

Three Senses of ‘Explanation’

Jonathan Waskan (waskan@illinois.edu)

Ian Harmon (iharmon2@illinois.edu)

Andrew Higgins (higgins9@illinois.edu)

Joseph Spino (spino2@illinois.edu)

Department of Philosophy, 810 S. Wright Street

Urbana, IL 61801 USA

Abstract

‘Explanation’ appears to be ambiguous between a representational-artifact, an objective, and a doxastic sense. That the distinctions between the three are still poorly understood we regard as an impediment to progress in the philosophy of science and as a source of the field’s resistance to greater integration with experimental psychology. We begin to elucidate the overlapping contours of the three sense of ‘explanation’ using a variation on Powell & Horne’s Semantic Integration paradigm, showing that both laypeople and scientists regard doxastic explanations as constitutive of representational-artifact, but not of objective, explanations and accuracy as closely connected to objective, but not representational-artifact, explanations.

Keywords: Explanation; understanding; philosophy of science

Investigating Conceptions of Explanation

In the early 20th century, many positivistic philosophers claimed that science rightly concerns itself only with questions of ‘what?’ not ‘why?’ because, as Stace (1935) put it, “‘why?’ does not proceed from the intellect, but from the emotions.” The general sentiment was that answers to why-questions engender mere *feelings* of understanding, empathy, or familiarity. Nevertheless, by the mid-20th century, there was general agreement that science has the function of answering both sorts of questions, that it not only describes but also explains (Hempel & Oppenheim 1948). This shift in attitude had much to do with the pioneering work of Hempel, who regarded explanations as *representational artifacts* – that is, sets of statements that answer ‘why?’ questions. On one common manner of speaking (e.g., “There is an explanation for the odd trait on page 25”) the noun ‘explanation’ does seem to refer to a set of representational-artifacts; let us call them explanations_{r-a}. Hempel (1965) countered worries about the affect-inducing character of explanations_{r-a} by claiming that they are doubly dissociable from psychological states, and he embraced the apparent corollary that the philosophical study of explanations is, like formal logic, autonomous from scientific psychology. Accuracy, on the other hand, he regarded as essential for explanation.

Many subsequent philosophers would agree with Hempel’s anti-psychologism and his claims about accuracy. One holdout is Achinstein (1984), who contends that some statements E constitute an explanation_{r-a} for q only if an explainer could cite E and thereby render q intelligible to an

audience. On his view, E need not have actually been used to render q intelligible; it suffices that it *could* be used in this way. Nor, for Achinstein, need E be accurate.

Others made a cleaner break with positivism. Salmon (1984), for instance, claimed that explanations are sets of objective happenings, what we will call explanations_o. Here too we agree that there is a common manner of speaking (e.g., “The explanation for combustion is oxidation”) on which ‘explanation’ appears to refer to a set of objective happenings. Similar to Hempel, Salmon argued that explanations_o have nothing to do with what anyone feels or thinks and, as such, are non-psychological. As for accuracy, since explanations_o are not representations, they cannot be right or wrong, “they just are” (Craver 2007).

Meanwhile, as early as Craik (1943), psychologists have undertaken their own investigation of explanations, which they construe as mental states, or more particularly as having a belief about what might have produced the target phenomenon (see Waskan et al. 2013). There does appear to be a common manner of speaking (e.g., “Lavoisier had an explanation for combustion” or “The preverbal infant had an explanation for the meowing coming from the closet”) on which ‘explanation’ refers to a doxastic state; let us call it explanations_d. The relevant beliefs are those in virtue of which one understands how or why, at least possibly, the phenomenon came about or, more succinctly, in virtue of which one finds the phenomenon intelligible. Psychologists often study the explanations_d of children and science-naïve laypeople, whom they often regard as misrepresenting the state of the world, the tendency thus being to disregard accuracy as an important feature of explanation_d.

It may be that all sides are correct in that the noun ‘explanation’ is actually ambiguous between at least these three senses. We are not the first to propose this. Craver (2014), for instance, claims that ‘explanation’ can refer to “a representation or text...a cognitive act, and...an objective structure” (also see Waskan 2006). Salmon (1998) likewise shows sensitivity to “The radical ambiguities of ‘explanation,’” which, he claims, “create almost endless opportunities for obfuscation and confusion.”

Though ‘explanation’ seems to designate either artifacts, objective happenings, or mental states, the three senses of the term are closely intertwined. For instance, what explanations_d we have may depend upon what explanations_{r-a} we have read (e.g., phlogiston or oxidation theory). And what we believe to be the explanation_o for an occurrence (e.g., combustion) may depend upon what theories we take

to be true; though what the objective explanation *is* will presumably not depend upon what we believe or what theories we have encountered (e.g., in print). These interconnections complicate the project of disentangling the various uses of ‘explanation.’ Even so, we regard this as a worthy pursuit given the role that explanations play in our lay and scientific lives. Indeed, with regard to science, it is hard to see how one could effect the larger epistemological project of determining what makes it such an exemplary epistemic exercise without first understanding what, precisely, ‘explanation’ refers to in its various senses.

Philosophers of science sometimes seek answers to questions about how ‘explanation’ is used, and its referent conceived of, by consulting judgments about cases. For instance, in order to show that explanation_{r-a} is in no way a psychological category, one might imagine a well-supported hyper-complex model of a phenomenon that is so complex that it is incapable of rendering that phenomenon intelligible to anyone and then judge that this model is still an obvious or clear-cut case of explanation (Craver 2007; Trout 2007). This practice carries with it the implicit assumption that others will concur on the relevant judgments. As Hempel puts it with regard to his own theory, “the construal here put forward has to be justified by ... [showing that it] does justice to such accounts as *are generally agreed* to be instances of scientific explanation...” (italics ours, 1965). Given that the subject matter here is *scientific* explanation, we take it that the agreement in question ought to extend, at a minimum, to practicing scientists. Obviously there is some risk that the philosopher’s own judgments will not match those of scientists, either because they have not mastered science’s sociolinguistic norms or because their judgments are colored by the very theories of explanation they are trying to prove (Cummins 1998). The latter worry is much the same as the one driving the utilization of theoretically naïve coders for classifying qualitative data in psychology.

Another staple method in the philosophy of science is to comb the historical record for evidence that supports or undermines a given theory. To do this effectively, one must be careful that one’s sample is adequately large, representative, and unbiased (see Thagard 2012). But these restrictions are seldom met in practice. Often samples are small and focus on hand picked cases, revolutionary developments, or cases about which a theorist has special knowledge.

While by no means advocate the abandonment of traditional philosophical methods, the fecundity of which is beyond question, we think that the above worries suffice to motivate greater reliance upon other tools, so that the best descriptive theories of science can be shown to enjoy converging support from a variety of independent sources.

Bibliometric techniques would seem an almost mandatory addition to the philosopher’s toolkit, for these can shed clear (i.e., uncolored by theory) light on facts about linguistic usage. Overton (2013) has already carried out a bibliometric investigation of how ‘explain’ and its cognates are used in recent scientific articles. Unfortunately, he only considers

the semantics of explanation-talk in a small set of case studies and primarily relies on manual search. Nonetheless, his basic strategy could easily be augmented to bring the broader arsenal of techniques available at the intersection of information and computer science (e.g., machine learning algorithms, big data text mining) to bear on the analysis of a far more comprehensive sample of scientific articles.

The methods of experimental and social psychology offer another, independent way of studying a variety of descriptive questions about how scientists talk and think. These methods have already been employed to show that terms like ‘gene’ and ‘innate’ are used in varying ways by laypeople and scientists and subpopulations thereof (Stotz & Griffiths 2008). Here we pursue a related line of investigation, examining how laypeople and scientists conceive of explanation_{r-a} and explanation_o with an eye towards possible linkages between each and explanations_d.

Explanation_{r-a} and Explanation_o

As noted above, philosophers often regard explanation_{r-a} and explanation_o as dissociable from psychological states and, on this basis, they view the philosophical investigation of explanation as autonomous from work in experimental psychology. Our previous studies suggest, however, that laypeople and scientists consider it a central feature of explanations_{r-a} that they actually produce the psychological state of finding intelligible. This looks a lot like Achinstein’s view except that the mere capacity to render intelligible is not enough.

This finding held up under two very different paradigms. The first involved traditional brief vignettes (~200 words) followed by ratings questions (Braverman et al. 2012). In the second, rather than asking for an explicit rating, which we worry could prompt explicit deliberations and amateur philosophizing about category membership, we hid the explanation probe within a larger true-false comprehension test about the materials.

It seems plausible that insofar as one understands how or why-possibly a phenomenon occurred, one has an explanation_d for it, and vice versa. If that is the case, then explanation_{r-a} comes out looking not just inherently psychological, but actually constituted by explanation_d, and this brings unity to the psychological and (part of) the philosophical investigations of explanation. Now it could be that there is a similar underlying psychological component of explanation_o – that is, it could be that an objective process is only regarded as the explanation for a phenomenon if someone understands, or could understand, how it brought the phenomenon about. To find that this is so would further unify the philosophical and psychological projects. To find that it is not would reinforce the suggestion that there are distinct representational-artifact and objective senses of ‘explanation’ in play and it would highlight their distinguishing features.

Also of interest is whether or not accuracy is regarded as essential for explanation_{r-a}. Because explanations_o are ostensibly just the actual facts, then an accurate

explanation_{r-a} (or explanation_d) for a phenomenon should (with some caveats) elucidate the explanation_o for it. We thus undertake to determine if laypeople and scientists conceive of explanations_{r-a} as veridical representations of explanations_o.

Experiment 1: Explanation and Intelligibility

Our first experiment uses similar methods as (Waskan et al. 2013), except that a lengthy distractor (about neuroscience) is interposed between the target materials and the comprehension test. This brings our methods into closer alignment with Powell and Horne's (see Powell et al. 2013) Semantic Integration paradigm, which relies on the fact that, when remembering a passage of text after a delay, participants' memories often reflect their semantic interpretations rather than the actual sentences they read. We administered the same materials separately to two groups of participants, laypeople and scientists. Participants were presented with one of two variants on the gamma-ray article:

- Intelligible – the article describes a hyper-complex computer model and a published research article describing it (hereafter 'materials'), which have numerous theoretical virtues and have actually rendered the target occurrence intelligible to a scientist (Brown).
- Never Intelligible – same as Intelligible except the materials are described as prohibitively complex such that they are incapable of rendering the target occurrence intelligible to anyone.

We examined how often participants remembered that the story described an explanation_{r-a} or an explanation_o under each condition.

Participants

For lay participants, we recruited 115 workers through the Amazon Mechanical Turk (MTurk) work distribution website.¹ For scientists, we recruited 112 practicing scientists (faculty members, postdocs, and graduate students) affiliated with the Beckman Institute for Advanced Research at the University of Illinois at Urbana-Champaign. Data from MTurk workers and scientists were analyzed separately.

Materials and Procedure

The gamma-ray article (roughly 1000 words) was adapted from an article on a popular science website.² The original article describes the long history of research into gamma-ray bursts. Our version utilizes most of this material, but it culminates in a report on how a noted astrophysicist, Dr. Brown, produced a computer model of how distant type-B2 stars might produce gamma-ray bursts. The model, called

'B2-Evo,' incorporates information about the relevant physical laws at play inside of type-B2 stars to produce simulations of how they produce gamma-ray bursts. The model generates the very surprising prediction that gamma ray bursts are preceded by gamma-ray bubbles, a prediction later confirmed by telescopic observations. The model is so complex that Dr. Brown initially has difficulty wrapping his head around why the simulated stars produce gamma rays. Dr. Brown publishes a detailed report of his findings and provides open access to his computer model.

In the Intelligible condition, Dr. Brown eventually figures out a way to rein in the model's complexity in order to understand the possible origin of gamma-ray bursts. In the Never Intelligible condition, Dr. Brown eventually concludes that his B2-Evo model is far too complex for anyone to use it to understand the origin of gamma-ray bursts.

After reading the distractor article on neuroscience, participants in the R-A condition were asked to specify, granting that the article is accurate, whether or not the following statements appear to be true given what they had read in the gamma-ray article:³

- Brown's paper and the accompanying computer model constitute an explanation for why type-B2 stars produce gamma-ray bursts.
- Brown understands how type-B2 stars at least might produce gamma ray bursts.
- Gamma-ray bursts were first detected by satellites as part of an attempt to monitor nuclear weapons testing.
- Sato's team used the Arecibo telescope in Puerto Rico.
- Brown's paper was rejected by the Journal of the American Astrophysics Association.

The first statement references a set of representational artifacts, and so we take responses to it to provide an implicit measure of the semantic activation of participants' concept of explanation_{r-a}. In the Objective condition, the first statement was replaced by, "The explanation for gamma-ray bursts is a physical process which also produces gamma-ray bubbles." This statement refers to a physical process, and so we take responses to it to indicate activation of participants' concept of explanation_o.

The remaining items were included to measure whether participants had read the article with sufficient attention to detail. The second was used to specifically measure if they correctly remembered whether or not Dr. Brown understands why-possibly gamma ray bursts are produced. Participants were next asked to supply demographic information.

Results

Of the 115 lay participants (MTurk workers) who began the experiment, we analyzed data from 106 participants.⁴

¹ All MTurk workers for this study were in the United States and had an 80%+ approval rate. They received \$.40 compensation.

² <http://theconversation.edu.au/flash-aah-aah-could-a-gamma-ray-burst-eradicate-all-life-on-earth-5291>

³ Order of presentation for test items was randomized.

⁴ Data from 9 participants were excluded, 5 for missing 2+ comprehension questions and 4 for prior exposure to the materials.

We also analyzed data from 112 scientists. As shown in Figure 1, consistent with our earlier work, significantly more lay participants remembered that Dr. Brown's materials constitute an explanation_{r-a} for gamma-ray bursts in the Intelligible condition than did so in the Never Intelligible condition ($\chi^2(1, n = 53) = 13.117, p < .001$). In contrast, there was no significant difference between the Intelligible and Never Intelligible conditions in the number of lay participants who remembered that the process described by Dr. Brown's materials was the explanation_o for gamma-ray bursts ($\chi^2(1, n = 53) < 1, \text{ns}$).

Similar results held for scientists. Significantly more scientists remembered that Dr. Brown's materials constitute an explanation_{r-a} for gamma-ray bursts in the Intelligible condition than did so in the Never Intelligible condition ($\chi^2(1, n = 57) = 5.058, p < .05$). There was no significant difference between the two conditions in the number of lay participants who remembered that the process described by Dr. Brown's materials was the explanation_o for gamma-ray bursts ($\chi^2(1, n = 55) = 3.16, \text{ns}$).

There was a significant difference between the number of participants who remembered that Dr. Brown understands how type-B2 stars could produce gamma ray bursts across the Intelligible and Never Intelligible conditions for both laypeople ($\chi^2(1, 106) = 29.150, p < .001$) and scientists ($\chi^2(1, 112) = 24.938, p < .001$).

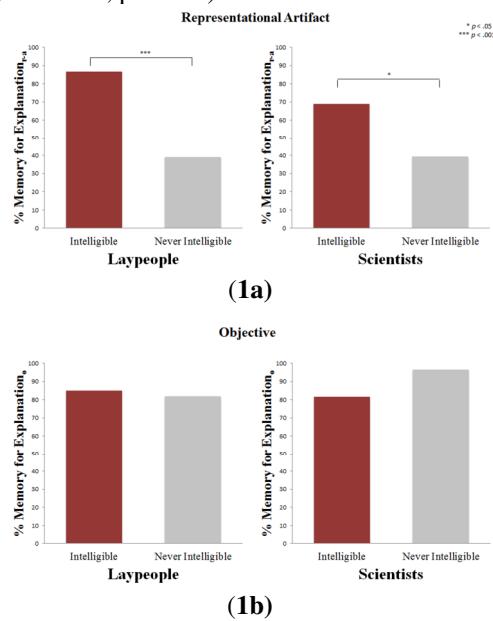


Figure 1. (1a) Percentage of participants remembering that the materials described constitute an explanation_{r-a}. (1b) Percentage remembering that the process described by the materials is the explanation_o.

Discussion of Experiment 1

Both lay and scientific participants were clearly sensitive to the main manipulation to the article – namely, the variation in whether or not Dr. Brown finds gamma-ray bursts intelligible. Experiment 1 shows that, for both populations, while memory for the explanation_{r-a} claim was

affected by the presence or absence of intelligibility, memory for the explanation_o claim was not. Our findings thus suggest that finding intelligible is central to explanation_{r-a} but not to explanation_o. It is noteworthy that the lay and scientific communities seem to be in agreement on this. Also important is that, in the Intelligible condition, a large percentage from both populations remembered, of the very same article, that either Dr. Brown's materials or the objective process described by them constitute an explanation. In other words, both populations were comfortable classifying representational artifacts and objective processes as explanations, supporting the notion that 'explanation' is ambiguous between these two senses.

Experiment 2: Explanation and Accuracy

Our second experiment studies the importance of accuracy to explanation_{r-a} and explanation_o. As mentioned, many philosophers consider accuracy to be an essential feature of explanation_{r-a}. Explanations_o, by contrast, are ostensibly just the actual facts (e.g., about how a phenomenon came about). Taken together, this suggests that an explanation_{r-a} will (with some caveats) necessarily elucidate the explanation_o for a phenomenon. One should expect, then, that varying the accuracy of Dr. Brown's materials will alter participants' memories for both explanation_{r-a} statements and explanation_o statements.

Participants

Participants in this study were 151 workers recruited using MTurk.⁵ For scientists, we recruited 111 practicing scientists (faculty members, postdocs, and graduate students) affiliated with the Beckman Institute for Advanced Research at the University of Illinois at Urbana-Champaign. As with Experiment 1, data from MTurk workers and scientists were analyzed separately.

Materials and Procedure

Materials were much like as in Experiment 1, except that the story did not end with Dr. Brown discussing B2-Evo's capacity to render gamma-ray bursts intelligible, but rather with mention of B2-Evo's competitor, Eigen-burst, and of how the two make very different predictions about the phenomena that accompany gamma ray bursts. These predictions were said to be tested, and depending on the condition Dr. Brown's model either passed (Crucial Test Pass) or fails (Crucial Test Fail) this test. This provides participants with potent indirect evidence of the accuracy of the materials provided by Dr. Brown.

The only other difference was that the second comprehension item from Experiment 1 was replaced with:

- Brown's B2-Evo model of gamma-ray burst formation is accurate.

⁵ Eligibility and compensation were as in Experiment 1.

Results

Of the 151 lay participants (MTurk workers) who began the experiment, we analyzed data from 128 participants.⁶ We also analyzed data from 111 of the 139 scientists who began the experiment.⁷

As shown in Figure 2, significantly more lay participants remembered that the process described by Dr. Brown's materials was the explanation_o for gamma-ray bursts in the Crucial Test Pass condition than in the Crucial Test Fail condition ($\chi^2(1, n = 63) = 6.87, p < .01$). The manipulation did not significantly affect how many participants remembered that Dr. Brown's materials constitute an explanation_{ra} for gamma-ray bursts ($\chi^2(1, n = 65) < 1, \text{ns}$).

As with laypeople, significantly more scientists remembered that the process described by Dr. Brown's materials was the explanation_o for gamma-ray bursts in the Crucial Test Pass condition than in the Crucial Test Fail condition ($\chi^2(1, n = 54) = 12.946, p < .001$). The manipulation did not significantly affect how many participants remembered that Dr. Brown's materials constitute an explanation_{ra} for gamma-ray bursts ($\chi^2(1, n = 57) < 1, \text{ns}$).

Significantly more participants remembered that the B2-Evo model as being accurate in the Crucial Test Pass than in the Crucial Test Fail condition, a finding that held for both laypeople ($\chi^2(1, 128) = 42.482, p < .001$) and scientists ($\chi^2(1, 111) = 37.282, p < .001$).

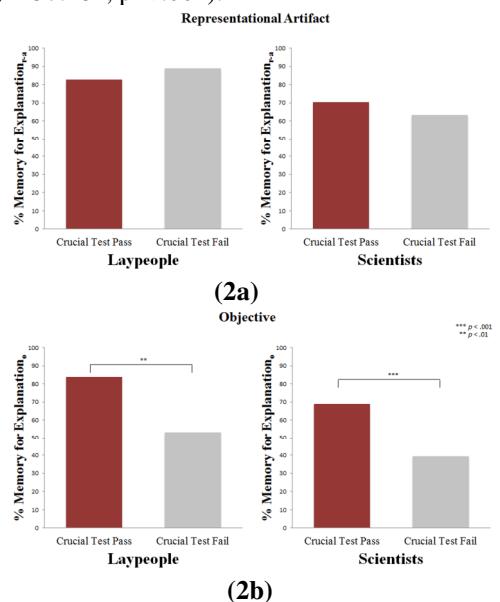


Figure 2. (2a) Percentage of participants remembering that the materials described constitute an explanation_{ra}. (2b) Percentage remembering that the process described by the materials is the explanation_o.

⁶ Data from 23 participants were excluded, 14 for missing 2+ comprehension questions and 9 for prior exposure to the materials.

⁷ 28 scientists were mistakenly given the comprehension item regarding intelligibility instead of accuracy. We only analyzed their memories for explanation_{ra} and explanation_o.

Discussion of Experiment 2

These results appear to contradict traditional philosophical claims that accuracy is necessary for explanation_{ra}. They also further distinguish explanation_{ra} from explanation_o in that representing the explanation_o appears not to be crucial for explanation_{ra}. One concern is that responses from participants in the Objective condition simply reflect memory that the model described in the story is accurate. There is something to this worry, for as noted above, a plausible a priori position is that people will not regard some specific process as the explanation_o for an occurrence unless there is some explanation_d or explanation_{ra} of it making reference to the process that we believe is accurate. However, it seems unlikely that subjects are *merely* reporting on whether or not the model is accurate given that the prompt never mentions a model at all and that the term 'explanation' is explicitly described as a 'physical process.' If using 'explanation' in this way were semantically infelicitous, then subjects would be far less likely than they were to remember the claim as being advanced in the story. It would be no part of the story, and nonsense to boot.

General Discussion

We noted at the outset that 'explanation' seems to be ambiguous between explanation_{ra}, explanation_o, and explanation_d. Some philosophers claim that the first two senses are dissociable from psychological states and that psychology can thus add little to the philosophical investigation of them. Some claim as well that accuracy is constitutive of explanation_{ra}, which seems to entail that any explanation_{ra} for a phenomenon accurately represents the explanation_o for it. Psychologists meanwhile have focused their efforts explanation_d.

Our goal is to bring clarity to the ambiguities of 'explanation' by supplementing tradition philosophical research with the methods of experimental psychology. Our previous research suggests that laypeople and scientists regard it as central to something being an explanation_{ra} that it has actually rendered the phenomenon intelligible to someone, which suggests that explanation_d is constitutive of explanation_{ra}. We wondered if explanation_d might also be constitutive of explanation_o, and so we examined, using more rigorous methods than employed previously, the importance of finding intelligible to explanation_{ra} and explanation_o. Our findings suggest that, among laypeople and scientists, finding intelligible is considered central to explanation_{ra} but not to explanation_o. In this regard explanation_{ra} and explanation_o appear to come apart.

We wondered as well if accuracy might provide a link between explanation_{ra} and explanation_o. We found, however, that laypeople and scientists are no less inclined to classify a representational artifact as an explanation_{ra} when it fails to track the explanation_o than when it succeeds. Those same populations were, however, less inclined to classify the objective conditions designated by a representational artifact as the explanation_o when that

artifact was inaccurate than when it was accurate. In this respect too, explanation_{r-a} seems closer to explanation_d than to explanation_o. These two sets of findings appear to be at odds with how many philosophers conceive of explanations. However, our studies utilize implicit measures of how large samples of laypeople and scientists conceive of explanations. We think that findings arrived at through studies such as ours will generally trump those arrived at through the more traditional methods discussed above.

Overall, our research bears out the intuitively plausible suggestion that there are distinct representational-artifact and objective senses of ‘explanation’ in use among laypeople and scientists, and it goes some way towards elucidating in what ways they are conceived of as being distinct. As concerns accuracy and intelligibility, we see no qualitative difference in how laypeople and professional scientists conceive of the different senses of ‘explanation.’ This brings to mind Wilfred Sellars’ (1956) claim that “science is continuous with common sense, and the ways in which the scientist seeks to explain empirical phenomena are refinements of the ways in which plain men, however crudely and schematically, have attempted to understand their environment and their fellow men...”

More work is clearly required in order to fully clarify the overlapping contours of each conception of explanation, especially that of explanation_d. The results of this empirical work clearly have much importance to philosophy of science, not only because a big part of the philosophical project is to determine how science actually works. For even knowing this, one might find, for instance, that one or more conceptions of explanation are ill suited to play the kind of roles they are slated for in science. This would provide some mandate for a revision to science’s conceptual or methodological practices, but this again presupposes as its starting point the kind of clear appreciation for standard conceptual practices revealed by studies such as ours. Another interesting possibility is that, once we have thoroughly clarified the various conceptions, an underlying unity to the philosophical and psychological investigations of explanation will become even more apparent. Indeed, it is worth stressing that we have not ruled out the possibility that there is an important psychological dimension even to explanation_o. It could be, for instance, that where an explanation_o for gamma-ray bursts defies human comprehension, the boundaries drawn around the specific set of objective conditions constituting the explanation_o are grounded in the interests of those asking the original “why?” question. It could be, in other words, that the boundaries around objectives explanation are set partly by nature and partly by *us*. This too is a hypothesis that lends itself to empirical study. Indeed, to the extent that one is concerned with the factual matter of how a given population conceives of explanations, psychological methods will be useful, and that is true even if the categories under scrutiny are purely objective.

- Achinstein, P. (1984). The Pragmatic Character of Explanation. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association: Symposia and Invited Papers*, 2, 275-292.
- Braverman, M., J. Clevenger, I. Harmon, A. Higgins, Z. Horne, J. Spino, and J. Waskan. (2012). Intelligibility is necessary for explanation but accuracy may not be. *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society*.
- Craik, K.J.W. (1943). *The Nature of Explanation*. Cambridge, UK: Cambridge University Press.
- Craver, C.F. (2007). *Explaining the Brain*. New York: Oxford University Press.
- Craver, C. (2014). The Ontic Account of Scientific Explanation. *Synthese Library* 367, DOI 10.1007/978-94-007-7563-3_2, 27-52.
- Cummins, R. (1998). Reflections on reflective equilibrium. In M. Depaul and W. Ramsey (Eds.), *Rethinking Intuition*. New York: Roman and Littlefield.
- Hempel, C.G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.
- Hempel, C.G., and P. Oppenheim. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15 (2), 135-175.
- Overton, J. 2013. "Explain" in scientific discourse. *Synthese*, 190, 1383-1405.
- Powell, D., Z. Horne, and N. Pinillos. (forthcoming). Semantic Integration as a method for investigating concepts. In J. Beebe (Ed), *Advances in Experimental Epistemology*. New York: Continuum Press.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
- Salmon, W. (1998). *Causality and explanation*. New York: Oxford University Press.
- Stace, W. T. (1935). Science and the explanation of phenomena. *Philosophy*, 10 (40), 409-427.
- Stotz, K., and P.E. Griffiths. (2008). Experimental philosophy of science. *Philosophy Compass* 3 (3), 507-521.
- Thagard, P. (2012). *The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change*. Cambridge: The MIT Press.
- Trout, J.D. 2007. The psychology of scientific explanation. *Philosophy Compass*, 2 (3), 564-591.
- Waskan, J. (2006). *Models and Cognition*. Cambridge, MA: MIT.
- Waskan, J., I. Harmon, Z. Horne, J. Spino, & J. Clevenger. (2013). Explanatory Anti-psychologism Overturned by Lay and Scientific Case Classifications. *Synthese*, (DOI 10.1007/s11229-013-0304-2).