A New Approach to Argument by Analogy: Extrapolation and Chain Graphs

Daniel Steel[†‡]

This essay sketches an account of extrapolation in terms of analogical reasoning that generalizes the notion of comparative process tracing. The account offered here relies on something known as chain graphs, which are used to explicate the role of "fingerprints," or distinctive markers, as a strategy for avoiding an underdetermination problem having to do with spurious analogies. Moreover, I show how the extrapolator's circle, one of the central challenges for extrapolation, is closely tied to distinctive markers and the Markov condition as it applies to chain graphs.

1. Introduction. In order to make scientific results relevant to practical decision making, it is often necessary to transfer a result obtained in one set of circumstances—an animal model, a computer simulation, an economic experiment—to another that may differ in relevant respects—for example, to humans, the global climate, or an auction. Such inferences, which we can call *extrapolations*, are a type of argument by analogy. This essay sketches a new approach to analogical inference that utilizes chain graphs, which resemble directed acyclic graphs (DAGs) except in allowing that nodes may be connected by lines as well as arrows. This chain graph approach generalizes the account of extrapolation provided in Steel (2008) and leads to new insights that integrate the contributions of the other participants of this symposium. More specifically, this approach explicates the role of "fingerprints," or distinctive markers, as a strategy for avoiding

†To contact the author, please write to: 503 S. Kedzie Hall, Michigan State University, East Lansing, MI 48824; e-mail: steel@msu.edu.

‡I would like to thank an anonymous referee, Francesco Guala, Wendy Parker, Marcel Weber, and participants in at the 2008 PSA symposium on Extrapolation and Public Policy for helpful comments on earlier drafts.

Philosophy of Science, 77 (December 2010) pp. 1058–1069. 0031-8248/2010/7705-0005$10.00
Copyright 2010 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/656543 Published online by Cambridge University Press

an underdetermination problem having to do with spurious analogies. Moreover, it shows how the extrapolator's circle, one of the central challenges for extrapolation, is closely tied to distinctive markers and the Markov condition as it applies to chain graphs. Finally, the approach suggests additional ways in which investigations of a model can provide information about a target that are illustrated by examples concerning nanomaterials in sunscreens and Parker's discussion of fingerprints in climate science.

2. Difference and the Extrapolator's Circle. Any adequate account of extrapolation must answer a pair of challenges, namely, the problem of difference and the extrapolator's circle. The problem of difference is that some causally relevant differences between the model and target are inevitable in most interesting examples of extrapolation in the biological and social sciences. Thus, an adequate account of extrapolation must explain how extrapolation can be justified even when some causally relevant differences are present. The extrapolator's circle rests on the observation that evidence is needed to show that the model is relevantly similar to the target and hence a good basis for extrapolation. For example, one might suppose that a good model should possess the same or very similar mechanisms as the target. But then showing that the model is relevantly similar might seem to require first learning the mechanism in both and then comparing them to demonstrate the similarity. But in that case, extrapolation from the model to the target would be redundant. The challenge posed by the extrapolator's circle, then, is to explain how it is possible to establish the similarity of the model and target without already knowing what one wants to extrapolate. The extrapolator's circle and the problem of difference are presented by Hugh LaFollette and Niall Shanks (1996) as arguments for the claim that extrapolation from animal models to humans is never justified.

One answer to the problem of difference is that the closeness of match required between the model and target depends on the specificity of the causal claim that one wishes to extrapolate (Steel 2008, 93–94). The greater the specificity of the causal claim, the closer the match between model and target must be. If the goal is to extrapolate the exact probability distribution of the effect when the value of the cause is set by an intervention, then a complete absence of causally relevant differences may indeed be required. But for some types of probabilistic causal claim—particularly claims about positive causal relevance—extrapolation can be justified even when a good deal of causally relevant difference is present. Thus, even if we cannot extrapolate the exact causal effect from model to target, we might still be justified in extrapolating the claim that (say) exposure to aflatoxin B_1 increases the chance of liver cancer.

https://doi.org/10.1086/656543 Published online by Cambridge University Press

An answer to the extrapolator's circle needs to explain how limited information concerning the target can be efficiently used to support the relevant similarity of the model. This can be done within the context of a mechanisms-based approach to extrapolation by taking note of two key strategies for reducing the nodes or stages of the mechanisms that need to be compared in the model and target. First, we might possess background knowledge according to which causally relevant disanalogies are likely to be found at some stages of the mechanism and not others. If this is so, we can make our comparison more efficient by focusing only on the points of likely relevant difference. However, in many and perhaps most cases of extrapolation in biology and social science, data about the target needed for even this more limited set of comparisons may not be accessible. That leads to the second strategy. A difference in a mechanism matters to the outcome only if it has an impact on subsequent steps along the way. Hence, comparisons of model and target mechanisms will be more efficient if they focus on mechanism activities and components that are downstream in the sense of being more direct causes of the outcome. In some cases, it may be possible to measure a key downstream stage upon which any upstream difference must leave its mark. In that case, only one point of comparison between model and target mechanisms may be needed. These two strategies for making model-target mechanism comparisons more efficient are discussed in Steel (2008, 88–92) under the label "comparative process tracing."

Since extrapolation is a type of analogy, it is worth asking how comparative process tracing relates to analogical inference. This question is not merely interesting from a theoretical perspective. It is also relevant to extending comparative process tracing to further cases, such as climate change and experimental economics.

3. Analogy and Chain Graphs. In this section, I present an account of analogical reasoning that provides a more general perspective on extrapolation than comparative process tracing. One central theme of this account is a distinction between what I call the basis of an analogy and types of evidence that could be given to show that the basis is present. I then explain how the notion of a basis of an analogy can be usefully represented by means of something called chain graphs, which then can be used to clarify a number of aspects of analogical reasoning that are pertinent to extrapolation.

Stripped to its bare bones, the logic of arguments by analogy is the following:

1. The model, m, and target, t, are relevantly similar.
2. Φ is true of m.

https://doi.org/10.1086/656543 Published online by Cambridge University Press

3. Therefore, Φ is probably also true of *t*.

There are two central questions for analogical inferences: what does it mean for the model and target to be relevantly similar? and how do we know whether that relevant similarity, whatever it consists in, is present? I will call whatever features constitute the relevant similarity of the model and target the "basis" of the analogy. Clearly, the basis varies with Φ. Some similarities can always be found among any pair of objects, but the basis of an analogy requires a very particular sort of similarity, namely, one that matters to Φ. In the ensuing discussion, I will assume that Φ describes a cause-and-effect relationship. In this case, the relevant similarities would be features that make a difference to that causal link. Since mechanisms are thought to underlie causal relationships in complex systems such as organisms or social groups, they are a sensible candidate for the basis. Which mechanisms and which aspects of those mechanisms form the basis of the analogy would depend on further details of Φ. Moreover, as discussed in the foregoing section, how similar is "similar enough" will often depend on the specificity of that causal claim. I do not assert that mechanisms are the only possible basis for an analogy. In some cases, the relevant similarity might be largely functional. For instance, familiar technological devices—such as televisions or cars—often adopt similar user-interface designs even when underlying mechanisms differ. Hence, knowledge about the effects of, say, pressing the pedal farthest to the right on the driver's side (i.e., the accelerator) can be extrapolated among cars with distinct types of engines. However, for the purposes of this discussion, I will assume that the basis for the analogy consists of relevantly similar mechanisms.

A good argument by analogy should make reasonably clear what the basis is and provide evidence that this basis is present in the model and target. The most direct approach for establishing similarity of mechanisms is a step-by-step comparison of components and interactions among those components. For example, suppose you wanted to know whether two cars of distinct brands and appearance are powered by similar types of engines. The most direct way to answer this question would be to look under the hoods of the two cars and check for a correspondence in the arrangement of the working parts. But in many cases this direct "look and see" approach is not feasible, and it may be necessary to rely on more indirect evidence of similarities of mechanisms. For example, instead of directly comparing a subprocess of the mechanisms in the model and target, one might compare a side effect of that subprocess. That idea is illustrated by the studies of the carcinogen aflatoxin B_1, wherein by-products of the mutagenic effects of that compound (DNA adducts circulating in the bloodstream) served a point of comparison among humans and several

https://doi.org/10.1086/656543 Published online by Cambridge University Press

animal models (Steel 2008, 91–92). The DNA adducts were especially valuable in this case because they were indicators of a crucial downstream stage of the carcinogenic mechanism. Guala's discussion of the outer continental shelf (OCS) example in his contribution to this symposium is similar insofar as focusing on characteristic effects of auction rules, which he represents as a downstream stage of the mechanism in that case (see his fig. 2).[1]

In some cases, effects of mechanisms might constitute evidence of relevant similarity or difference even when we are not able to link those effects to a particular subprocess of the mechanism. For example, the clatter of a diesel versus the nearly inaudible whir of an electric motor might enable you to distinguish one type of engine from the other even if you are ignorant of the components and activities of the mechanisms that are responsible for these distinct sounds. Thus, one might infer by their sounds that two engines are of the same type and then proceed to use that similarity as a basis for analogical inferences. In addition, similar causes as well as similar effects can provide evidence that the basis is present in both model and target. For instance, we might think that two cars possess the same type of engine because they rolled off the same assembly line of the same factory.

There are, then, a variety of types of evidence that could be called upon to support the basis for an analogy. In this section, I will focus on a challenge for establishing a basis of an analogy that arises when evidence consists of similar effects of the model and target. Similar effects of model and target provide evidence of a particular basis only if there is reason to think that those effects are consequences of that basis and not something else. More precisely, if B is the proposed basis of the analogy, then a spurious analogy is a similarity of effects of the model and target that is not a result of B holding true of both model and target. For example, a whirring sound emanating from two cars might be taken as evidence that both have battery-powered electric engines when, in fact, one is powered by batteries and the other by hydrogen fuel cells. Thus, if B is the presence of a battery-powered electric motor and the effect is the whirring sound, then this is a case of a spurious analogy. This spurious analogy might lead to various mistaken inferences, for instance, that the hydrogen fuel cell car should be plugged into an electrical outlet overnight to recharge.

The concept of spurious analogy can be perspicuously represented by means of chain graphs. Chain graphs are a type of graphical represen-

1. However, this point about downstream stages and their relevance to the extrapolator's circle is not explained in Guala (2005). Therefore, I stand by my claim (Steel 2008, 150–53) that Guala had provided no answer to the extrapolator's circle.

https://doi.org/10.1086/656543 Published online by Cambridge University Press

tation in which nodes may be connected either by lines (undirected edges) or arrows (directed edges), for example, $A \leftarrow B - C \rightarrow D$. Chain graphs have received far less attention than DAGs, which may be because it is unclear just what chain graphs ought to be used for. One proposal is that chain graphs can represent equilibrium relations of dynamic feedback models (cf. Lauritzen and Richardson 2002). Here I propose that chain graphs may also be useful for representing analogical inferences. Analogies involve judgments that one thing is similar to another, and hence that information about the one tells you something about the other. For example, if A is similar to B, then learning the causes and effects of A may tell you something about B and its likely causes and effects. Yet, similarity is plainly a symmetrical relationship and is therefore much more naturally represented by a line than an arrow. Chain graphs, then, are a natural format for representing analogical reasoning, which involves judgments about similarity as well as causation.

It is important to distinguish the sense of similarity pertinent to analogy, which is represented by a line linking two nodes in a chain graph, from a mere coincidence. Similarities of the sort relevant to analogical reasoning differ from mere coincidences in that they provide information about one another. For example, if I know that Tom and Fred have similar temperaments, and I subsequently learn that Tom is prone to emotional outbursts, then I will think it more probable that Fred is prone to emotional outbursts as well. I will refer to mutually informative similarities of this sort as "analogies." Coincidences, by contrast, are similar outcomes resulting from unrelated or independent processes. For instance, if you and I both flip a coin, there is a one-quarter chance that both will come up heads, but this outcome would not constitute an analogy. Prior to learning the outcome of my coin toss, the probability of heads in your toss is one-half, and after I flip my coin and see that it is heads, the probability that yours will come up heads remains one-half. Coincidental outcomes, then, happen to be similar in a particular case, but one outcome makes no difference to the probability of the other. In addition, a line drawn between two nodes in a chain graph represents an analogy that is unmediated by other nodes in the graph. For instance, in the chain graph $A \leftarrow B - C \rightarrow D$, the analogy between B and C makes A and D mutually informative. Nevertheless, so long as this relationship between A and D is due solely to their connections to B and C, respectively, no additional line is drawn directly between A and D.

Consider the use of chain graphs to represent analogical reasoning conjunction with the concept of a basis of an analogy. Let $B(m)$ be a variable that indicates the basis in the model, m. If the basis for the analogy is a particular type of mechanism, then $B(m)$ might equal one if that mechanism is present and zero otherwise. For example, consider the OCS

https://doi.org/10.1086/656543 Published online by Cambridge University Press

case mentioned above (Guala 2005, 189–92). This example concerned the explanation of lower-than-anticipated profits ensuing from offshore oil-drilling leases purchased at auctions. One explanation was the "winner's curse," according to which bidders fail to adequately take into account the fact that, if they are the highest bidder, then it is likely that they have overestimated the value of the lease. However, the winner's curse assumes that bidders violate some rationality constraints often presumed in economic models of what are known as "common value auctions" (i.e., auctions in which the item on the block has a constant but unknown value for all bidders). Thus, in this example, $B(m) = 1$ if the winner's curse explanation is correct in the model, while $B(m) = 0$ if the rationality assumptions typical of economic models of common value auctions are correct in the model. The basis for the target, $B(t)$, would be defined similarly.[2] If there is an analogy between model and target, then learning $B(m) = 1$ increases the probability that the same is true of $B(t)$.

What evidence, then, would justify drawing a line between $B(m)$ and $B(t)$ in the chain graph and thereby asserting a basis for analogy? In the OCS example, the models were economic experiments conducted in laboratory settings, while the targets were real auctions of leases for offshore oil drilling. The basis in this case consisted of the auctioning mechanism, which depended in turn on the auction procedures, the type of item being auctioned, as well as the cognitive processes of the bidders. Some of these, such as the auction procedures, can be verified directly, but the same is not true for others, especially the reasoning processes of those making the bids. Thus, this key aspect of the analogy between $B(m)$ and $B(t)$ would have to be supported by more indirect evidence. One possible type of evidence would be similar effects of $B(m)$ and $B(t)$, particularly effects that are linked to whether the bidders make adjustments for the possibility that their information overstates the value of the item. Guala describes two effects of this kind that correspond to differing predictions made by the two potential mechanisms (2005, 191–93). The first is that larger numbers of bidders in the auction are associated with higher bids. Let $E_1(m) = 1$ indicate the presence of this association in the experimental model and $E_1(m) = 0$ its absence, and likewise for $E_1(t)$ with regard to the real world auctions. The winner's curse mechanism predicts that this effect is present, while the rational choice model predicts that this effect is absent. The second effect is that public information concerning the value of the item increases profits. Again, let $E_2(m) = 1$ indicate that this effect is present in the experimental model and $E_2(m) = 0$ indicate its absence,

2. For simplicity and ease of illustration, I assume here that the basis has only two possible values. However, $B(m)$ and $B(t)$ could potentially have as many values as one would like.

https://doi.org/10.1086/656543 Published online by Cambridge University Press

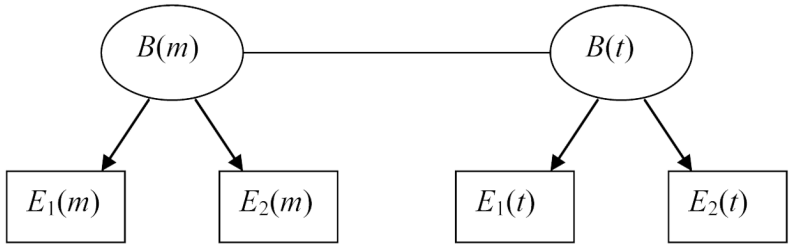


Figure 1. Chain graph to represent the analogical reasoning in the outer continental shelf example. Variables in rectangles are directly observed, while those in ovals are not.

and likewise for $E_1(t)$. As before, the winner's curse hypothesis predicts that this effect is present, while the rational choice alternative predicts the opposite. Both of these effects were found in laboratory experiments and in field data concerning real-life auctions of oil-drilling leases (Guala 2005, 192). These results can be explained by the chain graph in figure 1. According this graph, the analogous winner's curse mechanisms in the model and target systems account for the similar results in each. One important consequence of this graph is that further effects of the winner's curse found in the model may provide evidence for similar effects in the target.

However, the chain graph in figure 1 is not the only possible explanation in this case. One alternative would be that the common effects are a mere coincidence and hence not due to any underlying similarity of the model and target at all. In that case, the similarity of E_1 and E_2 in the model and target would provide no basis for expecting correspondences in other effects. But even if the similarity of effects in the model and target is not a coincidence, it could still be due to an analogy other than the winner's curse. In this case, there may be some grounds for predicting similarities with regard to further effects, although not necessarily the same effects as if the winner's curse were the shared basis. These two alternatives are represented by the chain graphs in figures 2 and 3, respectively. In these graphs, $A(m)$ and $A(t)$ represent alternative causes that impact E_1 and E_2. In both of these cases, the analogy would be spurious.

Both of these alternatives depend on the possibility that there exist causes of E_1 and E_2 in addition to the winner's curse. If we knew or could somehow show that such alternative causes do not exist, then we could be confident that the chain graph in figure 1 is correct. This, I think, is precisely where Parker's concept of fingerprints, or distinctive markers, of a causal process enters the picture. A distinctive marker is a telltale indicator of a particular cause, as fingerprints are indicators of the manual

https://doi.org/10.1086/656543 Published online by Cambridge University Press

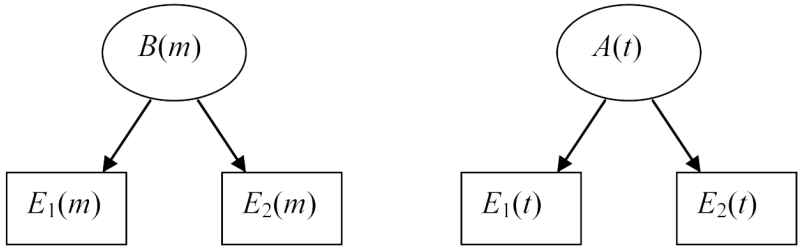


Figure 2. Corresponding effects in the model and target are a coincidence, that is, the result of very different mechanisms.

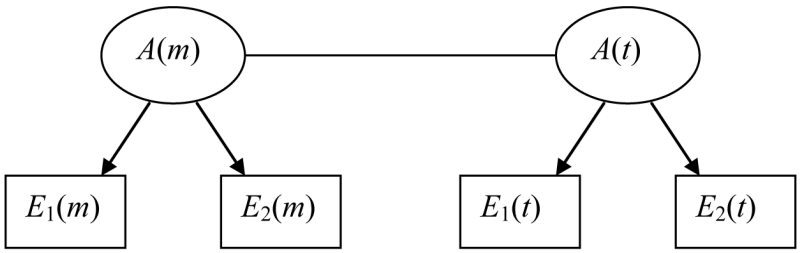


Figure 3. Corresponding effects in model and target are not due to an analogy between $B(m)$ and $B(t)$ but rather to a distinct analogy between $A(m)$ and $A(t)$.

contact of a particular person. Thus, if E_1 and E_2 are distinctive markers of the basis B, then we can be practically certain that B is present in the model and target when both have these effects. But on what basis could we claim that an effect, or set of effects, is a distinctive marker? In the next section, I consider that question and show that there is a surprising connection between distinctive markers and the extrapolator's circle.

4. Fingerprints and Circles. One reason to establish an analogy, like the one represented by the chain graph in figure 1, is to enable the model to serve as grounds for further inferences about the target. This idea is represented in figure 4. As before, values of variables inside rectangles are observed, while the values of those inside ovals are inferred, so in figure 4, an additional effect, E_3, has been observed in the model but not in the target. Thus, given the analogy between $B(m)$ and $B(t)$, the observation of the value $E_3(m)$ provides evidence concerning the value of $E_3(t)$. The usefulness of the chain graph representation of analogical inference becomes especially apparent at this point, since it provides a framework for

https://doi.org/10.1086/656543 Published online by Cambridge University Press

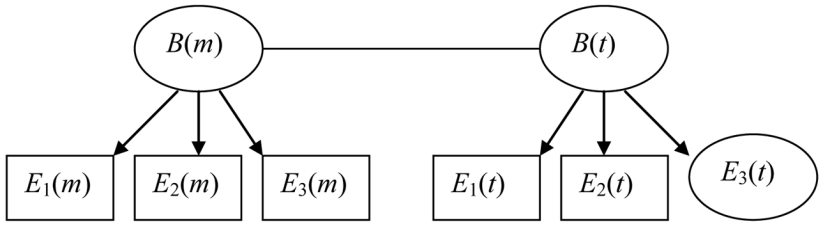


Figure 4. Analogy between $B(m)$ and $B(t)$ may allow an observation of the value of $E_3(m)$ to predict the value of $E_3(t)$.

deciding which variables in a graph are probabilistically associated with which others. This leads to a more precise formulation of the extrapolator's circle; demonstrates a close tie between the extrapolator's circle and fingerprints; and, finally, suggests an additional way in which models can provide useful information about a target.

In the previous section, we noted that fingerprints, or distinctive markers, suggest an appealing strategy for avoiding spurious analogies, which are understood as a type of underdetermination problem within our chain graph approach. A distinctive marker is an effect that could only have resulted from the hypothesized basis of analogy. Hence, if E_1 and E_2 are distinctive markers of B, then they make B practically certain. However, this quickly leads to a difficulty for the idea that learning further effects in the model will tell us something about further effects in the target. In particular, it is natural to suppose that probability distributions associated with chain graphs will obey a version of the Markov condition (cf. Lauritzen and Richardson 2002). One familiar consequence of the Markov condition for DAGs is that common causes screen off their effects. For example, in the graph $A \leftarrow B \rightarrow C$, A and C are probabilistically independent conditional on B, or more colloquially, once we know B, learning A provides no additional information about C, and vice versa. Although the Markov condition for chain graphs differs from that for DAGs, it has a very similar consequence, which we can state as "a basis of an analogy screens off its effects." For instance, in the chain graph $A \leftarrow B - C \rightarrow D$, A and D are probabilistically independent conditional on B. Once we know B, A no longer tells us anything about D, and vice versa. Similarly, in figure 4, $B(t)$ screens off $E_3(m)$ from $E_3(t)$. But if the previous two effects are distinctive markers of B, then we know the value of $B(t)$ with practical certainty, which means that subsequently learning the value of $E_3(m)$ provides no further information about $E_3(t)$.

This result makes the same point as the extrapolator's circle: evidence sufficient to establish the analogy between model and target may also

https://doi.org/10.1086/656543 Published online by Cambridge University Press

render any further information about the model redundant. The solution to the extrapolator's circle proposed in Steel (2008) pointed out that the effects may be indicators only for a key stage or component of the mechanism rather than an indicator of the whole thing. In the aflatoxin B_1 example, the effect consisted of the DNA adducts circulating in the bloodstream among those exposed to aflatoxin B_1. These DNA adducts consist of an aflatoxin B_1 molecule covalently bonded to a fragment of DNA and hence can reasonably be treated as distinctive markers of an interaction between aflatoxin B_1 and the DNA of the organism. However, the DNA adducts are not distinctive markers of the mechanism as a whole, since they do not tell us what carcinogenic effects, if any, that interaction has.

So far, we have assumed that we knew about the effects of the basis but were unsure about the analogy between the model and target. Hence, we tried to establish the analogy by comparing effects. Another possibility is that we have strong antecedent reasons for the hypothesized analogy between model and target but are uncertain of the effects of the basis. In this situation, the purpose of studying the model would be to discover what the effects of the basis are. Once those effects are discovered in the model, we can infer from the analogy that they are likely to be present in the target as well. This idea is nicely illustrated by a recent study on, of all things, the effects of sunscreens containing nanoscale materials on prefabricated painted metal roofing in Australia.

Since around 2006, complaints of unsightly blemishes and patches and unusually rapid deterioration have arisen from Australians whose homes were fitted with newly installed prepainted steel roofs. The blotches often took the shape of boot or handprints, clearly suggesting a cause involving contact with workers responsible for the installation. One possible explanation was that the premature damage resulted from new materials contained in sunscreens that may have been used by the workers. Two Australian industrial scientists, Philip Barker and Amos Branch, constructed an experimental model to test this hypothesis: they purchased a representative sample of over-the-counter sunscreens and applied them to commercially available prepainted steel roofs that were then exposed to the Australian sun (Barker and Branch 2008). There can be little serious doubt about the analogy between the experimental model and the roofs that constituted the target in this case, since the materials were obtained from the same sources and subjected to the same environmental conditions. Indeed, this example nicely illustrates the observation made in the previous section that a common origin of the model and target can constitute evidence for an analogy. The question in this case was whether the damage to the metal roofs was due to some ingredient in the sunscreen and, if so, which one. Barker and Branch found that the premature weathering oc-

https://doi.org/10.1086/656543 Published online by Cambridge University Press

curred only in roofing materials treated with sunscreens containing nanoscale titanium or zinc oxide, and they carefully traced the mechanism through which this damage occurred.

Consider how this example contrasts with the aflatoxin B_1 example. In the aflatoxin example, the status of DNA adducts as a distinctive marker of an interaction between aflatoxin B_1 and the organism's DNA could be inferred from already accepted background knowledge. Comparisons of this distinctive marker, then, were used to support the analogy between the model and target. In the roofing example, the analogy was antecedently well founded, but experimentation on the model was necessary to discover that the "unsightly defects" in the prepainted steel roofing were distinctive markers of a photocatalytic mechanism involving nanoscale materials included as ingredients in some modern sunscreens. Thus, prior knowledge of a distinctive marker can be used to provide evidence of an analogy, and prior knowledge of an analogy can be used to discover that an effect is a distinctive marker of a particular cause. I think that the argumentative strategy of climate scientists described in Parker's contribution combines elements of both of these cases. The climate scientists attempt to discover distinctive markers by means of experimentation with their models, and they also use comparisons involving those distinctive markers to support the analogy between the climate models and the real climate.

5. Conclusion. Extrapolation is a type of reasoning by analogy that is important for policy relevant science. Therefore, a better understanding of the logic of analogical reasoning— especially as it relates to the problem of difference and the extrapolator's circle—is of philosophical and practical value. In this essay, I have attempted to make some initial steps in this direction by explaining how chain graphs are a useful format for representing analogical reasoning.

REFERENCES

Barker, Philip, and Amos Branch. 2008. "The Interaction of Modern Sunscreen Formulations with Surface Coatings." *Progress in Organic Coatings* 62:313–20.

Guala, Francesco. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

LaFollette, Hugh, and Niall Shanks. 1996. *Brute Science: Dilemmas of Animal Extrapolation*. New York: Routledge.

Lauritzen, Steffen, and Thomas Richardson. 2002. "Chain Graph Models and Their Causal Interpretations." *Journal of the Royal Statistical Society* 64 (3): q321–61.

Steel, Daniel. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. New York: Oxford University Press.

https://doi.org/10.1086/656543 Published online by Cambridge University Press