

EXPLANATION AND MODELIZATION IN A COMPREHENSIVE INFERENTIAL ACCOUNT

Xavier Donato Rodríguez
Jesús Zamora Bonilla

1. Introduction

In the present paper, we defend an inferential account both of explanation and scientific modelling. Our account is “comprehensive” in the sense that we assume a pragmatic perspective that tries to capture the intrinsic versatility scientific models and explanations may adopt in the course of scientific discourse. This inferential-pragmatic view is essentially inspired by the work of Robert Brandom in the philosophy of language (see Brandom 1994 and 2000), but takes elements from other authors, mainly from argumentation theory and epistemology. As many philosophers of science that favour an inferential perspective, we see scientific models as inferential tools that help to extract inferences about the target in relation to specific goals. In this sense, models can be understood as consisting of sets of interconnected commitments and inferential norms allowing us to explaining and predicting phenomena in a relevant way (we develop this in de Donato and Zamora-Bonilla 2009). Likewise, explanation can be seen as a particular form of speech act understood according to a pragmatic-inferential view that allows to capturing the versatility of explanation. This is our main goal in the present paper.

2. An inferential approach to scientific discourse and inquiry

Scientific discourse, like all kind of discursive practice understood as the game of giving and asking for reasons (to use a well-known expression introduced by Wilfrid Sellars), is implicitly normative, as it includes assessments of moves as correct or incorrect, appropriate or inappropriate (see Brandom 1994, 159). It would not seem a too unnatural move to apply Brandom’s distinction to scientific discourse and this is in fact what we aim to do. We will distinguish between:

- *commitment*: something that a scientist is committed to believe because it is a principle, a general law or a methodological rule unanimously recognised as such by the community which she belongs to, or because it logically follows from other commitments.
- *entitlement*: something legitimate for a scientist to be believed or claimed because there are good reasons (including those provided by inductive methods, analogy or abduction and those provided by testimony and authority).

According to this inferential view of the “game of science”, there is a kind of internal normativity in scientific discourse governing the significance of changes of deontic status, which specify the consequences of such changes, link various statuses and attitudes into systems of interdefined practices. As Brandom (1994) puts it, acknowledging the under-taking of an assertional commitment has the *social*

consequence of licensing or entitling others to *attribute* that commitment (this is crucial to the rational debate between groups of scientists). Brandom distinguishes intrapersonal as well as interpersonal inferences. There is indeed a kind of social inheritance of entitlement and, in this process, the content of the commitments is preserved intact. To adopt a certain attitude has consequences for the kind of commitments the participants in a scientific discussion are entitled to assume. In this sense, participants in a rational discussion are committed to do certain moves and exclude others provided the context of the discussion and the situation which they depart from. Following the discussion in a rational way involves in part taking these commitments into account and preserving them. Not surprisingly, this perspective (or a form which is very near to it) has also been adopted in argumentation theory. For example, Walton and Krabbe say in a book on *Commitment in dialogue*, highlighting the normative character of interpersonal reasoning:

“Therefore, in order for argumentation to have enough platform of stability and consistency to make sense or be reasonable (and, in fact, rational) as a dialogue unfolds, it is necessary that an arguer’s commitments be fixed in place. If someone keeps denying propositions she has just asserted, or keeps hedging and evading commitment whenever it appears, her argument may be automatically defeated, since you can never get anywhere with this arguer. Arguing with her becomes pointless.” (Walton and Krabbe 1995, 10).

Within the pragma-dialectical model of argumentation defended by van Eemeren and Grootendorst (2004), an approach that has become very influential in argumentation theory, externalization of commitments is achieved by investigating exactly what obligations are created by performing certain speech acts in the specific context of an argumentative discourse or debate. Acceptance or disagreement are not just states of a particular mind, but they stand primarily for undertaking public commitments that are assumed in a rational discussion and can be externalized from the discourse. So they have a social and pragmatic dimension (see van Eemeren and Grootendorst 2004, chap. 3, 42 and ff.)¹. As Walton and Krabbe, van Eemeren and Grootendorst want to insist on the normative character of the exchange of reasons and arguments undertaken in rational discussions.

If we speak of the “game of science”, different kind of norms may be distinguished (see Zamora-Bonilla 2006): (i) internal norms: inferential norms and epistemic norms evaluating the epistemic quality of our claims and theories; (ii) entry norms: norms about authority and about evidence gathering (i.e. how empirical observations are to be carried out and interpreted); (iii) exit norms: norms that regulate things like publication, jobs, prizes, funding, or academic recognition. About epistemic norms, in de Donato and Zamora-Bonilla 2009, we have developed an inferential picture of scientific modelling according to which models consist of sets of interconnected doxastic commitments and inferential norms, some of the latter made explicit by formal rules and other taking the form of practices. We characterize there a *model* (or *representans*) as an *inferential machine* to which an intended interpretation is associated and that represents a certain *real system* (or *representandum*). This relation between *representans* and *representandum* can be modeled in different ways.

¹ See Kibble (2006) for an attempt of formalization; of course, certain aspects of Hintikka’s interrogative model of research can be seen as belonging to this family of approaches.

There is already a well known literature defending an inferential view of scientific models (see, for example, Swoyer 1991, Hughes 1997, Suárez 2004, Contessa 2007, Bueno and Colyvan 2009). Our main contribution is to defend an inferential account of models according to a pragmatist view similar to that of Brandom. We have applied it to the process by which a scientific claim becomes accepted among the relevant research community². Taking this perspective into account, it is seen as a desideratum that the addition of a model (of a new claim or hypothesis) to the corpus of our commitments serve to increase the ratio of successful inferences to not so successful ones. At the same time, the model should increase the number and variety of inferences we were able to draw from the rest of our commitments and help us to reduce the cognitive (computational) costs of the activity of drawing consequences, either doxastic or practical. In general, the value of a new commitment (or of the inferential patterns that generates it) will depend on its ‘size’ (the number of questions it answers), its coherence (especially between those commitments derived from others, and those derived from perception), and its manageability (models should allow to draw many consequences at a low cognitive cost). In the same paper, we also give a list of possible virtues a good model should have: adequacy, versatility, ergonomics, and compatibility. According to this picture, a model might not only be good for explaining a certain set of phenomena, but also for being capable of making more coherent and workable our network of previous commitments. We label this capacity as *enlightening*. More about this comes in the next section. Before we go into it, let us make two important remarks regarding the status of commitments: firstly, commitments can be explicitly *acknowledged* or *consequentially* (and therefore implicitly) assumed; secondly, the status of a commitment to which the participant in a debate is *prima facie* entitled is not permanent: the legitimacy of an assertive commitment can be challenged at any time, whenever it is challenged in an appropriate, reasoned way.

This extrapolation of the idea of commitment to the context of scientific knowledge is not new. Polanyi (1958) argued that scientists must be committed to the cause of science and that commitment plays an essential role in the actual pursuit of scientific knowledge, in contexts such as discovery, observation, verification, or theory construction. In fact, he contends that scientific inquiry is possible only in such a context of commitment. It is even passionate commitment what has led to scientific controversy (in the sense of rational discussion of alternative and opposite hypotheses) and the growth of scientific knowledge³.

3. Explanation in scientific dialogues: credibility vs. enlightening

We depart from a pragmatic and dialectical perspective of explanation. Walton (2004 and 2007) offers a good example of what we have in mind. Basically, if you are taking part in a dialogue (i.e., in any linguistically mediated practical situation), your commitments are those propositions the other speakers are allowed to take as premises in their reasoned attempts to persuade you of other sentences, or, as it often occurs, in their attempts to persuade you of performing some actions. The dialogue takes place according to a set of *norms* that tell what inferential links between commitments (and

² See de Donato and Zamora-Bonilla (2009).

³ See Hall (1982) for a comparison between Polanyi’s account of commitment in science and Popper’s methodology.

between commitments and actions or events) are appropriate, and that partly depend of the type of dialogue we are dealing with. Of course, the process of the dialogue can force the agents to retract some commitments or adopt new ones. Furthermore, and importantly, it is not strictly necessary that a speaker *believes* all her commitments, though we may assume that the norms of a dialogue should tend to make consistent one's commitments with one's beliefs. The basic difference is that we cannot have a belief at will, whereas we often undertake or retract some commitments *voluntarily*, and we even have a certain degree of control over the *set* of our commitments by strategically selecting those ones that (according to the inferential norms of the dialogue we are immersed into) will lead us to have the most beneficial set of commitments (ask any politician if you are in doubt).

Walton's paradigmatic example of explanation is the case where somebody wants to do something (e.g. a photocopy with a new machine) and does not know how, and then asks somebody else to 'explain' her how to do it. A better example for our purposes would be the case where the photocopy does not come out well even when you think you are strictly following the machine's instructions, and then you ask someone more expert for an explanation of your failure. This case shows that the reason why you want an explanation is because you *expected* the machine to work well. In the terms introduced before, the expert and you are committed (or so you think) to certain propositions describing the proper functioning of the machine; from these commitments, together with your actions and the inferential rules applicable to the case, you are led in principle to a new commitment of the form 'the photocopy *should* come out well', but the observation of the actual result commits you to the proposition that it comes out badly. What you demand from the expert is some information about which of your previous *commitments* you have to retract of ('it's not here where you must put the paper'), what new ones you have to add ('you also have to press here'), what of your *inferences* were ill grounded, which new inferences you can do from formerly entertained commitments,⁴ or what *facts* about the machine are making it not to function properly. So, in general, we can say that, in a dialogue, it is justifiable for agent A (the *explainee*) to ask other agent B (the *explainer*)⁵ for an explanation of a proposition P if:

- (a) A and B are committed to the truth of P, and
- (b) some of the other common commitments (or at least what A takes these commitments to be) would make proposition P untenable.

By 'untenable' we mean that P stands in some relation of opposition or incompatibility with respect to other statements, though this opposition need not be as strong as a logical contradiction. It can be just that the explainee has some reason to think that P 'should not' be the case, or even that he does not *want* P to be the case. It is relevant to note here that the inferential norms governing actual dialogues do not only reduce to the principles of formal logic, but also facilitate many *material* inferences. This is in part what explains that we are so often in the situation of having to accept incompatible propositions, because, if we only were allowed to use the rules of logic,

⁴ In contrast with what Hilpinen (2005, 202) suggests, an explanation can start from some premises you already knew, just that you didn't notice that the explanandum followed from them. As a matter of fact, many of the 'explanations' offered in economics about the recent changes of economic variables comes from known premises.

⁵ The terms of 'explainer' and 'explainee' are used, for example, by Goldberg (1965) and other authors.

plus assertions describing our direct experiences as input-commitments (as the first logical positivist dreamed of), then no contradiction could ever exist in our set of commitments. On the other hand, it is because actual inferential rules are informationally much stronger than mere logical rules, that we can have the amazingly rich sets of commitments (i.e., ‘knowledge’) we have. But, of course, it is this richness what very often produces the kind of incoherence that leads us to ask for explanations. Of course, other types of speech acts can count as ‘explanations’ in some cases, but we shall assume in the next sections that this function of ‘incompatibility removing’ is the most characteristic one in the case of science, and we can even perhaps defend that, in any other linguistic or cognitive process that we can reasonably call ‘explanation’, there is an essential room for the *unblocking of the smooth flow of inferences* that incongruence removal tends to facilitate. For example, in a recent survey on the psychology of explanation, Keil (2006, 133ff.) makes a casual list of things daily explanations are for (favouring prediction and diagnosis, rationalising actions, and allowing responsibility adjudications) in all of which it is clear that, *once* we are provided with an explanation, we have an improved capacity of making *new* inferences (as Keil says, by knowing “how to weight information or how to allocate attention when approaching a situation”, Keil 2006, 234). Even in the remaining case quoted by Keil (aesthetic pleasure, in which he includes merely epistemic explanations), it would serve to “increase appreciation (...), providing (us) with a better polished lens through which to view the explanandum” (*ibid.*).

4. Explanation in scientific dialogues: credibility vs enlightening.

One aspect that could seem odd of Walton’s dialectic theory of explanation (DTE), particularly when applied to science, is the fact that it understands explanation mainly as a relation between two agents (the explainer and the explainee), rather than between two sets of propositions (the traditional *explanans* and *explanandum*). Certainly, Walton’s theory assumes that the explainer *already knows* an explanation the explainee ignores, and, hence, the theory is mainly devoted to analyze the intricacy of the norms and strategies the agents have to follow in order for the dialogue to be successful under those circumstances. But in the case of science the most essential point is obviously the *search* for explanations. Nevertheless, that this difference is not as important as it might seem at first sight. For, first, also in non-scientific cases it is often the case that the explainer needs to look for an ignored explanation. And more importantly, in looking for an unknown *explanans*, a scientist has to take into account whether her proposed explanation will count as acceptable when displayed to her colleagues, and hence, *having an idea of what makes an explanatory speech act successful is as relevant during the dialogue that before*. We have argued elsewhere that it is useful to understand the process of scientific research as a dialogical interaction, at least in order to make the epistemic rationality of science consistent with the fact that the pursuit of recognition is a basic force behind the decisions of scientists concerning their research and publication strategies (see Zamora-Bonilla 2002 and 2006). The question we are dealing with here refers to the *types of reasons* you can give in order to persuade your colleagues that your proposal is acceptable. Seeing explanation as an argument whose logical dynamics ‘goes’ from the *explanans* to the *explanandum*, either in the traditional accounts of explanation as in Walton’s DTE, tends to obscure the fact that, in most cases, scientists use such a kind of arguments to persuade her colleagues of accepting the particular *explanans* they are proposing. In this game of persuasion,

explanations are used as moves directed towards the acceptability of an explanatory hypothesis or model. This is why, even in spite of its logical or mathematical shortcomings, ‘inference to the best explanation’, or something similar to it, is a strategy frequently used by scientists: from their *pragmatic* point of view, the problem is *not only* whether T’s being the best available explanation of some facts ‘proves’ that T is true or highly probable, but also whether T surpasses or not the quality level that the norms governing the dynamics of commitments establish for a theory to become *acceptable*.⁶ So, in the game of persuasion, being T an explanation of some fact is just one of the possible ‘good reasons’ that you are accumulating in order to persuade your colleagues that T must be accepted.

The question is, hence, *what role do explanatory arguments play within the dialogues science consists in?* Our proposal is to look at this question by discussing in the first place what is, for a scientific community, the *value* of having a specific set of commitments (or, in old fashioned language, a certain ‘corpus of knowledge’). In everyday situations, this value derives from the degree of success of those *actions* our *practical* commitments command us (or allow us) to perform, though it can also be true that having some merely *doxastic* commitments also gives us some kind of cognitive satisfaction when they are successful (e.g., when our predictions are fulfilled, even if we don’t win anything else from it), and sometimes independently of any kind of empirical validity (e.g., when we adopt a certain religious dogma). In the case of science, however, the relative weight of practical and epistemic criteria for assessing the value of a set of commitments are probably reversed to some degree, though this does not mean that practical success is unimportant; actually, from a *social* point of view the value that most citizens attach to science, and what justifies for them that their income is partly transferred to scientific research, comes out of the practical things scientific knowledge *allows* to do. So, *ceteris paribus* a scientific community will attach a higher value to a set of epistemic commitments if the latter allows to deriving more practical or technological *right* predictions. Of course, predictive capacity is an internal goal as well, not related in this sense to technological applications; this has to do in part with the fact that scientists derive a satisfaction from having been lead to the *right* commitments, for successful prediction is a way of judging just this. By the way, we might call ‘prediction’, in the most possible general sense, any propositional commitment that can be derived from *more than one* subset of the commitments you have, or in more than one independent way, according to the inferential rules which are applicable (when one of these alternative lines of derivation is from our commitment to accept some empirical results, we would talk of an ‘empirical prediction’).⁷ So, *giving us the capacity of making useful and right ‘predictions’*, in this general sense, would be the main factor explaining the value of a set of commitments (see section 2 of this paper).

On the other hand, predictions have to be carried out by means of *arguments*, and performing arguments in a proper way, according to the relevant inferential norms, can be a very *difficult* and *costly* activity, both psychologically and technically speaking. So, a second ingredient of the value of a set of commitments is the relative *easiness* with which it allows the inferential moves that lead us from a set of

⁶ See Zamora-Bonilla (2002 and 2006) for a model about the choice of that quality level.

⁷ In this sense, even the proof of a mathematical theorem is a ‘prediction’, since by considering it a theorem the mathematical community ‘bets’ that a counterproof will not be discovered, and that alternative proofs can be envisaged.

commitments to others. Take also into account that one commitment can be more or less *strong*, i.e., its acceptance within a community can be *more or less* compulsory, and the inferential links will work more smoothly (at least in the sense of allowing less disputes) the stronger the relevant commitments are, *ceteris paribus*. Hence, the more contested by anomalies a theoretical commitment is, the less confidence the scientists will have in adding to their corpus of commitments the ones entailed by the former (its ‘predictions’).

This ‘ergonomic’ aspect of our sets of commitments is what we suggest to identify with the notion of *understanding*: we understand something (better) when we are able of smoothly inserting it into a strong network of (as many as possible) inferential links, in such a way that many new (right) consequences are derivable thanks to it.⁸ According to this view, it is reasonable to assume that scientific communities will have developed inferential norms that make it more acceptable those theories, hypotheses, or models that are more efficient in allowing, *in the first place*, to derive the highest possible number of right predictions (in the general sense indicated above), and *in the second place*, to increase in the least possible way the inferential complexity of the arguments employed in the discipline, or even to reduce it from its current level of complexity and difficult. The differences in the norms of different communities can be due in part to the relative hardness of attaining these two goals, or the specificities of the methods which are more efficient for doing it in every field of research. So, in general, when a scientist shows that the hypothesis she is proposing (H) ‘explains’ a known fact or collection of facts (F), this can make her model more acceptable because of *two* reasons: first, that H entails F is just an argument for the empirical validity of H;⁹ second, if F was an ‘anomalous’ fact, i.e., one colliding what other accepted models or theories in the field, then H’s entailing F allows to introduce the latter into a more coherent network of inferential moves, reinforcing by the way the strength of the core theoretical commitments of the discipline. Stated in other words, the value of the new theoretical commitment H will depend on two sets of considerations: in the first place, those that make it for a research community reasonable to accept the new commitment *because of its truth* (depending on the field, different degrees or types of approximate truth will be allowed, of course), and in the second place, those that make it reasonable to accept it *because of its capacity in making more coherent and workable the network of commitments and inferential links of the discipline*. We suggest to call “*credibility*” and “*enlightening*” these two general factors, respectively. In principle, ‘good’ theories must be good in both senses, though obviously some theories can be better in one of them than in the other, and under some circumstances a very high performance in one sense can compensate for a not so high level attained in the other. And, of course, different disciplines can attach a higher value to one of the reasons (as, by the way, they can attach different weights to the factors determining the value of each reason), reflecting the relative difficultness or benefits associated to each in the corresponding field.

A reason why a pragmatic-inferential approach to explanation should deserve attention will be that it allows to understanding why so many different “theories” of

⁸ Robert Brandom, by equating the *meaning* of a concept to its set of inferential links to other concepts, explains also the notion of ‘*understanding* a concept’ just as the capacity of performing in a proper way those inferential moves. See Brandom (1994, 85 and ff.).

⁹ By the way, what is important here is just that F logically follows from H, not that the latter explains the former in a causal, functional, or any another specific sense of ‘explaining’.

explanation have had some conceptual appeal. Think of theories identifying explanation with, for example, understanding, deduction from laws, identification of causal processes, identification of functional or identification of intentional processes. These different perspective may well be and have been shown to be appropriate depending on the context, but none could have been erected as *the* correct approach, subsuming all the others. In this sense, there is a part of truth in the nihilistic theories of explanation, according to which there is nothing substantial which is common to all kinds of explanation that we may correctly distinguish. Our pragmatic-inferential approach would be able to account, in inferential-pragmatic terms, for the reasonability of each of the different approaches just considered, because they would be easily interpretable in terms of nets of appropriate inferential links. A pragmatic-inferential approach as the one defended here could interpret each kind of explanation as saying us which is the relevant inferential link in order to use it in an appropriate argument in a given context, i.e. which kind of inferential link are we entitled to use in a given argumentative situation.

5. Conclusion

We have endorsed a pragmatic-inferential view of models and explanation. We have sketched the main line of this view in the precedent sections. To conclude, let us now say something about which general implications for epistemology this view amounts to. Elsewhere, we have endorsed a view of epistemic values according to which good epistemic procedures are not those that create the value of their outputs, but those that help us to identify the outputs that are *valuable in order to attain our expected goals* (provide right, simple and accurate predictions, to give appropriate explanations for surprising phenomena, to develop faster computer machines allowing us to do specific operations...). We should perhaps give up the Cartesian project of giving universal norms and criteria for the value of knowledge and concentrate our efforts as epistemologists in trying to delimit our world of rights and duties taken the appropriate context into account. According to this picture, epistemic norms are not universal, but they depend on our epistemic standards and specific goals. So they can change, as well as our inferential norms, if necessary, in favour of other norms that allow us to better results. A crucial aspect is that epistemic normativity has essentially a social dimension and that the epistemic agents are groups and institutions rather than mere individuals. What is the value of a particular norm in a given context, when we face a particular problem, should be a point of discussion of the members of these groups and institutions and cannot be decided *a priori* by conceptual analysis.

Bibliography

- Brandom, R. (1994): *Making it explicit*, Harvard University Press, Cambridge.
- Brandom, R. (2000): *Articulating Reasons*, Harvard University Press, Cambridge.
- Bueno, O. And M. Colyvan (forthcoming): "An inferential conception of the application of mathematics", *Nous*.
- Contessa, G. (2007): "Scientific representation, interpretation, and surrogate reasoning", *Philosophy of Science*, 74, 48–68.
- de Donato, X. and J. Zamora-Bonilla (2009): "Credibility, Idealisation, and Model Building: An Inferential Approach", *Erkenntnis* 70, 101-118.

- Goldberg, L. (1965): *An inquiry into the nature of accounting*, Arno Press, Wisconsin, 1980 (reprint of the original edition).
- Hall, R.L. (1982): "The Role of Commitment in Scientific Inquiry: Polanyi or Popper?", *Human Studies*, vol. 5, no. 1, 45-60.
- Hintikka, J. (1986): "Logic of Conversation as a Logic of Dialogue", in R. Grandy and Warner (eds.): *Intentions, Categories and Ends*, Clarendon Press, Oxford, 259-272.
- Hughes, R. I. G. (1997): "Models and representation", *Philosophy of Science*, 64, 325-336.
- Keil, F.C. (2006): "Explanation and Understanding", *Annual Review of Psychology* 57, 227-254.
- Kibble, R. (2006): *Reasoning about propositional commitments in dialogue*, Springer, Amsterdam, Netherlands.
- Polanyi, M. (1958): *Personal Knowledge. Towards a Post-Critical Philosophy*, Routledge and Kegan Paul, London (corrected edition, 1962).
- Suárez, M. (2004): "An inferential conception of scientific representation", *Philosophy of Science*, 71, 767-779.
- Swoyer, C. (1991): "Structural representation and surrogate reasoning", *Synthese*, 87, 449-508.
- van Eemeren, F.H. and R. Grootendorst (2004): *A systematic theory of argumentation: the pragma-dialectical approach*, Cambridge University Press, Cambridge.
- Walton, D.N. and E.C.W. Krabbe (1995): *Commitment in Dialogue*, State University of New York Press.
- Walton, D. (2004): *Abductive Reasoning*, University of Alabama Press.
- Walton, D. (2007): "Dialogical Models of Explanation", in *Explanation-Aware Computing: Papers from the 2007 AAI Workshop*, Association for the Advancement of Artificial Intelligence, Technical Report WS-07-06, AAAI Press, 2007,1-9.
- Zamora-Bonilla, J. (2002): "Scientific Inference and the Pursuit of Fame: A Contractarian Approach", *Philosophy of Science* 69, 300-323.
- Zamora-Bonilla, J. (2006): "Science as a Persuasion Game", *Episteme* 2, 189-201.