectral analysis is needed before judging the explanatory power of an argument in which entities of different scale and order are involved. Thus, before engaging in a debate on the causal role of mental states and the modular or holistic architecture of the brain, this analysis would be helpful for avoiding the risk of an exclusive focus on a purely top-down analysis of language while at the same time steering clear of the bottom-up constraints on the emergence of meanings, and the converse risk.

6. Missing Modality

We are able to recognise a proof even with no explicit formulation of the logical rules used in it. Once the implicit is made explicit, in order to ensure that a given sequence of sentences is a proof, one has to check the correctness of each step. Checking the correctness of a definition is different but follows a similar line. Passing to the modal notions of provability and definability means a jump in logic, which leads to a finer analysis of the structure of a given proof and definition and also to a precise comparison of different sets of principles used in proving and defining.

Can the same be said of the step leading from the checking of the explanatory character of a given argument to the analysis of explainability? Though different kinds and models of explanation have been made explicit and compared, the answer seems to be in the negative: there was no jump for what concerns results about the set of explanations possible relatively, say, to a pair $<T, U>$, where T is a theory expressible in a specified formal language and D is an empirical domain to which T refers and into which one or more models of T (we conjecture) can be embedded. If the DN model is adopted, one would reply that such results exist, being the same as for provability and logical consequence, now suitably applied to T, to which the laws belong, plus (say) the diagram of D (a notion external to the language of T) to which the list of specific conditions belong, to keep them simple. But this reply is of little help with specific features of scientific explanations for real events. On the other hand, as we introduce more and more constraints, they seem to diverge from one topic to another; and also in the case of causal explanation we have to consider the above dialectics, which varies with the type of emergent system, so that in the end we can hope to construct a framework suitable for a specific theory.

Nevertheless, we cannot claim to have a theory of explainability, even though for a long time we have had various philosophical doctrines about the demarcation line between natural science and metaphysics (theologians address to provide explanations).

No such theory can be obtained by induction, say by examining, one after the other, every existing scientific theory, since the form of theories changes with scientific progress. And if we think explainability (and not only its range) is strictly dependent on the resources of the theories we accept, the same context-oriented point made for explanation can be repeated, and then such strict dependence might lead the subjectivist to modally extend the contextual fission of any con-

cept and the objectivist to dispense with causal or cause-free explanation, at least in principle. For it would be convenient for anyone to keep as meaningful the talk about explanations and causes for Picwickian practice in ordinary language.

But, apart from the modal notion of explainability, the modal notion of necessity had been recognised as a key ingredient of the two main patterns of explanation, i.e. the DN model and the causal one, considered here. (For brevity's sake, I don't consider other alternative patterns that have been proposed, but I don't think they solve the issues listed: they only call for mild modifications of the arguments put forward here.) In fact, necessity is involved in *both* the "nomological" character of general hypotheses supposed to explain something *and* in the causal relation between two events. If we subscribe the research program according to which any talk of necessity for what concerns the cause-effect pair hides an appeal to nomological hypotheses, the second conjunct reduces to the first. In such a case, it remains unclear how can the resort to counterfactuals be internalised, i.e. re-absorbed within a non-metaphysical space of "possible worlds".

7. Permutational Democracy of Causes

The ordinary view is that any event E is the effect of the combined action of many events, say $x_1, ...x_n$, each of which is considered as concurrent to produce E as a result. But which of them is "the cause" of E? The classical way of looking at "the" cause is by means Bacon's twofold "vintage", namely by adding some x_{n+1}, or removing some xi to the tentative list $x_1, ...x_n$ of concurrent causes of E ($1 \leq i \leq n$). One can obviously refine each "vintage" by probabilistic considerations. The issue is how categorical the list has to be.

There are cases in which each action might be performed by other events too, e.g., when in order for E to occur, some source is required which feeds a quantity z of energy, with no need to consider the "quality" of z and its potential influence on other concomitant, or concurrent, causes. In other cases, a similar replacement of source for x_i may not be allowed and E can be very sensitive in this respect. Let's write $[x]_E$ for the class of events that are equivalent to x relative to E, i.e. all those x'that can replace x while providing the same contribution as x to E.

Sensitivity to substitutions clearly depends on the degree of specificity of the description through which E is identified. For instance if E is just my eating a cake, there are many substitutions for the cake's ingredients which preserve the truth of E, while if E is my eating a cheesecake the number of substitutions

diminishes and so on, possibly until one arrives at a *unique* event which admits no substitutions (as in love songs). In the latter case, the equivalence class of the set of causes is also a singleton. Here *uniqueness* is intended as conceptually grounded, thus no spatiotemporal deixis is admitted in the description of events (Leibniz *docet*): coordinates should play no essential role in the explanation, as reality has no preferred coordinatisation (reference frame). If we endorse Aristotle's claim that there is no science of individuals and the explanation for an event-type here and now must work for the "same" (on conceptual grounds) event-type elsewhere and then, singleton classes in an explanation need to be managed with care.

Now take the substitutivity classes $[c_1], ...[c_n]$ for events at time t supposed to be the concurrent causes of a given E at time $t' > t$, where the classes are assumed to be mutually irreducible relative to E. Suppose that the c_i are read as specific conditions (by translating the c-states into state-descriptions), to which a set of laws can be added such that there is a proof p which has the conjunction of conditions and laws as premises and E as conclusion. Then, by passing to equivalence classes $[c_i]$, let's select as representant c_i* of the class the one providing the minimal amount of information sufficient to work for p. This p-preserving minimality of information for each c_i* keeps the relevance objection to the DN-model in standby. As for a p-preserving substitution, let's define [0] as the class of events having no causal role in E (obviously this class varies with E). The set $[c_1] \cup ... \cup [c_n] \cup [0]$ is a partition of events at time t for what concerns E at time t'. If substitutions have to respect a given partition, the size of each class will diminish. If not, the partition will change as an alternative $c_{i'}$ replacing c_i might prevent the occurrence of some c_j.

We may also make a different start by taking "the" cause of E to be the set of concurrent causes, up to substitution, depending on the specificity of E's description.

In either case, we can choose to be democratic or not on the "rights" of different concurrent events to be "responsible" for E. Causal democracy is "flat" in the sense that each concurrent state (think of it as a vote for E) is assigned a weight equal to the weight of any other. As a matter of fact, we tend not to endorse flatness, for a number of reasons. This seems to lend support to the subjectivist party (with emphasis on different contexts and interests) but each reason is ultimately tied to what we, as observers, can identify as a figure against a background, and this is inbuilt biological machinery, thus an objectively emergent factor at work in our partitioning of what happens into types of

events, each type having a corresponding figure. In any case, we may ask, if all of our explanations were flat, could we achieve any of them?

But if the observers are considered part of the domain of explanation, so that their figure-capacity is objectively taken into account, different kinds of observer could have a different stock of figures associated with non-flatness. This adds to the partition of causes up to equivalence another partition of evidence according to equivalence classes of event-figures. Thus, objectivity would not consist in recovering flatness (possibly preferred by a view from nowhere of an external observer) but rather in identifying what is (internally) stable under refiguring. Is such stability beyond human knowledge? Or merely empty? To face this issue, it will be convenient to start by finding the possible event-figures relative to a sufficiently simple explainer-system (different from us), classifying them and arguing why, internally, a given partition of events is accessible and another is not, and how this affects explanations in terms of causes.

Finally, note that if the "no singleton" demand is consistently respected, it covers not just C but also L and E, and thus if anything is determined up to equivalence the by now classical arguments about "empirically equivalent theories" have a different force from what was intended. In fact, to deal in a uniform way with multiple equivalences, a general notion of isomorphism should be employed in constructing any kind of quotient and this suggests a category-theoretic reformulation of the whole topic.

8. Comparing Explanations

Given two logically inequivalent explanations of something, which is the better? The question presupposes that both explanations are correctly argued and we cannot determine which is right and which is wrong, for it goes without saying that right is better than wrong: all actual evidence at our disposal is supposed to be consistent with both. (Future evidence making a difference might depend on which explanation we choose, so we should be careful to refer to it, and talk of "all possible evidence" is more proper to a god-like explainer than a human one).

If both explanations are phrased in terms of causes, their ordering can only rely on two different sets of causes (up to substitution), i.e. two partitions, or two different possible processes starting from one and the same set of causes (for example, imagine having to explain how a certain chemical compound was obtained, knowing that two different sequences of chemical reactions could have produced that compound). In either case, the state of the system beyond E

in some or all the times in (t-, t) will differ, and if any action is by contact, there will be a neighborhood U of E on which the two alternative explanations will differ. So if such U can be accessed, the comparison is no longer an issue. It is an issue only if we don't have such access, that is, if we lack some information (and possibly the gap can't be filled).

If one explanation is causal and the other is not, comparison would be non-homogeneous. One can deal with it in a general or a specific way. The general way appeals to an argument in favour of one *kind* of explanation over the other: in one direction one prefers the arguments for, say, the DN model (or one of its refinements or a cause-free alternative model), in the other direction one prefers arguments in favour of the ineliminability of causal talk. (If what is preferred in principle is actually lacking in some cases, the comparative judgment reduces to a sort of modal desideratum.) The specific way refers to the gain of one explanation with respect to other for what concerns a specific class of facts. The specific way, in principle, admits the coexistence of both models (for different kinds of facts) but then one has to explain how this admission can be part of a coherent general picture, otherwise it is just an instance of making some non-granted virtue out of a present necessity. So the suggestion would be that of translating one specific explanation into one of the other kind and then compare. Finally, if both explanations are DN-like and either specific conditions or the laws are the same, the better one could be identified by a minimax principle.

The by now classical argument based on "empirical equivalence", leading to a more or less mild pragmatic choice, can be applied to any pair of explanations, be they homogeneous or not; it makes any of the previous preference judgments non-objective, by means of a sudden leap to the *totality of possible experience*. But we can only rely on arguments concerning a stable totality we can identify and relatively to which change is defined. As a matter of fact the boundaries of possible experience have changed through the progress of science and thus that totality is neither stable nor identifiable. This does not prevent us from using equivalence arguments with respect to a given stage of knowledge, but then their consequences are no longer those expected. If for any stage of knowledge there are two theories (or two models of explanation) referring to one and the same domain D such that the first can't be claimed to be better than the second in the light of the actual evidence, (rational) pluralism (followed by pragmatic selection) becomes too cheap. It is only if the domain is global and the union of all theories about subdomains is a theory, supposed to come from choosing, for each D, in a covering of all of our evidence, one

in each pair of D-consistent and D-equivalent theories and thus in agreement with the totality of experience, that similar arguments can work. However, it should be noted that i) a collection of theories retains the unification feature which is implicitly associated with "theory" is far from having a clear meaning, since "theory" is not a mass noun; and ii) the result of combining theories, with *necessarily* overlapping subdomains of application, is consistent and uniquely defined is a strong assumption, which I endorsed in Peruzzi (2009) but is not in line with the approach under discussion here.

Finally, there is the problem of determining whether an explanation is better than another. This is a step which is supposed to be completed in order to start any IBE ("Inference to the Best Explanation"), which is in fact an inference to truth from the best explanation, see Lipton (1991). Were it not, it would be the best inference in unconvincing power, for one can list infinitely many truths from which you get no explanation. If all local pragmatic preferences about what is "the better" can be combined, the determination of the global "best" can be reached. Otherwise, it cannot. If the pragmatic line is restricted to a specific domain or to each of them but unpretentious to cover them all, it dispenses with such a task by tagging it as *hic sunt leones.* If it not so restricted, it implies a commitment to an ontological hypothesis. The tag assigns a limit to a pragmatic view of explanation, the commitment to ontology signs its inconsistency.

9. Other Issues

There are other issues too, which cannot be dealt with in this short paper. Let me sketch just three of them.

The first concerns the missing logic of explanation. The common idea is that explainability lives on top of provability: in other words, the relation "B can be explained from $A_1, ..., A_k$" depends on (though it does not reduce to) a logical notion, namely, that of "B can be proved from $A_1, ..., A_k$", which in standard notation is written $A_1, ..., A_k \vdash B$. But we could also imagine dealing with the former relation in an independent way, paralleling $\vdash$ with an autonomous new symbol, say $\wr$, so that "B can be explained from $A_1, ..., A_k$" is now written $A_1, ..., A_k \wr B$. One can easily check that some usual rules for $\vdash$ are preserved, while others are not. It is not clear how much the resulting theory would contribute to eliminating the difficulties examined so far.

A further issue has to do with probability and its multiple but mutually

inconsistent uses: inductive-statistical explanations generally rely on the frequentist interpretation, whereas, when rational agents are involved in the domain of events to be explained, the subjectivist interpretation is favoured. To this it should be also added that there are further approaches, as in Dempster-Shafer theory, about epistemic uncertainty.

The third issue, the absence of which is (rightly) regarded as a major gap in this paper, concerns the status to be assigned to IBE. Its brief mention in § 8 is simply a pointer to the link between explanation and truth, and to its multiple aspects, one of which is the acknowledgement of an essential role to abduction. In Peruzzi (2009) I argued for a weakening of IBE which is, however, abduction-free, which makes it look like a non-standard point of view. I now think that argument needs revision, which is also related to a different formalisation (yet a work in progress) of the structure of an explanation.

REFERENCES

Craver, C., & Bechtel, W. (2007). Top-down Causation without Top-down Causes. *Biology and Philosophy*, 22, 547–563.

Ellis, G. (2012). Top-down Causation and Emergence. *Interface Focus*, 2, 126–140.

Hempel, C.G. (1965). *Aspects of Scientific Explanation*. New York: Free Press.

Lipton, P. (1991). *Inference to the Best Explanation*. London: Routledge.

Peruzzi, A. (2001). Hartmann's Stratified Reality. *Axiomathes*, 12, 227–260.

Peruzzi, A. (2002). ILGE-interference Patterns in Semantics and Epistemology. *Axiomathes*, 13, 39–64.

Peruzzi, A. (2004). Causality in the Texture of Mind. In A. Peruzzi (Ed.), *Mind and Causality*. Amsterdam: John Benjamins, 199–228.

Peruzzi, A. (2009). *Modelli della spiegazione scientifica*. Firenze: Firenze University Press.

Salmon, W. (1998). *Causality and Explanation*. New York: Oxford University Press.

van Fraassen, B. (1977). The Pragmatics of Explanation. *American Philosophical Quarterly*, 14, 43–150.