In theories of scientific representation and investigation, metaphor has long been treated as a form of alchemy, with one of two divergent attitudes. The celebratory camp, represented by the likes of Shelley, Nietzsche, and Mary Hesse, takes metaphor be distinctively equipped to achieve a mystical communion with Nature, not just unlocking the universe’s secrets but creating new worlds. Often, this group takes all language and thought to be ultimately metaphorical, or at least takes metaphor to be the truest embodiment of the basic mechanisms by which reference, truth, and understanding are achieved. Against such seeming excesses of representation and ontology, the dismissive camp, represented by the likes of Hobbes, Locke, and Zenon Pylyshyn, treats metaphor superstitiously positing occult entities and forces that don’t correspond to anything real. At best, metaphor is a decorative trope, or a causal source of inspiration; at its worst, it spins a bubble of self-confirming pseudo-science.

This opposition appears especially stark in the context of a positivist conception of science as the logical subsumption of observation sentences under general theoretical laws. Few endorse this conception today. Since well before Kuhn (1962), philosophers of science have noted that scientists bring a host of only partially articulated theoretical and empirical assumptions to bear in investigating the world, and that distinct patterns of attention and explanation strongly influence the interpretation of any particular bit of data. A more recent trend, exemplified by Ronald Giere, Peter Godfrey-Smith, and Michael Weisberg, points to the crucial role of intermediate constructions – ‘models’ – which are known to differ from the actual world in various significant respects.

Both developments have had the salutary effects of dispelling a false picture of science as transparent description embedded in logical relations, and of enriching our understanding of scientific investigation, representation, and justification to bring them more in line with actual scientific practice. Less directly, they have also enriched our understanding of rationality, by demonstrating an essential role for imagination within a paradigm case of rational inquiry. However, theorists who advocate a less simplistic view of scientific theorizing often lump together multiple types of indirect representation under a general undifferentiated banner of ‘models’. Further, some of these theorists, in their zeal to oppose a naively descriptivist realism, have sometimes concluded that all theories are mere fictions levied in the service of competing pragmatic interests.

1 Thanks for useful and enjoyable discussion to audiences at the philosophy departments at Indiana, Harvard, St. Andrews, and LOGOS Barcelona, at the Rutgers Philosophy of Science Reading Group, and at the Metaphors in Use Conference (Lehigh) and Varieties of Understanding Conference (Fordham). Individual thanks to Jordi Cat, Catherine Elgin, Peter Godfrey-Smith, Arnon Levy, Deborah Marber, Matthew Slater, and Isaac Wilhelm for helpful discussion. Special thanks to Michael Weisberg for many illuminating conversations about models, metaphors, and science over multiple years. Finally, thanks to Stephen Laurence for drawing the especially elegant, easily reproducible version of Figure 1.
In this chapter, I distinguish among a range of distinct representational devices, which I call ‘frames’, all of which are used to guide our overall interpretation of a subject, by providing an intuitive principle for noticing, explaining, and responding to it. Frames play a theoretical role closely akin to that ascribed to models. But where much discussion of models focuses on their ontological status and representational relation to reality, I focus more the cognitive structures and abilities that frames generate, and on the imaginative activities that exploit them. Further, where many theorists of modeling have aimed to explain models in terms of a unified type of representational relation, I focus on identifying distinct ways that scientists fruitfully depart from representing ‘the truth, the whole truth and nothing but the truth’. In particular, where much recent discussion of models has drawn inspiration from fiction, I focus on metaphor.

My aim here is primarily descriptive: I want to explain what frames have in common which makes them powerful cognitive tools; the various ways in which they can work; and the commonalities and differences between their application in everyday cognition and in scientific inquiry. I believe that the account I provide here (drawing on previous work) also provides the resources to establish norms for evaluating the aptness of particular frames, which we can use to diagnose the epistemic benefits and perils of frames in general. I also believe that this account justifies an important epistemic role of frames within the process of scientific inquiry – and even at the nominal end of inquiry. But establishing these normative consequences is a task for another occasion.

The chapter proceeds as follows. In §1, I use metaphor to introduce the broader family of frames. In §2, I distinguish metaphor from some of its close cousins, especially telling details, just-so stories, and analogies, as they function in the context of ordinary cognition and communication. In §3, I illustrate these various species of frame within the context of scientific inquiry, and use them to illustrate important differences in the sorts of gaps that models open up between representation and reality. I conclude briefly in §4 by advocating a form of mild ecumenicalism about scientific models: although models are deployed in support of importantly similar cognitive and epistemic functions, there is no single ontological status or representational relation common to all models.

1. Frames and Characterizations

Begin with perhaps the most influential metaphor about metaphor in recent analytic philosophy, from Max Black (1954):

Suppose I look at the night sky through a piece of heavily smoked glass on which certain lines have been left clear. Then I shall see only the stars that can be made to lie on the lines previously prepared upon the screen, and the stars I do see will be seen as organised by the screen’s structure. We can think of a metaphor as such a screen, and the system of ‘associated commonplaces’ of the focal word as the network of lines upon the screen. We can say that the principal subject is ‘seen through’ the metaphorical expression – or, if we prefer, that the principal subject is ‘projected upon’ the field of
I think that this passage expresses an insightful and basically correct view of metaphor. But it is unsatisfying, in two ways. First, there is the problem of explicitness. Because it is itself a metaphor, Black’s image of smoked glass etched with clear lines does not directly articulate a claim about how metaphor works; further, the subsequent paraphrases or elucidations introduce distinct metaphors, not all clearly consistent. So at a minimum, we need to spell out what talk of ‘screens’ and ‘projections’, ‘seeing through’ and ‘organizing structure’, amount to.

Second, there is the problem of distinctiveness. In the paragraph preceding the quoted passage, Black articulates the core idea in less metaphorical language, saying that “The …metaphor suppresses some details, emphasizes others – in short, organizes our view of [the topic].” While this is more explicit, it also characterizes a range of other rhetorical tropes that ‘frame’ and ‘filter’ thought. I think this is an important insight to be gleaned from Black’s remarks, not (just) a weakness. In this section, I spell out Black’s talk of frames as ‘organizing structures’ in my own terms, as it applies to all these cases.

In everyday cognition, we frequently engage with the world using complex, intuitive ways of thinking about a subject, which I call characterizations (Camp 2003, 2015). The most obvious instances are stereotypes – Black’s ‘systems of associated commonplaces’. But where stereotypes are culturally ubiquitous, characterizations can be much more culturally restricted: limited to a sub-discipline, a clique, even two interlocutors in a particular conversation. In many cases, especially those relevant to science, characterizations are close to what philosophers sometimes call ‘conceptions’: a set of beliefs about an individual or a kind, which need not be extension-determining, or constitutive of conceptual competence, or even reflectively endorsed by the agent; but which are easily evoked in thinking about the subject, and which provide the intuitive “mental setting” (Woodfield 1991, 551) or background against which specific beliefs and questions are formulated.

Most characterizations are relatively inchoate and largely tacit: an intuitive patchwork of more or less unreflective assumptions. They also tend to be highly malleable, depending on the issues, interests, and contrasts that happen to be operative within the current context. In order to impose more coherence and stability on our own intuitive thinking, and in order to coordinate intuitive assumptions within communication, we frequently employ interpretive frames. As I will use the term, frames are representational vehicles – for example, a slogan, a diagram, or a caricaturing cartoon – under an intended interpretation, where that interpretation itself functions as an open-ended principle for understanding their target subjects.

Metaphors constitute a canonical class of framing device. But there are many other types of frames, even just among verbal representations. Among the most notable other verbal framing devices are slurs, as in ‘Well you know, he’s an S’ (Camp 2013); ‘telling details’, as in ‘Well, I’m just saying that
Obama’s middle name is Hussein’ (Camp 2008); and ‘just so stories’, as in ‘It’s as if John lost out on his prom date to the football captain and has been trying to make up for it ever since’ (Camp 2009). Although these rhetorical tropes differ in important ways, what they all have in common, in virtue of which they function as frames, is that they proffer an overarching interpretive principle which organizes one’s intuitive thinking about the target subject. More specifically, all frames perform three related functions: they guide what information an agent remembers about the subject, how the agent assimilates and explains new information into their existing assumptions, and in turn how they evaluate and respond to that information.

Frames are ubiquitous in ordinary life: we observe them at work in political discourse, in intimate interpersonal arguments, in informal commentaries on movies – anywhere interpretation is at stake and potentially contested. Three features of frames are especially important, both for understanding their operations in general and for explaining their functions within science in particular.

First, a frame presupposes a taxonomy: a basic level of analysis which partitions the domain of relevant entities into a space of contrasting possibilities, and which often also entails superordinate and subordinate classifications relative to that basic level (Rosch 1978). This taxonomy in turn determines, at least roughly, what sorts of features are relevant for classifying individuals and kinds within the taxonomy, and which features can and should be ignored. Such taxonomic presupposition is inherent in all conceptual representation. But because ordinary thought and talk tends to be especially concerned with people and ‘middle-sized dry goods’, ordinary frames typically assume a fairly similar level of analysis, even if they differ significantly in how they partition objects at that level, and what explanatory weight they assign to those partitions. (So, for instance, slurs and thick terms like ‘slut’ are objectionable partly in virtue of their focusing on race, sex, or other properties as purportedly useful classifications.) Within science, different disciplines and sub-disciplines employ more radically divergent taxonomies, both in their operative scale and in the partition they employ at a given scale. Thus, one obvious source of diversity among scientific frames, which can make conflicts between apparently competing claims especially difficult to resolve, is the presupposition of dramatically distinct taxonomies.

Second, in everyday cognition, frames often raise to attention and/or impute experientially vivid representations of highly specific features: for instance, that George has, or that people of that kind tend to have, this sort of nose or eyes. Further, characterizations often represent features in ways that are affectively and evaluatively loaded: that such noses are elegant, or haughty; or that George is handsome, or sleazy. Different frames ‘color’ the features they attribute to their subjects differently, by linking experiential, affective, and evaluative responses in intimate, intuitive ways (Camp 2015). Because scientific investigation and theorizing attempts to eschew affect and normative evaluation, images, feelings, and norms tend to play a comparatively diminished role in scientific thought. However,
scientists do still often interact with their target subjects in experientially rich, sustained, and highly invested ways. Given this, it is unsurprising that their characterizations of those subjects still involve images, feelings and norms at least to some degree – whether we take the epistemic consequences of this to be positive, negative, or mixed.

Third and most importantly, frames structure our intuitive thinking about a subject. Merely presupposing a taxonomy and/or coloring some of a subject’s features doesn’t suffice for framing, at least not in the rich sense at stake here. Rather, a representation functions as a frame when an agent uses it to organize their overall intuitive thinking about it. A frame doesn’t merely select a certain feature or features from the teeming mass of details about an object or kind as relevant for classification; nor does it merely evaluate or color some particular feature(s). Instead, a frame functions as an overarching, open-ended interpretive principle: it purports to determine, for any property that might be ascribed to the subject, both whether and how it matters.

To understand what cognitive and epistemic implications such an overarching interpretive principle might have, we need to say more about how frames select and structure a disparate set of features into a coherent whole. There are (at least) two distinct ways in which a feature can matter in an agent’s thinking about a subject (Camp 2003, 2013, 2015). First, some features are more prominent than others, in the sense of being more noticeable, or quicker to come to mind. I take prominence to be equivalent to what Tversky (1977) calls ‘salience’, which he in turn analyzes as a function of two factors, each contextually relative in a different respect. On the one hand, a feature is diagnostic to the extent that it is useful for classifying objects in a given context, as the elliptical shape of a snake’s pupils might be useful for determining whether it is poisonous. Because diagnosticity is taxonomy-relative, frames employing distinct taxonomies draw intuitive attention to distinct features, and/or assign distinct diagnostic implications to the same feature. On the other hand, a feature is intense to the extent that it has a high signal-to-noise ratio. What counts as ‘noise’ – the background against which the signal is measured – varies widely in interpretation, both in how locally restricted the relevant background is and in how interpretively rich or cognitively mediated it is. So, for instance, the intensity of a light’s brightness relative to the ambient lighting in a room is likely to be measured against a background that is both highly local and directly physical; while the intensity of a pigment’s tonal saturation in a painting might be measured not just relative to the rest of the picture in which it occurs, but also against the agent’s assumptions about the saturation levels typical of other paintings within a broader genre, or from other historical periods. The total prominence of a given feature, then is a function of both diagnosticity and intensity, with these interacting in potentially complex ways given the cognitive mediation involved in intensity.
The second dimension along which features may matter, in addition to prominence, is \textit{centrality}: how richly connected a given feature $f$ is to other features in the characterization, in terms of explaining, motivating, and justifying them, or of being explained, motivated or justified by them in turn. Some such connections among features may be conceptual, roughly in the sense of being inferences that a competent thinker would find compelling (Peacocke 1992). However, conceptual status is neither necessary nor sufficient for a connection to be part of a frame or its resulting characterization; many robustly conceptual inferences are too obvious and general to be relevant for explaining why that particular target subject is as it is. In ordinary cognition, we intuitively connect features in various ways, including on emotional, normative, and aesthetic grounds. But the paradigmatic form of explanatory connection is causal. A good measure of causal centrality is \textit{mutability}: how much the agent’s overall thinking about the subject would alter or need to be revised if the feature $f$ were removed (Murphy and Medin 1985, Thagard 1989, Sloman, Love and Ahn 1998). At least in a scientific context, such a psychological criterion fits smoothly with an analysis of causal explanation at least partly in terms of ‘difference makers’ (Woodward 2003, Strevens 2008): roughly, those features the changing of which would alter the event, individual, or kind in ways relevant to the operative taxonomy. Even here, though, it is worth noting that empirical work suggests that agents frequently impute causes where they perceive emotional or normative connections (e.g. Tiedens and Linton 2001, Lerner et al 2003, Small et al 2006). Given this, once again we should not be surprised if scientists sometimes impute causal structures on the basis of affective and normative intuitions.

Prominence and centrality are structurally distinct senses in which a feature can intuitively matter. Just as Barak Obama’s ears, or Donald Trump’s hair, may be highly prominent without being, or being represented as being, at all central, so too might a species of fox or horse tend to exhibit patches of highly noticeable white fur for no interesting reason, just because of a random mutation within a limited gene pool. But the two dimensions are not entirely disconnected. In particular, if a feature $f$’s intensity departs markedly from the contextually determined baseline, this fact calls out for some sort of explanation. Sometimes we dismiss such departures as mere anomalies; and sometimes such dismissal is justified, both in science and in ordinary life. But more often, we seek or posit an explanation grounded in the subject’s other features. Thus, some people intuitively connect Obama’s protruding ears with his Spock-like nerdiness, or Trump’s swooping hair with his grandiosity. More seriously, depigmentation has been correlated with hormonal and neurochemical changes associated with docility (Belyaev 1978, Trut 1999). The desire to explain prominent but apparently non-central outliers may produce a novel characterization in which $f$ itself becomes more central, or in which some other feature $g$, which itself explains $f$, is promoted in centrality.
These two interacting dimensions of ‘mattering’ generate a complex, intuitive organizational structure. However, most ordinary characterizations are only loosely organized: various features are distinguished in prominence and connected in centrality, but those distinctions and connections are inchoate, jumbled, and – as attested by the many experiments on priming – highly malleable (Camp 2015). By contrast, a frame constitutes a more unified interpretive principle which, when applied, organizes the characterization of its subject into a more coherent, contextually stable whole. That interpretive principle may be explicit in, or follow immediately from, the semantic meaning of the frame itself, for instance as when Bayes’ Theorem – the very principle articulated in the equation – is deployed as a tool to frame human decision-making. More typically, though, the operative interpretive principle is itself a complex, intuitive and partially inchoate characterization rather than a particular, specific proposition; in that case, the associated characterization itself brings along, and imposes, a complex structure on its target domain. Further, in some of these cases, the association between the framing device – that is, the sentence, picture, or diagram that functions as the frame – and the relevant characterization is conventional, as when a bigot uses a slur to evoke a characterization of members of a racial group (Camp 2013). In many other cases, though, most obviously with metaphor, the association between the intended framing characterization and the vehicle that serves as the framing device is merely pragmatic. Finally, in addition to variation in the conventionality of the association between framing vehicle and framing proposition or characterization, and in the richness and coherence of that framing characterization, the agent may also have a more or less coherent, rich, and cross-contextually stable antecedent characterization of the target subject itself. All of these differences will make a significant difference for the kind of cognitive effort required, and the kind of cognitive effect produced, by using the frame to guide one’s intuitive thinking about the subject.

So far, I’ve been talking about characterizations’ organizational structures, and the imposition of a new, more coherent structure. But what does it mean to say that a structure is implemented in or imposed on a characterization? The crucial insight that I take to be implicit in the quote from Black above, and in the near-ubiquitous talk of framing in general and metaphor in particular as a form of ‘seeing as’ or ‘perspective-taking’, is that a frame doesn’t merely represent its organizational structure. Rather, it implements it within cognition, so that more prominent features really do stick out in one’s attention and more central features are intuitively associated with others, by appearing to motivate or explain them. In this sense, frames provide cognitive Gestalts, much as the concepts ‘old lady’ and ‘young lady’ provide perceptual Gestalts for Figure 1.
This means that both characterizations and frames are in an important sense non-propositional. In principle, with sufficient reflection and effort, we might be able to explicitly articulate the complete set of features we intuitively associate with a given subject. Likewise, in principle, with even more reflection and effort, we might spell out the structure in which we intuitively arrange those features, weighting them by relative prominence and articulating their explanatory connections. However, it is neither necessary nor sufficient for employing a given frame that one endorse – or even explicitly entertain – the propositions that specify those constituent elements and relational structure. Instead, what matters is that one ‘get’ it, so that it actually structures one’s intuitive thinking about the subject in the relevant way. Moreover, ‘getting’ a frame is partly but not entirely under voluntary control. Sometimes, as with slurs, insinuations, and stereotype threat, frames may impose themselves when we would rather resist. Conversely, we may endorse a frame’s cognitive utility but be unable, at least as yet, to deploy it intuitively for ourselves. Indeed, first encounters with new scientific frames, such as Feynman diagrams, are frequently laborious, even when their primary advantage is facilitating an intuitive grasp of the topic. In cases where we want to but can’t intuitively ‘get’ a frame, any finite bit of new information about how to apply it in a particular case – for instance, being told that the young lady’s necklace in Figure 1 is the old lady’s mouth – may help it to ‘click’; but no one such bit is guaranteed to succeed.

In virtue of its intuitive Gestalt function, applying a frame is thus importantly a matter of imagination, but primarily in the synthetic sense most familiar from Kant, of uniting a manifold of disparate elements into a coherent whole. It is distinct from the sort of imagination typically discussed by philosophers interested in make-believe or pretense (e.g. Walton 1988, Currie and Ravenscroft 2003, Friend 2008). In particular, where make-believe is a matter of conjuring contents, in either an experiential or a propositional mode, which are known not to be actual, playing with a frame involves trying on a new mode of interpretation for a body of assumptions that are taken to be fixed (Camp 2009): as Wittgenstein says of Jastrow’s duck-rabbit figure, “I see that it has not changed, and yet I see it differently” (1953, 193). Often, the effects of this difference in the application of a new frame for how one sees or cognizes a subject extends beyond a realignment of the structure in which the basic elements are embedded, to produce alterations in the significance of those basic features themselves. So, for instance, in ordinary life someone’s teasing remarks may be interpreted as bumbling attempts at intimacy
or malicious bullying, depending on whether we take their producer to be more insecure or narcissistic.
Likewise, in psychology, an episode of lethargy might be interpreted as a sign of emotional paralysis or of
a low cortisol level, depending on whether we take depression to be a form of repressed, self-directed
anger or a biochemical imbalance.

§2: Metaphors and Other Framing Devices

So far, I've been deploying Black’s leading metaphor for metaphor – of an etched, smoked
glass – to explicate the idea of frames in general. Theorists who draw attention to the selective,
interpretive, and imaginative aspects of scientific theorizing sometimes treat all frames as equivalent.
Thus, Mary Hesse appears to treat models, narratives, fictions, analogies, and metaphors as fundamentally
equivalent and essential to scientific theory when she writes that “Scientific theories are models or
narratives, initially freely imagined stories about the natural world, within a particular set of categories
and presuppositions which depend on a relation of analogy with the real world as revealed by our
perceptions” (1993, 51; emphasis in original). While I share Hesse’s focus on the role of imagination and
presupposition in scientific theorizing, and while I agree that models, fictions, metaphors, and analogies
all employ imagination and presupposition to frame their subjects, I reject the assumption that all
scientific theorizing inherently involves modeling or framing in a substantive sense of the term. More
importantly, I think there are important differences among these species of frame. In this section I
identify some the key differences, and argue that these differences make a difference for how different
types of frame guide thought in everyday cognition and communication. In §3, I apply these distinctions
to a variety of examples of scientific models.

2.1 External and Internal Frames

All frames provide overarching principles of interpretation for their target subjects. A crucial
feature of metaphors as opposed to other frames is that they frame their subjects in terms of something
else (Camp 2003, 2006, 2008). Broadly, I advocate a story roughly along the lines of Black’s
‘interactionism’, which has important affinities to Dedre Gentner’s ‘structure-mapping’ account of
analogy and metaphor (e.g. Gentner and Markman 1993). On my view, a metaphor is a representation
which triggers initial characterizations of both a subject, A, and a frame, F: to take the canonical example,
the sentence ‘Juliet is the sun’ triggers characterizations of Juliet and the sun. Co-extensive expressions
(e.g. ‘sweat’ and ‘perspire’) may be associated with distinct characterizations; and the same expression
may trigger at least somewhat different characterizations in distinct conversational contexts. The
metaphor works by taking the most prominent and central features in the characterization of F and
seeking matches to them within the characterization of A, for as long as cognitive and/or conversational
interest warrants effort. Matched features are raised in prominence and centrality, resulting in a restructured characterization of \( A \) (and also, to a lesser extent, of \( F \)). In certain circumstances, when it would be plausible for \( A \) to possess a feature \( f \) that would be matched to a prominent and central \( F \)-feature, but where no \( f \)-like feature is currently included in the \( A \)-characterization, that feature \( f \) may be introduced into the characterization of \( A \). When a metaphor is used assertorically, the speaker claims that \( A \) possesses those features that are most tightly matched to the most prominent and/or central features of \( F \).

Not all frames work by matching between distinct characterizations in this way. At the broadest level, I think we need to distinguish both metaphor and analogy, as ‘external’ frames, from ‘internal’ frames, where the latter directly attribute a feature \( f \) to the subject \( A \), raise that very feature to prominence, and present it as maximally central for thinking about \( A \). The most straightforward type of internal frame is the ‘telling detail’ (Camp 2008), as vividly exemplified by certain types of insinuation. So, for instance, a speaker who utters ‘Obama’s middle name is ‘Hussein’’ overtly asserts a fact that is itself undeniable, and thereby implicates that Barak Obama instantiates a cloud of additional, much more controversial features that are implicitly associated with people named ‘Hussein’. In effect, the speaker presents a particular fact – Obama’s middle name – as symbolic or exemplary of a more general reality: who Obama is. The name (and further, its being a middle name, typically unmentioned) is purportedly a key to unlocking what truly matters about Obama, by highlighting known but otherwise unnoticed features and suggesting other, more sinister ones. Many insinuations, like this one, are insidious. But other invocations of ‘telling details’, both in everyday discourse and in science, are less underhanded: they explicitly present the mentioned feature \( f \) as central to \( A \), and may even spell out its most important implications. Thus, a primatologist might utter ‘Trump is a primate’, and go on to spell out just how Trump’s behavior can be explained and predicted by an analysis in terms of notable, relevant, and causally influential properties of primates, including especially social dominance (Camp 2008).

Although both metaphors and telling details provide interpretive frames, they do so in quite different ways. For one thing, metaphors don’t typically offer up a single feature as unlocking an entire characterization; instead, most metaphors map many subsidiary features from the framing characterization onto the subject. Second and more importantly, metaphors are unlike telling details in that they operate ‘from the outside’, in the sense that there remains a felt gap or contrast between frame and subject. Where telling details are interpretive keys inserted into the subject characterization, metaphors are colored lenses – sometimes kaleidoscopes – through which we view the subject from elsewhere.

Thus, to stick with the canonical example: Romeo doesn’t claim that Juliet is the sun, in the way that the speaker above claims that Obama is named ‘Hussein’. Instead, Romeo invites us to attend to and/or introduce features of Juliet that are like important features of the sun in relevant respects. In particular, the most prominent feature in our characterization of the sun – and the central feature from
which the metaphor unfolds – is its intense luminosity. But Romeo doesn’t claim that Juliet actually glows; rather, as his subsequent paraphrase spells out, the sun’s luminosity is matched to the quite distinct feature of Juliet’s (purported) beauty.

This match between the sun’s luminosity and Juliet’s beauty is *indirect* in the sense that features which are qualitatively distinct are presented as sharing relevant higher-order properties that are embedded in a common structure. In particular, both the sun’s luminosity and Juliet’s beauty are highly intense: they are instantiated to a degree that far exceeds the baseline exemplified by other individuals of the relevant class, and that makes them qualitatively different from other individuals in that class. However, the scale of intensity, the kind of intensity, and the individuals against which that intensity is measured are quite different in each case: the sun is brighter than the moon, Venus, or Saturn; while Juliet is (purportedly) more beautiful than Rosalind or any other girl of Verona. The sun’s luminosity and Juliet’s beauty share other important features as well: both are natural, and a source of energy and life; and both produce a feeling of warmth. Again, however, these common higher-level features are implemented in different ways within each subject.

### 2.2: Metaphor and Fiction

Thus, an internal frame operates directly and ‘from the inside’ of the subject characterization, by introducing and highlighting shared features; while an external frame operates indirectly and ‘from the outside’, by matching distinct features. So far, this might just seem like a new label for the difference between being literally true and literally false: when literal truth falls away, at most indirect or non-literal truth remains. Against this, I think at least some frames that are known to be literally false are nonetheless still internal, because they function imaginatively *as if* they were true. In particular, I take it that ‘just so’ stories are clearly fictional, insofar as they ascribe properties that the target subject is taken not to actually possess; but they function imaginatively much like ‘telling facts’, and importantly differently from metaphor (Camp 2009). Thus, someone might say that Donald Trump acts *as if* he was denied admission to Harvard as an undergraduate and has been trying to compensate ever since, while explicitly acknowledging that this *it not true*.\(^2\) Intuitively, such a speaker invites their hearer to pretend that Trump, in all his specificity – having been raised in Queens, with a real estate mogul father, etc. – really does possess that very feature – of having been denied admission to Harvard – and to treat that very feature itself as an imaginative key that unlocks what really does matter about him. That is, the hearer is supposed to pretend that the fictional feature \(f\) is actually instantiated and is explanatorily central, and to restructure their overall characterization of the subject \(A\) in this light, by introducing and elevating

other features in the $F$-characterization which $A$ really would possess if it did instantiate $f$. Once this is done, they drop the pretended ascription of $f$, leaving the subject characterization as close as possible to what it would be if $A$ really were $F$.

The contrast between the sort of imaginative activity involved in internal fictional frames and in metaphors is clearest in cases where a single sentence, like ‘Jane is a nurse,’ can plausibly be deployed as either species of frame. On the one hand, employing the sentence as a just-so story involves pretending that Jane really is a nurse. Here, what we might call the ‘direction of imaginative fit’ is from the actual reality to an imagined possibility (Levin 1988). That is, the interpreter starts with actual-Jane and uses her as a imaginative prop for the fiction. This involves transforming Jane imaginatively in two ways: first, by adding features that nurses do actually prominently possess, such as listening to various people’s symptoms, monitoring vital signs, administering medicine, perhaps being on call at inconvenient times, answering to imperious bosses, and juggling many patients; and second, by downplaying features of actual Jane that conflict with these prominent and central nurse-features, such as her actual incompetence with machines or the fact that she works regular business hours. A speaker might offer ‘Jane is a nurse’ as a just-so story in order to elucidate first-order respects in which Jane really does function basically as a nurse, even though she doesn’t have a BSN or RN.

On the other hand, a speaker might employ the same sentence as a metaphor. In that case, interpretation begins with a characterization of nurses and seeks to identify respects in which Jane, as she already currently actually is, is nurse-like. Rather than attributing actual nurse-features to pretend-Jane, the interpreter of a metaphor construes actual-Jane in a nurse-like way. As with Juliet, this imaginative activity focuses attention on actual features of Jane’s which nurses don’t generally actually possess, but which share higher-order similarities with prominent and central features in the stereotypical characterization of nurses, such as consistently lending a sympathetic ear, but for friends rather than assigned patients; checking on those friends’ emotional and psychological well-being, rather than on their physical symptoms and statistics; or nudging them toward avenues of emotional and psychological improvement, rather than delivering pills and injections. (Anyone who has spent a sustained period of time in a hospital will realize that this stereotype, like any other, is not universally applicable.)

In a case of sheer escapist fiction, an imaginative ‘prop’ like Jane is merely a springboard for make-believe. Some fictions, such as just-so stories, are “prop-oriented” (Walton 1990): we engage in the pretense with the aim of learning something more about the prop itself. What we aim to learn might be something about its counterfactual possibilities: about how it could be even though it isn’t. Or it might be something about what it’s actually already like, which makes it apt to serve as a prop in this particular game of make-believe. In focusing imaginative attention on their props, just-so stories are importantly like metaphors; as a result, both can serve as effective, informative frames for investigating and informing
others about their subjects. Partly for this reason, Kendall Walton (1993) argues that metaphors are invitations to engage in prop-oriented make-believe by pretending that the subject possesses the feature explicitly mentioned in the metaphorical sentence (see also Hills 1997 and Yablo 2001 for elaborations of this view).

I agree that the kinds of imagination overlap, and that many utterances invite a mixture of both modes of interpretation. However, I think there is still an important difference between the two tropes and the two sorts of imaginative activity, at least in paradigmatic cases. With a just-so story, we temporarily transform the prop A, by directly imbuing it imaginatively with features that are part of our characterization of Fs; only then do we consider what this reveals about A as it actually is. By contrast, with metaphor we hold our understanding of how A actually is as fixed as possible. Both frames are indirect, in one sense: we imaginatively step away from our actual assumptions about A. And both are guided by our intuitive characterizations about A and Fs. But with just-so stories, the indirectness arises via a temporary metaphysical transformation in imagination into A’s being F. By contrast, with metaphor the indirectness is generated at the interpretive level, and involves matching features of A and Fs that are merely similar to one another. Because they differ in direction and directness in this way, the two types of frames often end up highlighting and introducing different features within the ultimate characterizations of their subjects (Camp 2009).

That said, as I said, I do also acknowledge that in practice many cases of framing mix the two modes of imagination, or fall somewhere in between. For instance, when Wordsworth describes daffodils as ‘dancing’ in “I Wandered Lonely as a Cloud,” it initially appears to be an unexceptionable metaphor attributing regular motion induced by a breeze. But in the context of the entire poem, and of Wordsworth’s Romantic, pantheist philosophy, the metaphor becomes increasingly literal, more like a surreal fiction. This interpretive shift arises because, and to the extent that, the reader doesn’t merely select a few features of dancers and match them to merely analogous but distinct features of daffodils, but imaginatively transforms daffodils, as fully as possible, into agents who actually, literally dance – and Wordsworth, by contrast, into an empty drifting cloud (Camp 2009).

2.3: Metaphor and Analogy

In drawing the contrast between ‘external’ frames and ‘internal’ ones, I have distinguished both metaphors and analogies, as external frames, from telling details and just-so stories, as internal ones. In particular, I have emphasized that metaphors are indirect in the sense of relying on abstract similarities between frame and subject, so that distinct lower-level features share common higher-order properties and are often embedded within common higher-order structures. In this respect, my view is closely akin to Dedre Gentner’s “structure-mapping” theory of analogy. At the same time, there are also important
differences between metaphor and analogy. Specifically (as Gentner herself points out), metaphors are significantly less constrained than analogies, in at least two ways.

First, metaphors are less constrained in not being confined to matching identical higher-order, purely ‘structural’ features: they also frequently employ qualitative matches between base-level features. Thus, Romeo’s metaphor suggests that being near Juliet produces a physical feeling in him that is qualitatively, and not merely structurally, similar to the glow produced by the sun on a warm spring day. Likewise, the metaphor of Jane as nurse highlights base-level features of Jane, such as being compassionate, which nurses stereotypically actually possess. Briefly, we can classify the types of matches that metaphors rely on into four basic categories (Camp 2003). Some matching A-features are qualitatively identical to the matched F-feature: my beloved’s hair really is the color of ebony, or copper. Others differ in degree along a common qualitative dimension: an admired colleague is like Einstein in being really smart, but not that smart. A third class of matches, like that between the sun’s luminosity and Juliet’s beauty, are analogical, in sharing precise higher-order properties. Finally, some matches may themselves be metaphorical, albeit in a way that is less novel and robust than the overall framing metaphor itself: for instance, ‘Sally is a block of ice’ communicates a temperamental attitude, of being unemotional, that relies on a ‘deep’ conceptual association between temperament and temperature (Searle 1979, 78), which is cultural and/or biological experience (Lakoff and Johnson 1980).

Second, metaphors are less constrained in being more permissive in the tightness of the structure that must be preserved in the mapping from framing to subject characterization. In analogy, potential matches that are isolated by not being embedded within a more complex structure tend to be ignored, even if they are topically relevant (Gentner and Jeziorski 1993); by contrast, metaphors often happily permit isolated matches. Analogies also require consistency in mapping: the operative structure in the frame must be replicated within the subject for the analogy to be considered sound; and known failures of match within relevant aspects of the structure compromise the analogy’s plausibility. By contrast, metaphors are often quite unsystematic. Thus, Othello’s description of Desdemona as being “false as water” suggests myriad distinct respects in which Desdemona is deceptive: formless and unstable; running whichever way is easiest; reflecting whatever is around her; showing things within as different than they really are (as water does a bent stick); seemingly clear but potentially poisonous. These various matches mount up to a richly condemnatory portrait, but they don’t align neatly with one another, nor are they connected within any clear structure, causal or otherwise. Despite the lack of systematicity here, the metaphor is still an effective frame, in virtue of the rich range of resulting matches, and their affective and imagistic contents.

Because metaphors are more permissive than analogies, their interpretation also tends to be more imaginatively intuitive and holistic. Rather than puzzling out a precise, consistent formal mapping
between complex articulate structures, we feel our way through tacit clusters of matches involving largely inchoate features at a variety of levels, drawing on images and attitudes, and coloring and connecting those features – along with other, unmatched features that intuitively ‘fit’ with them – often without any reflective awareness of how we do so. Individual matches that are especially relevant for conversational or investigative purposes leap out at us and motivate intuitively related matches, even if these are not connected to or even logically consistent with the initial match; and clusters of such matches reconfigure both subject and frame to motivate further matches, in a snowball effect that can overwrite marked antecedent differences between the two characterizations.

§3. Metaphors and Other Frames in Scientific Inquiry

In §1, I described framing devices in general. In §2, I distinguished metaphor from three of its close cousins – telling details, just-so stories, and analogies – in terms of the direction, directness, level, and systematicity of imaginative fit between frame and subject. How do these differences play out in a scientific context? More specifically, what implications do they have for discussions of models and modeling in science? Although use of the term ‘model’ is both various and contentious, I think we can illuminate the utility and effects of at least many models by treating them as frames in the sense described in §1: as representational vehicles that embody interpretive principles which guide overall thinking about the target system to which they apply, by determining both what (purportedly) matters about that subject relative to a presupposed taxonomy, and also how those features that do matter are (purportedly) connected within an explanatory structure. Beyond this, our tour through various species of frame in §2 puts us in a position to identify important sources of variation among scientific models, while also illuminating their functional commonalities. In this section, I take a tour through some important types of frames in science, with particular attention to the sort of gap they assume between representation and reality, and to their different ways of bridging that gap to illuminate the target system.

3.1: Telling Details and Telling Instances

Many scientific theories employ what are in effect telling details: they explain a complex phenomenon by treating a single feature, which is itself relatively uncontroversially true and also associated with a rich set of assumptions, as being maximally explanatorily central. For instance, Longino and Doell (1983) contrast androcentric and gynocentric theories of tool use in hunter-gatherer societies within anthropology. Both theories agree that men hunted and women gathered, and both invoke tool use to explain the development of cognitive characteristics like flexible intelligence and instrumental reasoning. But the two theories disagree at a higher level, about which of these facts matter, and which data are exemplary of more general patterns that in turn reveal underlying causal structure.
Androcentric theories focus on hunting behavior and the relative efficacy of stone tools over sticks; while gynecentric focus on the nutritional stresses of pregnancy and lactation, and on the basic utility of sticks and reeds for digging, carrying and food preparation. These different frames weigh additional data differently, generate different chronologies and causal histories, and implicitly (and sometimes explicitly) suggest different predictions about and affective and normative responses to contemporary sex, tool use, and intelligence. Insofar as the primary locus of disagreement lies at this higher-order, interpretive level, it is difficult to adjudicate between the two theories directly, at the level of lower-level, demonstrable facts; each theory has its own way of taxonomizing and explaining any given bit of information, and can dismiss distinct isolated chunks of (putative) data as anomalies or as true but marginal.

Like the telling detail in everyday life, the ‘telling fact’ as scientific frame takes a feature $F$ that is uncontroversially assumed to be instantiated by a subject $A$ and treats it as maximally prominent and central within one’s intuitive thinking about $A$, relying on an assumed background characterization of $F$ and $F$-ish things. A closely related but distinct class of internal frame focuses directly on a single or limited class of instances – a vial of water, say, or a population of mice, or a patch of forest – and treats that instance, $a$, as exemplary of a more general kind $F$. Catherine Elgin aptly calls such samples “telling instances,” and points out that they serve many of the functions I have identified for frames: as she says, the sample $a$ “exemplifies, highlights, displays or conveys the features or properties it is a sample of” (2006, 12), in a richly context-sensitive way; as a result, it functions as “a symbol that refers to some of the properties it instantiates” (2006, 13).

Both of these types of ‘telling’ frames focus on a fact or feature that the target subject is presumed to actually possess. The difference between them is at least in part a matter of the level and direction of interpretive attention. First, the ‘telling fact’ operates at a theoretical level, by re-structuring the overall characterization of the target subject (say, the evolution of tool use) in terms of a characterization of a particular fact about that subject (say, that women used sticks to dig for roots). The frame’s main interpretive work lies in teasing out the theoretical consequences of taking this fact to be central for thinking about this subject: as a key that unlocks what matters about $A$. By contrast, the ‘telling instance’ or sample is itself concrete, and investigation involves probing it directly, in concrete ways.

Second, the two types of ‘telling’ frame differ in the direction of interpretive attention. In the case of the ‘telling fact’, the overall target subject $A$ – say, tool use – is framed by a characterization of a particular fact $f$. This involves making $f$ itself prominent and central within the characterization of $A$; and this in turn introduces or elevates further features $f_1, f_2, f_3, \ldots$ that play a central and prominent role within the $F$-characterization, possibly thereby suggesting causal connections between those $F$-features and further, non-$F$ features within $A$. By contrast, in the case of a ‘telling instance’, the primary locus of
investigative attention is the subject \( a \) itself. The sample \( a \) matters itself matters only as an instance of the general kind \( F \); assumptions and questions about \( F \)-ness fix an initial characterization of \( a \), selecting only some, taxonomically relevant features as being exemplified. And the ultimate goal is to ‘read back’ any newly discovered, relevant features and causal structures from \( a \) to other instances of \( F \). However, the investigation proceeds by probing \( a \) itself.

3.2: Abstraction and Idealization

Both ‘telling facts’ and ‘telling instances’ are intuitively true, in the sense that the predicate \( F \) cited in the framing sentence does indeed apply to the subject \( A \). Rejecting this basic intuition, some theorists, like Hesse (1993) and Elgin (2006), take the selectivity inherent in all classification, and the high degree of abstraction involved in modeling in particular, to render all theories and models literally false – more or less pragmatically efficacious fictions. I agree that selection and abstraction play a pervasive role both in cognition generally and in science specifically. Indeed, they are plausibly conditions on the very possibility of conceptual thought: applying a concept is a matter of classifying multiple distinct entities together as alike in some selected respect, or the same entity as recurring on multiple occasions, which thereby involves ignoring or abstracting away from other respects in which those entities or occasions differ from one another (Camp 2015). Further, we do regularly, and rightly, criticize representations (and representers) for neglecting relevant features. Some such criticisms accuse a representation of ignoring features that are diagnostic relative to its own presupposed taxonomy; while others challenge the taxonomy itself, either for embodying unwarranted assumptions about what kinds of features tend to cluster together or be causally efficacious, or for serving ethically or practically dubious ends. However, I do not think that representational silence, in the form of either selectivity or abstraction, constitutes falsity. A representation can only be false if it positively represents a state of affairs as obtaining which does not obtain.\(^3\) Moreover, because assessment for falsity can only take place against the background of a presupposed taxonomy, the very assumption of a taxonomy cannot itself be grounds for falsity, only for inappropriateness of some other variety.

Insofar as abstraction does not itself introduce falsity, then, it differs importantly from idealization. Both abstraction and idealization involve “imagining away” known facts that are assumed to be irrelevant (Godfrey-Smith 2009), either temporarily, in the service of practical tractability, or permanently, as a way to isolate key causal factors (Elliott-Graves and Weisberg 2014). But where abstraction engages in “mere omission” (Thomson-Jones 2005), remaining silent about features that are

---

\(^3\) Speakers are especially likely to exploit, and insist on, the difference between misrepresentation and non-representation in strategic conversational contexts (Camp 2017). Assessment for falsity becomes more complex in the context of extended conversations, where representations are embedded within structures of presupposition and relevance.
known to be possessed, idealization imagines features which are known to have one value to have a
different, more tractable value, as when the amount of friction between an inclined plane and a rolling
ball is imagined to be zero, or the number of possible mates in a population is imagined to be infinite.
When distinct features are causally interdependent, idealizing in one respect will affect the values of
distinct but related features. Because these interdependent effects may not be transparent to the agents
doing the idealizing, it is especially important to actively probe for unintentional, unrecognized falsities
that might in fact matter relative to the operative taxonomy and explanatory purposes.

The contrast between abstraction and idealization emerges especially clearly in the case of telling
facts and telling instances. As Elgin emphasizes, treating a ‘telling instance’ a as a sample of F employs
abstraction in an inevitable and pervasive way: only a limited subset of a’s features warrant investigation
and are eventually ‘read back’ into the characterization of Fs in general. But the very concreteness of
‘telling instances’ means that their use as models combines uneasily, if at all, with idealization. Because
idealization imaginatively constructs an entity that differs from the actual target subject, rather than
simply filtering out irrelevant features, it inevitably shifts attention away from probing the sample itself.
By contrast, the use of ‘telling facts’ as frames is fully compatible with idealization: for instance, both
‘androcentric’ and ‘gynocentric’ theories of the evolution of tool use might acknowledge that gender roles
were in fact more fluid than a strict segregation into male hunters and female gatherers suggests, while
still holding on to sharply contrasting models of the evolution of tool use based on starkly differentiated
‘male’ and ‘female’ roles. Likewise, each might invoke highly idealized “agent-based models” which
compute the long-term dynamic effects of repeated interactions between individuals defined in terms of
just a few traits, which are themselves derived from assumed gender roles.

3.3: Fact and Fiction

If idealization, unlike abstraction, introduces a form of known falsity, are all idealizations
therefore fictions? Here too, I think we should resist assimilation. The falsification introduced by
idealization is still like abstraction in functioning to ignore complexities of the target subject that are at
least purportedly irrelevant. By contrast, fictions paradigmatically introduce features that are known not
to apply: the subject is transformed in imagination, in ways that are known to be relevant. While the line
between merely ‘smoothing out’ irrelevant complexities and actively introducing alternative properties is
not sharp, fictionalization tends to involve both a more substantive qualitative departure from how the
subject is assumed to actually be. It also involves greater attention to the fictionalized subject in its own
right, so that implications for the actual target system are derived as a second step of comparison between
fiction and fact.
Maxwell’s Demon provides an illustrative case of how such active fictionalization differs from mere idealization. The second law of thermodynamics, that entropy in a closed system never decreases, had often been interpreted as an absolute law grounded in the nature of ‘caloric’. As a counterexample to such an interpretation and in support of the molecular theory of heat, Maxwell (1871) suggested that “we conceive a being” whose perceptual faculties are “so sharpened that he can follow every molecule in its course,” but “whose attributes are still as essentially finite as our own.” If this being were stationed at a door which divided a vessel into two chambers, he could produce a difference in the temperature of the chambers “without expenditure of work,” just by opening and closing the door to allow swift molecules into one chamber and slow molecules into the other. From the coherence of this possibility, Maxwell concluded that the second law holds only at a statistical level – that is, “as long as we can deal with bodies only in mass, and have no power of perceiving or handling the separate molecules of which they are made up” (Maxwell 1871, 338-9). In short, by asking us to imagine a scenario that is obviously false, but in (purportedly) merely contingent respects – the demon is just like us, shrunk to a molecular scale – Maxwell claims to demonstrate that a perpetual ‘heat engine’ is merely extremely unlikely, not physically or metaphysically impossible.

Unlike paradigmatic cases of idealization as ignoring or “imagining away,” Maxwell’s thought experiment directs attention toward an overtly counterfactual situation. Much as with a just-so story, we are asked to imagine that the very situation described is true; and in both cases, the result is to draw our attention to other features which follow directly from that fictional proposition, and which purportedly are actually true. Assessing the fiction’s aptness as a frame is thus a matter of determining two things: first, what is true in the fiction: what unstated implications follow from the propositions that are literally and explicitly articulated by the fiction, given the fiction’s operative ‘principles of generation’ (Walton 1990); and second, whether the real world is indeed like the fiction in these unarticulated respects (Frigg 2010, 260). The question of just what the operative principles of generation are for fiction is a vexed one in general (Lewis 1978, Walton 1990, Byrne 1993); and different genres employ importantly different principles of generation. However, these issues are considerably less pressing in a scientific context: because the fiction is ultimately just a tool for highlighting propositions that are true in both fiction and reality, the primary issue is just whether there is a relevant but unacknowledged gap between fiction and reality. Whether that gap arises from inappropriately unrealistic principles of generation or from inappropriately concluding that the real world shares those unstated features with the fiction is effectively moot. So, for instance, subsequent discussion of Maxwell’s Demon has challenged Maxwell’s conclusion that the demon’s operating the door – or more importantly, his measuring individual molecules’ speed – does not itself constitute “expenditure of work,” with little attention to precisely how that conclusion is generated.
3.4: Metaphor and/versus Analogy

In effect, we have now seen that abstraction, idealization, and fictionalization involve successively greater departures from stating ‘the whole truth and nothing but the truth’ about the target subject. But all of these departures arise in the service of focusing attention on features in the model or framing representation that are also (purportedly) relevantly instantiated by the actual target. ‘External’ frames like metaphor and analogy take a further step, of ‘telling the truth but telling it slant’, as Emily Dickinson puts it. In these cases, as I argued in §2.2, we do not pretend, even temporarily, that the world really is as the representation literally describes. Instead, we leave frame and target distinct in imagination, and seek to identify respects in which they are alike, where these similarities may be both highly selective and indirect.

The history of competing models of atomic structure provides an illuminating case of the selective and indirect mapping employed by external frames, and their difference from fiction. A key problem for early atomic theory was how to reconcile the stability of atoms, which are neutrally charged, with the negative charge of their constituent electrons. Thomson’s (1904) “plum pudding” model of the (hydrogen) atom achieved this reconciliation by embedding those electrons within a uniform sphere of positive charge, much as the batter in a Christmas pudding contains raisins. In understanding Thomson’s model, we are not asked to even imagine that atoms are bowls of raisin-studded pudding, in the way Maxwell asks us to imagine a microscopic demon opening a tiny door. Rather, we are asked to posit, and treat as central, a sphere of positive electric charge that is like a pudding in the respect of functioning as a diffuse stabilizing medium.

Rutherford’s (1911) discovery of the existence of a small nucleus of intense positive charge falsified Thomson’s ‘diffuse’ model of positive charge, and provided an empirical basis for the alternative model of an atomic core. It thereby provided support for Nagaoka’s (1904) “Saturnian” metaphor of electrons as akin to the rings around Saturn, which Nagaoka had proposed on distinct theoretical grounds, based on the impenetrability of opposite charges. Bohr’s (1913) “solar” model then extended and refined Nagaoka’s Saturnian metaphor by suggesting that the negative electrons orbit the massive positive core, just as the planets in the solar system revolve around the sun; and that electrons are attracted to the nucleus by electrostatic forces, akin to the sun’s gravitational force. Bohr’s model is a theoretical improvement in part because it subsumes the disparate empirical results which supported the earlier models into a coherent larger model, and also because it suggests a casual mechanism by which those effects are produced. In particular, shifting to the solar model introduces an orbit as a discrete, stable path, where previous models were unable to explain either atomic stability or discreteness of energy levels.
In all of these models of the atom, the mappings from frame to target are highly selective, highly abstract, and closely focused on structure, in the manner characteristic of analogy (Gentner and Jeziorski 1993, 449). In particular, Bohr’s model identifies an identical higher-level relational feature, of an attractive force causing rotation, which is in turn instantiated by quite different lower-level features within each of the frame and target: where gravity causes the planets to orbit the sun, electrostatics causes electrons to orbit the nucleus. And it ignores myriad possible matches, like color and relative temperature, that are not implicated in this causal structure.

As we saw in §2.2, this selective focus on “common relational abstractions” (Gentner and Jeziorski 1993, 448) as opposed to lower-order common features differentiates both metaphor and analogy from fiction. A scientific fiction, as Elgin (2006, 16) says, “sheds light on the way the world actually is” by “exemplifying features that diverge (at most) negligibly from the phenomena it concerns.” In this respect, Elgin argues, fictions are like samples – indeed, because she assimilates abstraction and idealization to fictionalization, she argues that samples, such as paint chips, are fictions. While I want to reject Elgin’s assimilation, I agree that scientific fictions function to draw attention to features that really are exemplified both in the fiction and in the actual world, or that diverge negligibly. By contrast, metaphors and analogies shed light on the world by exemplifying common structures that diverge substantively and relevantly in the way they are implemented in frame and target.

The difference between fiction and metaphor or analogy becomes especially palpable if we contrast Maxwell’s original thought experiment with its subsequent deployment as a metaphor for other topics. For instance, Pierre Bourdieu argues that the (French) educational system functions as an entropy-reversing mechanism that maintains social structures of “difference and order, which would otherwise tend to be annihilated,” by sorting students at an individual level in terms of their possession of cultural capital (1998, 20). Bourdieu ignores Maxwell’s ultimate point: that the second law of thermodynamics does in fact hold at a global, statistical level because there actually is no demon. But his metaphor does identify a common structure for Maxwell’s imagined, fictional situation and actual schools: of an entropy-reversing and therefore ‘unnatural’ mechanism which produces global effects by sorting individuals. However, where Maxwell’s fiction directed attention at the target phenomenon itself – the trajectory of distribution of heat in a closed volume – and asked us to imagine something literal but counterfactual about it, Bourdieu applies that structure to a very different domain.

It is true that in analogy, and to a lesser degree metaphor, the frame and target are presented as possessing identical higher-level features: in Bohr’s model, an attractive rotation-causing force; in Bourdieu’s metaphor, a entropy-reversing mechanism for sorting individuals. In this sense, at a suitably high level of abstraction an analogical or metaphorical frame does impute features that are actually possessed by the framing subject to the target subject – just as a just-so story imputes features that are
possessed by the fictional subject to the target. The proponent of a unified account of scientific models as fictions might thus propose that the difference between metaphor and fiction is just a difference of the level at which common features are imputed, rather than a difference between pretending that a nonfactual feature really does apply in order to draw attention to and/or introduce other features that would directly follow from that feature, and identifying matches between merely similar features, as I have claimed. However, an analysis of metaphor and analogy purely in terms of higher-order common features distorts the representational import of analogy and metaphor, both in everyday discourse and in science. The claim made by a metaphor or analogy is not just that the target is like the frame in this common, highly abstract respect.

That is, scientific metaphors and analogies do not typically function like abstract models, like the Lotka-Volterra equations describing the effects of predator-prey dynamics on population distribution. Abstract models of this sort prescind from messy detail in order to focus attention entirely on high-level, general, structural features. By contrast, in metaphor and analogy, the shared high-level, structural features warrant attention only instrumentally, as a way to identify the more specific lower-level feature within the target that instantiates it. In a pedagogical context – for instance, when explaining electrical current via an analogy to the flow of water through a pipe – the speaker will explicitly identify, or will ask listeners to identify for themselves, the lower-level instantiating features of the target. But in a context of discovery, the investigators are typically ignorant of what those instantiating features are, and employ the possibility of a match between frame and target via the relevant higher-order feature as a principle to guide investigation. In either case, attention is directed at the basic-level features, using the structural match as a guide.

So far, I’ve been emphasizing ways in which both metaphor and analogy differ from fiction, arguing that they involve a qualitatively greater gap between representation and reality than fiction, idealization or abstraction, in virtue of relying on indirect matches between the target system and a frame which is understood as something else. But as we saw in §2.3, metaphor and analogy also differ from one another in important ways in everyday life. Gentner and Jeziorski (1993) argue that contemporary scientific practice valorizes analogies, like Bohr’s solar model, because they employ precise, consistent, systematic matches between complex, causally-connected systems of features, over metaphors. However, they also claim that this is distinctive to modern Western science. They show that alchemists up through the sixteenth century were much more promiscuous in their invocation of similarity, happily citing not just higher-order relational structures but also base-level qualitative similarities, such as the yellowness of both the sun and gold or the whiteness of the moon and silver, and invoking multiple disconnected or even incompatible matches. They argue that the birth of modern science is due in significant part to this shift from promiscuous similarity to higher-order structural matching. As a result, they effectively treat
metaphor in contemporary science as a poor cousin to analogy, as reflected in a pithy quote from George Polya (1954): “And remember, do not neglect vague analogies. But if you wish them respectable, try to clarify them.”

In emphasizing the way that both metaphor and analogy involve selective, structural, indirect mappings from frame to target, in contrast to scientific fictions which imagine known falsities to be true in order to highlight and discover directly entailed consequences that really are true, I have also emphasized the ways in which metaphor approximates to analogy. And Gentner and Jeziorski’s priority claim seems to be validated in other respects as well. First, metaphors in science, in contrast to literature, do often emphasize fewer, more consistent matches over richer, inconsistent ones; this is especially true in pedagogy and theoretical advocacy, which are the cases that Gentner and Jeziorski focus on almost exclusively. Further, it is widely agreed that the ultimate aim of science is to develop a precise, articulate understanding of objects, properties, and the relations between them, and that in order to accomplish this, we need to employ symbols whose interpretation is “univocal, determinate, and readily ascertained” (Elgin 2006, 17). Insofar as metaphors differ from analogies in relying on tacit, vague, and otherwise inarticulable intuitions of similarity, they are not fully representationally adequate as they stand.

More substantively, some of the most influential scientific metaphors of the twentieth and twenty-first centuries have aimed at identifying abstract, high-level properties, just as Gentner and Jeziorski predict. To take a pair of particularly apt examples, the computational model of mind and the code model of genetic potential both hypothesize key causal operations which are functionally analogous to the algorithmic execution of a computer program. One reason both metaphors have been so theoretically and empirically productive is that they encourage a focus on functional, structural relations while remaining relatively neutral about implementational mechanism, at least until the functional role and the lower-level physical phenomena are each understood better in their own terms. (Pylyshyn (1993, 551) describes such neutrality a case of the “principle of least commitment” or “principle of procrastination.”) And as we saw in §2.3, analogies are especially well-equipped to accomplish this sort of identification of novel structural properties, precisely because of their selectivity and abstractness.

What is right about Polya’s dictum is that science should, and does, aim to clarify metaphors. But metaphors often play a theoretically and empirically fruitful role in scientific inquiry precisely because they stand in need of clarification: because they are inchoate, intuitive, and only partly consistent. As we saw in §2.3, metaphors’ greater permissiveness in matching allows them to engage imagination in a richer, more intuitive, and flexible way. This means they can guide attention and suggest hypotheses in epistemic circumstances where a more precise structural analogy would be stymied. Early advocates of both a computational theory of mind, and of a code model of genetic potential and action, lacked clear, coherent characterizations of both their target systems and their framing subjects, since the notions of
computation and coding were themselves still nascent. Indeed, as Fox Keller (1995) argues, conceptual and empirical developments within computation and genetics where mutually supporting, with each serving as a frame for interpreting and investigating the other domain: at the same time as the metaphor of genes as self-replicating machines drove theoretical, empirical, and technological developments in molecular biology, so too did the metaphor of complex machines as organisms also orient important strands of research within systems analysis and cybernetics, which in turn reciprocally influenced theories of biological development and cellular coordination.

In effect, in each direction, the metaphor provided what Richard Boyd (1979/1993, 488) calls an “inductive open-endedness”: they guided research by pointing to an indeterminate range of possible matches which had not yet been fully articulated or investigated. Metaphors like mind as computer, genes as machines, and machines systems as organisms can play a “programmatic research-orienting” role (Boyd 1993, 489) not merely in spite of, but because they lack the “univocal, determinate, and readily ascertained” interpretations of paradigmatic scientific symbols: they guide research by pointing to an indeterminate but bounded range of possible matches. Gentner & Jeziorski’s emphasis on ‘respectable’ analogy in the explication and justification of contemporary scientific theories neglects the unruly but ineliminable role of imagination in scientific inquiry, by suggesting that the ‘proper’ scientific use of metaphor and analogy must implement a general, logical interpretive algorithm of structure-matching. We can grant, first, that the interpretation of both analogies and metaphors are ultimately computational or algorithmic at some level of tacit cognitive processing, and second, that the ultimate aim is to articulate, test, and justify the incipient mapping between framing and target characterizations. But an important source of the power of metaphor specifically as a framing device derives from the absence of a general interpretive procedure for achieving this.

Our discussion of framing devices in §1 allows us to make this point about interpretive indeterminacy in a somewhat more precise way. Both the constituent elements and organizational structure of characterizations are typically largely implicit and only partially subject to voluntary control. They are also highly dependent on context, with diagnosticity and centrality in particular depending on the agent’s interests and goals. As a result, different scientists are likely to bring different characterizations to the interpretive table, especially at the beginning of inquiry. Further, even given a fixed pair of characterizations, multiple plausible overall mappings between frame and target will nearly always be available at any given moment in inquiry, which trade off preferences for systematicity against directness in matching, and for identifying potential new features and explanatory connections in the subject domain against preserving features and connections which are already assumed to hold, in different ways. Beyond this, frames don’t just apply to what is known about their targets at a given moment: they provide open-ended tools for attending to and assimilating new information as it comes in,
and for generating hypotheses about as-yet undiscovered features and causal structures. Finally, in addition to all of these frame-internal factors contributing to interpretive indeterminacy, the actual application of any frame also depends in deep, important ways on external factors: on what alternative theories it is being compared to, and so what expressive and epistemic needs it distinctively addresses (Okruhlik 1994), as well as its interaction with current technological opportunities and limitations (Fox Keller 1995).

Perhaps the best way, then, to view the relationship between metaphor and analogy in much of contemporary scientific practice is to see metaphor as tracing a trajectory or “career” (Bowdle and Gentner 2005): as beginning with an intuitive, holistic, and open-ended, but also diffuse and relatively unarticulated mode of construing one subject in terms of another, where one or both domains may be only minimally understood; moving through a process of articulation, probing, and refinement of the characterizations of one or both domains and of the plausible matches between them; and ultimately settling into a more regimented, systematic, and selective analogical mapping. At that point, the analogy may remain as a useful pedagogical tool. Alternatively, the interpretation of the framing term may have morphed so as to become literarily applicable in a few restricted respects, as has arguably happened with ‘computation’ or with gravitational ‘wave’. And of course, the metaphor may ultimately be discarded. Perhaps, like the metaphor of evolution as climbing a ladder, it turns out to be misleading, because it directs attention toward features that are not as central as once anticipated, or imputes features that are not in fact possessed. Or perhaps although it was once genuinely fruitful, and continues to identify features which are both prominent and central, it has become too dominant and literalistic in its application, leading to the neglect of other important features; perhaps the metaphor of natural language as a logical calculus fits this description.

3. Models and Frames

Much current philosophical discussion about scientific models has focused on their ontological status – in particular, whether models are abstract structures or hypothetical, typically uninstantiated concrete entities – and in turn on whether the representational relation between model and target is one of direct instantiation or a more indirect one of similarity in relevant respects (Giere 1988, Godfrey-Smith 2006, 2009, Frigg 2010, Weisberg 2012). I have focused on the apparently distinct topic of frames. Although I can’t pretend to have surveyed, let alone explained, all the phenomena and functions of models and modeling in science, it does seem that models and frames display remarkably many of the same features, and tend to be used for many of the same epistemic purposes. One benefit of turning at least temporarily to an investigation of frames is that it integrates the use of models in science more smoothly into a broader theory of interpretation, and thereby into a theory of cognition and
communication, from which we can discern both commonalities and differences between the use of models and other interpretive strategies within science, and commonalities and differences between the practice and normative assessment of those strategies in science and in everyday cognition and communication.

Specifically, I have argued that frames are interpreted representational vehicles which provide an overarching principle for interpreting something else. All frames presuppose a taxonomy, which is necessarily selective and contrastive; all frames determine what matters about their subject, and how, along at least the two dimensions of prominence and centrality; and all frames are intuitive and non-propositional, in the sense of implementing rather than merely representing these interpretive structures.

However, within this general characterization, different species of frames work quite differently. Frames themselves can be more or less abstract, idealized, detailed, and affectively and experientially loaded. Some, like Bayes’ Theorem, express highly abstract structures that literally describe a few highly idealized features of the target domain; others, like vials of water or colonies of rats, constitute concrete exemplifications of their target subjects. Frames can also be more or less conventionally tied to their vehicles: some vehicles, like Bayes’ Theorem the Lotke-Volterra equations, constitute explicit semantic specifications of the relevant structures; but in many cases, like sex-based theories of the evolution of tool use or computational metaphors of mind, the connection is one of implicit, pragmatic association.

Whether the connection between the representational vehicle and framing principle is conventional, implementational, or merely associative, the interpreted representational vehicle generates a cognitive structure, which is then used to re-structure the cognitive structure of the target subject. The ensuing connection between frame and target can be more or less direct, instrumental, and systematic. Some frames, like sex-based theories of the evolution of tool use, assimilate the frame’s defining feature, and all of its subsidiary features, directly into the target subject. Others, like Maxwell’s Demon, also assimilate that feature directly but only temporarily and partially, in order to highlight or introduce subsidiary features that the target really would have if the framing feature were actually possessed. Finally, some frames, like Bohr’s solar model, export a selective structure from one domain to a distinct one, matching lower-level features in each domain indirectly, by way of coherent systems of relations; and others, like the computer model of gene reproduction, highlight, explain and restructure features of the target in a way that is also indirect, but, at least initially, much more inchoate, potentially inconsistent, and often imagistically-driven.

All of these forms of framing can naturally be described as employing models. But we miss both important commonalities and differences if we focus primarily on the representational entities that underwrite them. Attending to the practices and processes of modeling and framing affords a more perspicuous analysis (Godfrey-Smith 2006, Levy 2015). But a full understanding of those practices
requires attending to the cognitive structures and operations that make them natural and effective for agents with minds like ours. I have argued that although the various species of framing direct imaginative attention at different levels and bridge the gap between representation and reality in different ways, they all employ a synthetic, restructuring imagination to achieve a unified, open-ended, intuitive construal of their targets.
References


