forthcoming in *The Scientific Imagination,* ed. P. Godfrey-Smith and A. Levy (OUP). Penultimate version. To cite or circulate, please email me at elisabeth.camp@rutgers.edu.

Imaginative Frames for Scientific Inquiry: Metaphors, Telling Facts, and Just-So Stories[1]

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, led by the likes of Vico, Shelley, and Mary Hesse, takes metaphor to be distinctively equipped to achieve a mystical communion with nature – a mode of representation that unlocks the universe's secrets and even creates new worlds. Often, subscribers to this view take all language and thought to be ultimately metaphorical, or at least take metaphor to be the truest embodiment of the basic mechanisms by which reference, truth, and understanding are achieved. The dismissive camp, helmed by the likes of Hobbes, Locke, and Zenon Pylyshyn, rejects such representational and ontological profligacy, and instead treats metaphor as superstitiously positing occult, non-referring forces and entities. At best, metaphor is a decorative trope or a mechanism for inspiration; at worst, it spins bubbles of self-confirming pseudo-science.

This opposition appears especially stark given a positivistic conception of science as the logical subsumption of observation sentences under general theoretical laws. Few endorse this conception today. Since at least Quine (1951) and Kuhn (1962), philosophers have noted that scientists bring a host of only partially articulated theoretical, practical, and empirical assumptions to bear in investigating the world, and that distinct patterns of attention and explanation can motivate distinct interpretations of any given 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' – that are known to differ from the actual world in significant ways.

Both developments have had the salutary effects of dispelling a false picture of scientific theories as transparent descriptions embedded in purely logical structures, and of connecting our theoretical understanding of scientific investigation, representation, and justification more closely to 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 the general banner of 'models'. Further, some of these theorists, in their zeal to

[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, Mike Stuart, 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.

oppose a naively descriptivist realism, have sometimes concluded that all theories are mere fictions levied in the service of competing pragmatic interests. Thus, we seem to return full circle, to the claim that all representation is essentially figurative, but with fiction now occupying the preeminent role once accorded to metaphor.

In this chapter, I distinguish among a range of representational devices, which I call 'frames', all of which guide our overall interpretation of a subject by providing a *perspective*, or an intuitive principle for noticing, explaining, and responding to that subject. Frames play a theoretical role closely akin to that commonly ascribed to models. But where much of the discussion of models focuses on their ontological status and representational relation to reality, I focus on the cognitive structures and abilities that are generated by frames, and on the imaginative activities that exploit them. Further, where many theorists of modeling have aimed to explain models by positing a single common representational relation, I focus on distinct ways that scientific representations can fruitfully depart from representing 'the truth, the whole truth and nothing but the truth'. Specifically, where recent discussion of models draws inspiration from fiction, I focus on metaphor.

My aim here is primarily descriptive: I want to explain the shared features of frames that make them powerful interpretive tools; distinguish among various ways that they can work; and draw out similarities and differences between their application to everyday cognition and scientific inquiry. I believe the discussion of frames here also provides the resources for identifying central norms on frames' epistemic aptness, both in general and in particular cases. Further, I think that once we assess frames for epistemic aptness, we can justify a significant epistemic role for frames within scientific inquiry, and even at the putative end of inquiry. However, establishing these normative consequences remains a task for another occasion.

In §1, I use metaphor to introduce the broader family of perspectival 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 at work within scientific inquiry, and use them to illustrate important differences in the sorts of gaps that models can open up between representation and reality. I conclude in §4 by advocating a mild ecumenicalism about scientific models: although most models are deployed in support of importantly similar cognitive and epistemic functions, there is no single ontological status or representational relation common to all.

1. Frames, Perspectives, 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 the subsidiary subject (1954, 288).

I think that this passage expresses an insightful and basically correct view of metaphor. But it is unsatisfying as it stands, 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 further metaphors, not all of which are 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, including fictions, slurs, and telling details. I think this is an important positive insight to be gleaned from Black's remarks, rather than (just) a weakness. In this section, I spell out Black's talk of metaphors as 'organizing structures' in my own terms, as it applies to all these cases. In §2, I tackle the question of how to differentiate among them.

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 familiar 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 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 and unarticulated 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 on common intuitive assumptions in 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. Notable types include slurs, as in 'You know he's an *S*' (Camp 2013); 'telling details', as in '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). These tropes differ in their rhetorical operations and effects. But 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 – what I call a *perspective*. More specifically, perspectives perform three related functions: they determine what information an agent *notices* and remembers about the subject; they guide how the agent assimilates and *explains* new information within the context of their prior assumptions; and this in turn guides how the agent evaluates and *responds* to that information.

Perspectives are thus closely related to both frames and characterizations, and mediate between them. On the one hand, the function of frames is to express perspectives. Not all perspectives are so expressible; some are just too multivalent to be crystallized into a single slogan or image, or no one has yet happened or needed to do so. And when a frame does express a perspective, that perspective goes well beyond the representational content encoded by the framing vehicle: perspectives are intuitive principles for interpretation, rather than particular thoughts or contents in themselves. On the other hand, the function of perspectives is to generate characterizations, where a characterization is a particular intuitive structure of assumptions about a subject. But perspectives are open-ended, in two senses: they provide intuitive default principles for updating characterizations over time, as new information and experiences come in; and often they also generate characterizations of indefinitely many different specific subjects.

As devices for crystallizing perspectives in order to generate and regulate characterizations, frames are ubiquitous in ordinary life. We observe them at work in political discourse, in intimate interpersonal arguments, in informal commentaries on movies – anywhere that intuitive interpretation is at stake and potentially contested. Three features of frames, and the perspectives they express, 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. (For instance, slurs and thick terms like 'slut' are objectionable partly in virtue of focusing on race, sex, or other properties as purportedly 'deep', predictively 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. One obvious source of diversity among scientific frames, then, 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). Insofar as scientific investigation and theorizing attempts to eschew affect and to provide a general, systematic understanding of its subject, images and feelings tend to play – or at least are supposed to play – a comparatively smaller role in scientific thought than in literature, say, or political discourse. However, given that scientists do still interact with their target subjects in experientially rich, sustained, and highly normatively-laden ways, it is unsurprising that their characterizations of them often involve images, feelings and norms to a significant degree.

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 that subject, at least not in the rich sense at stake here. Rather, a representation, like a metaphor, slogan, or diagram, functions as a frame when an agent uses it to organize and regulate their overall intuitive thinking about the subject – that is, to characterize the subject. 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, as the expression of a coherent perspective, a frame purports to determine, for any property that might be ascribed to the subject, both *whether* and *how* it matters.

To understand the cognitive and epistemic implications that such an overarching perspective might have, we need to say more about how frames select and structure a set of disparate features into a

coherent whole. There are (at least) two distinct ways in which a feature can matter, or be more or less important, in an agent's characterization of a subject (Camp 2003, 2013, 2015). First, some of the features ascribed to a subject are more *prominent* than others, in the sense of being more initially noticeable and quicker to recall. I take prominence to be equivalent to what Tversky (1977) calls 'salience', which he in turn analyzes as a function of two factors, each of which is contextually relative in a different way. 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 that employ distinct taxonomies will 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 an agent counts as 'noise' – as the relevant background against which the current signal is measured – varies widely, both in how locally restricted and in how cognitively mediated it is. So, for instance, the perceptual intensity of a light's brightness relative to the ambient lighting in a room is fixed by a background that is both highly local and directly physical; while the intensity of a pigment's tonal saturation in a painting is likely to be assessed not just relative to the other colors in that particular picture, but also against the agent's assumptions about typical saturation levels in other paintings within that genre, or from other historical periods. The total prominence of a given feature in one's intuitive characterization of the subject is a function of both diagnosticity and intensity, where these interact with each other and with the larger context in complex ways.

Where prominence selects *which* features matter, the second dimension of significance, *centrality*, concerns *how* they matter. Intuitive characterizations don't just attribute piles of more or less prominent features; they connect them into explanatory networks, with more central features being more richly connected to other features. Some 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 feature to play a central role in an agent's intuitive characterization of the subject. On the one hand, many robustly conceptual inferences are too obvious and general to be relevant for explaining why a particular target subject is as it is. And on the other hand, we often intuitively connect features in ways that are contingent, sometimes highly so. In ordinary cognition, these connections are often emotional, ethical, and aesthetic. But typically, and especially in science, imputed explanatory connections are causal. A good measure of causal centrality is *mutability*: how much the agent's overall thinking about the subject would alter or need to be revised if

they no longer attributed a given feature *f* to the subject (Murphy and Medin 1985, Thagard 1989, Sloman, Love and Ahn 1998).[2]

Prominence and centrality are structurally distinct ways in which a feature can intuitively matter. Barak Obama's ears or Donald Trump's hair may be highly prominent in our thinking about them without being represented as at all central to who they are. So too, a species of fox or horse might 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, when a feature *f*'s intensity departs markedly from a contextually determined baseline, this fact intuitively 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 that is 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). Thus, 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. Conversely, when a feature *f* is highly central in our thinking, this also tends to make *f* diagnostically relevant, and may lead us to raise our intuitive assessment of its actual intensity or statistical frequency.

These two interacting dimensions of 'mattering' generate a complex, intuitive organizational structure which is exhibited by all characterizations. However, most ordinary characterizations are only loosely organized: different features have different weightings of prominence and are connected in various ways to more or fewer other features; but those differences and connections are inchoate, jumbled, and – as attested by the vast experimental literature on affective and cognitive priming – highly malleable (Camp 2015). By contrast, a frame constitutes a unified interpretive principle which, when applied, organizes the characterization of a particular subject into a more coherent, stable whole.

The relevant interpretive principle may be explicit in, or follow immediately from, the semantic meaning of the frame itself. A clear example of this might be the use of Bayes' Theorem – that is, the

[2] At least in a scientific context, such a psychological criterion fits smoothly with an analysis of causal explanation that invokes 'difference makers' (Woodward 2003, Strevens 2008): roughly, an agent treats *f* as causally important to a subject *A* if they treat *f* as making a difference to *A* in ways that matter given the presupposed taxonomy; and *f* is central to the agent's characterization to the extent that the agent takes its potential alteration to affect many features that matter. Even here, though, it is worth noting that empirical psychological results suggest that priming agents to make emotional or evaluative connections between features in a situation frequently leads them to also impute causal connections between those features (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, at least initially, on the basis of affective and normative intuitions.

very principle articulated in the equation – as an open-ended tool for explaining and predicting human decision-making. More typically, though, as with metaphors and slurs, the interpretive principle that is associated with a given frame is itself a complex, intuitive and partially inchoate characterization; thus, the metaphor 'Man is a wolf to man' frames social relations in terms of a complex 'system of commonplaces' intuitively associated with wolves. Further, for some frames, the association between the sentence, picture, or diagram which functions as a frame and the perspective it expresses is conventional, as when a bigot uses a slur to evoke a derogatory perspective on members of a racial group (Camp 2013); while in others, most obviously metaphor, the association between the framing vehicle and the characterization it is intended to evoke is merely pragmatic. Finally, in addition to variation in the interpretive richness of the perspectival principle evoked by the frame and in the conventionality of their association, the agent's antecedent characterization of the target subject may itself be more or less rich, coherent, stable, and/or explicitly articulated. 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 any one frame to guide one's intuitive thinking about any given subject.

So far, I have identified a type of organizational structure that is implemented by all characterizations, and claimed that as expressions of perspectives, frames offer interpretive principles for imposing a more coherent structure on characterizations. In particular, I have translated Black's evocation of a network of clear lines etched on smoked glass into talk of an overarching interpretive principle for selecting, classifying, and connecting a given subject's features into an overall, multi-dimensional 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, and more generally in the ubiquitous talk of metaphor and other tropes as producing 'seeing as' or 'perspectives', is that neither a characterization, nor the perspective expressed by a frame, *represents* an organizational structure. Rather, that structure must be implemented within the agent's intuitive cognitive processes, so that they really are more likely to notice and quicker to recall features that are weighted as more prominent; and so that they do intuitively connect or associate central features with many others. That is, frames offer cognitive Gestalts, much as the concepts 'old lady' and 'young lady' provide perceptual Gestalts for Figure 1.

The difference between representation and implementation introduces an important non-propositional dimension to characterizations, perspectives, and frames. In principle, with sufficient reflection and effort, an agent might be able to explicitly articulate the complete set of features they intuitively associate with a given subject. Likewise, in principle, with even more reflection and effort, they might be able to spell out the structure in which they intuitively arrange those features, assigning weights to reflect differences in prominence and drawing out explanatory connections. However, it is neither necessary nor sufficient for having a characterization 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 the operative characterization actually structures one's intuitive thinking about the subject in the relevant way. Perspectives and the frames that express them inherit this same implementational constraint, and add the non-propositional dimension of being open-ended tools for characterizing multiple subjects as information and experience comes in, whatever it may be.

Further, 'getting' either a characterization or a perspective is partly but not entirely under voluntary control. Sometimes, as with slurs, insinuations, and stereotype threat, frames may impose themselves on our thinking when we would rather resist (Camp 2013). 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 characterization, any finite bit of advice 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 importantly a matter of imagination, but primarily in the synthetic sense 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 experientially or abstractly conjuring contents that are taken not to be actually present, trying on a frame involves temporarily adopting a new perspective on a set 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). Altering the prominence or centrality of a single feature can often induce pervasive, complex alterations to the structural relations among other elements, by 'tipping' them into new clusters of explanatory and other dependence relations and new weightings of prominence. But the effects of applying a new frame may also extend beyond structural realignment, to produce alterations in the significance of the basic features themselves. In particular, shifts in prominence and explanation can affect classification by altering diagnostic import.

§2: Metaphors and Other Framing Devices

In §1, I deployed Black's central 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 assimilate all frames into a single type. Thus, Mary Hesse appears to treat models, narratives, fictions, analogies, and metaphors as fundamentally equivalent, and equally 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 emphasis 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 want to 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 various species of frame, and that only some rely on analogy in a significant way. In this section I identify some the key differences, and argue that they matter to how different frames guide everyday cognition and communication. In §3, I apply these distinctions to a variety of scientific models.

2.1 External and Internal Frames

While all frames provide overarching principles of interpretation for their target subjects, I take a crucial differentiating feature of metaphors to be 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 also 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 framing topic, *F*: in 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 lexical 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 then raised in prominence and centrality, resulting in a restructured characterization of *A* (and to a lesser extent of *F*). In certain circumstances, when it would be plausible for *A* to possess a feature *f* that could be matched to a prominent and central *F*-feature, but where no *f*-like feature is currently included in the *A*-characterization, *f* may be introduced into the

characterization of *A*. When a metaphor is employed 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 features between distinct characterizations in this way. At the broadest level, I think we should distinguish 'external' frames, which include metaphor, analogy, similes, and paratactic juxtapositions, from 'internal' frames, where the latter directly attribute a feature *f* to the subject *A* and raise that very feature to prominence and centrality in the *A*-characterization. The simplest internal frame is the 'telling detail' (Camp 2008), as vividly exemplified by classic cases of insinuation. So, for instance, a speaker who utters 'Obama's middle name is 'Hussein'' overtly asserts a fact that is itself undeniable, but thereby implicates that Barak Obama instantiates a cloud of more dubious features associated with a presupposed characterization of people named 'Hussein'. In effect, the speaker presents Obama's middle name as symbolic of a broader reality; focusing on it functions to highlight some known but otherwise unnoticed features and suggest other, as-yet unknown ones. While many insinuations, like this one, are insidious, other invocations of 'telling details' are less underhanded, and can be quite explicit. Thus, a primatologist might utter 'Trump is a primate', and go on to detail just how Trump's behavior can be explained and predicted by an analysis in terms of notable, relevant, causally influential properties of primates, especially social dominance (Camp 2008).

So, although both metaphors and telling details provide interpretive frames, they accomplish this in different ways. In particular, metaphors differ from telling details in operating 'from the outside', in the sense that there remains a felt gap or contrast between frame and subject. As we might put it, where telling details are interpretive keys inserted directly into the subject characterization, metaphors are colored lenses – sometimes, kaleidoscopes – through which we view the subject from elsewhere. Romeo, for instance, doesn't ask us to focus on the proposition that Juliet *is* the sun, in the way that the insinuating speaker focuses on the proposition that Obama is named 'Hussein'; instead, he invites us to attend to known and possible features of Juliet's that are relevantly *like* important features of the sun.

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 present the thought that Juliet actually *glows* (nor, as we will see in §2.2, does he ask us to pretend that she does). Rather, as his subsequent paraphrase spells out, the sun's luminosity is matched to the distinct feature of her (purported) beauty. That is, where the insinuated speaker attributes to Obama the very features purportedly possessed by most people named 'Hussein' – being foreign, dark-skinned, duplicitous, and opposed to core American values like freedom and beer – the match in Romeo's metaphor is *indirect*, in the sense that features which are themselves qualitatively distinct are related in virtue of sharing relevant higher-order properties.

Specifically, both the sun's luminosity and Juliet's beauty are highly intense: they are instantiated to a degree that so far exceeds the baseline exemplified by other individuals in their respective comparison classes that it constitutes a qualitative difference from those other individuals. However, the scale of intensity, the specific respect of intensity, and the comparison class are quite different: the sun is brighter than the moon, Venus, or Saturn; while Juliet is more beautiful than Rosalind or any other Veronese girl. The sun's luminosity and Juliet's beauty share other relevant features in addition to intensity: both are natural, and a source of energy and life; both produce a feeling of warmth. Again, however, these common higher-level features are implemented in qualitatively different ways.

2.2: Metaphor and Fiction

So far, I've argued that as an internal frame, a telling detail frames its subject directly and 'from inside', by highlighting and introducing features that *A* (purportedly) possesses in virtue of its possessing feature *F*; while an external frame operates indirectly, by matching distinct features. So far, this might just seem like a new label for the old difference between being literally true or false: absent literal truth, 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 internal, because they function in imagination *as if* they were true. In particular, I think 'just so' stories are clearly fictional but function much like 'telling details', and differently from metaphor (Camp 2009). Thus, a speaker might say that Trump acts *as if* he was denied admission to Harvard and has been compensating ever since, while explicitly acknowledging that this is not true.[3] Intuitively, such a speaker invites their hearer to pretend that Trump, in all his specificity – raised in Queens, having a real estate mogul father, etc. – really does possess the very feature of having been denied admission to Harvard, and to treat that in-fact-unrealized feature as an imaginative key to unlocking what really matters about him. That is, the hearer is supposed to pretend that the fictional feature *f* is actually instantiated by and explanatorily central to *A*, and to restructure their overall characterization of *A* by introducing and elevating other features from the *F*-characterization that *A* really would possess if it did instantiate *f*. Once this imaginative exercise is accomplished, the hearer drops the pretended ascription of *f*, leaving the characterization as close as possible to what it would be if *A* were in fact *F*.

The contrast between the sort of imaginative activity involved in fictional frames and metaphors is clearest when a single sentence, like 'Jane is a nurse,' can plausibly be deployed in either way. 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

[3] http://www.snopes.com/donald-trumps-harvard-rejection-letter/. Apocryphal facts are in effect just-so stories masquerading as telling details.

(Levin 1988). That is, the interpreter starts with actual-Jane and uses her as a imaginative prop to construct the fiction. This involves *transforming* Jane imaginatively in two ways: first, adding features that actual nurses do prominently possess: for instance, 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. Second, it involves downplaying features of actual Jane that conflict with these prominent and central nurse-features: for instance, her actual incompetence with machines or the fact that she works regular business hours. Once this imaginative transformation is accomplished, the pretense that Jane really is a nurse is dropped, but the highlighted features remain prominent and central. Thus, a natural use for offering 'Jane is a nurse' as a just-so story might be to elucidate first-order respects in which Jane really does perform the key functions of a nurse, even though she doesn't have a BSN or RN.

On the other hand, if the speaker employs the sentence as a metaphor, then interpretation begins with a characterization of nurses and seeks to identify respects in which Jane, as she already currently actually is, is nurse-*like*. That is, rather than directly attributing actual nurse-features to pretend-Jane, the interpreter of a metaphor construes actual-Jane in a nurse-like way. And as with Juliet, this imaginative activity focuses attention on actual features of Jane's that are not generally actually possessed by nurses, but instead share higher-order structural similarities with prominent and central features in the stereotypical characterization of nurses. Plausible features of this sort might include consistently lending a sympathetic ear, but for friends rather than assigned patients; checking on those friends' emotional and psychological well-being, rather than their physical symptoms and statistics; or nudging them toward avenues of emotional and psychological improvement, rather than delivering pills and injections. Many people who have spent sustained time in a hospital will attest that these stereotypical features of nurses are not universally instantiated: some nurses are not sympathetic, consistently attentive, or improvement-oriented. But even the many nurses who are stereotype-conforming are, in the nature of their jobs, focused on their patients' physical state, while Jane is presented as focused on friends' emotional health.

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 in order to learn something more about the prop itself. What we aim to learn might be something about its counterfactual possibilities, or about what it's actually 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. 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).

I agree that the kinds of imagination overlap, and that many utterances invite a mixture of both modes of interpretation (Camp 2009). Both frames are indirect, in the sense that we imaginatively step away from our actual assumptions about *A*; and both are guided by our intuitive characterizations about *A* and *F*s. However, as I've argued, there is an important difference between the two tropes and the sorts of imaginative activity they invite, at least in paradigmatic cases. With a just-so story, we temporarily transform the prop *A* into a counterfactual counterpart, by imputing actual *F*-features to *A*; 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; the indirectness is generated at the interpretive level, by matching features of *A* and *F*s that are merely similar. 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).

2.3: Metaphor and Analogy

In drawing the contrast between 'external' and 'internal' frames, I have distinguished metaphors and analogies, on the one hand, from telling details and just-so stories, on the other. In particular, I have emphasized that metaphors are indirect in relying on abstract structures of higher-order similarities between distinct lower-level features. In this way, my view is closely akin to Dedre Gentner's "structure-mapping" theory of analogy. In this section, I argue that there are also important differences between metaphor and analogy. Specifically, as Gentner points out, metaphors are significantly less constrained than analogies, in at least two ways.

First, where Gentner claims that analogies rely exclusively on abstract, higher-order similarities, metaphors frequently employ qualitative matches first-order features. Thus, although the match between the sun's luminosity and Juliet's beauty is a structural one, Romeo's metaphor also suggests that being near Juliet produces a physical feeling in him that is not just structurally but qualitatively similar to the glow produced by the sun on a warm spring day. Briefly, we can classify the types of matches that metaphors rely on into four broad categories (Camp 2003). Some matching *A*-features are qualitatively identical to the matched *F*-feature: thus, a beloved's hair really is the color of ebony, or copper. Others differ from the *F*-feature in degree along a common qualitative dimension: thus, an admired colleague might be 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 identical higher-order properties. Finally, some matches are themselves metaphorical, albeit less novel and rich than the overall framing metaphor: so, 'Sally is a block of ice' communicates a temperamental attitude, of being unemotional, that relies on a 'deep' conceptual association grounded in cultural and/or biological experience (Searle 1979, Lakoff and Johnson 1980).

Second, metaphors are also less constrained than analogies in permitting a looser preservation of structure between framing and subject characterizations. In analogy, potential matches that are not embedded within more complex structures 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 within the frame must be replicated in the subject for the analogy to be sound, and known, relevant failures of match compromise the analogy's plausibility. By contrast, metaphors are often fairly unsystematic. Thus, Othello's description of Desdemona as "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. The lack of systematicity does not undermine the metaphor's effectiveness as a frame, since it suggests such a rich range of matches with robust affective and imagistic elements.

The greater permissiveness of metaphors also makes their interpretation more imaginatively intuitive and holistic. Rather than puzzling out a precise, consistent formal mapping between complex, abstract, articulate structures, we more often 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. But how do these differences play out in a scientific context? More specifically, what implications do they have for models and modeling? As an initial point, 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: representational vehicles that express intuitive interpretive principles which guide overall thinking about a target system, by determining both *what* (purportedly) matters about that subject relative to a presupposed

taxonomy, and *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 illuminating their functional commonalities. In this section, I tour some important types of scientific frame, focusing on the different sorts of gap they assume between representation and reality, and their different ways of bridging that gap.

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 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 reveal relevant causal structure. While androcentric theories focus on hunting behavior and the relative efficacy of stone tools over sticks; gynecentric theories 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) offer different predictions about, and affective and normative responses to, sex, tool use, and intelligence in contemporary humans. Insofar as the primary locus of disagreement is a higher-order, interpretive one, 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 mere anomalies or as true but marginal.

Thus, 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: it functions as a key that unlocks what matters about *A*. By contrast, the 'telling instance' or sample is itself (at least typically) concrete, and investigation involves probing it directly, in (typically) concrete ways: say, by feeding the mouse, or half of the mice population, more saturated fat.

Second, the two types of 'telling' frame differ in the direction of interpretive attention. In the case of the 'telling fact', like the insinuation about Obama, 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*; 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, and possibly thereby suggests causal connections between those *F*-features and additional non-*F* features within *A*. By contrast, in the case of a 'telling instance', the primary focus of investigative attention is on the subject *a* itself. *F* provides the frame for thinking about *a*, insofar as the sample *a* matters itself matters only as an instance of the general kind *F*, and assumptions and questions about *F*-ness fix an initial characterization of *a*, by selecting only some, taxonomically relevant features as being exemplified. Further, 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, and using discoveries about *a* to illuminate *F* in general.

3.2: Abstraction and Idealization

Both 'telling facts' and 'telling instances' are intuitively true, in the basic sense that the predicate *F* cited in the framing sentence does indeed apply to the subject *A*. Rejecting this core 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, or at least not true – 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 entities together as alike in some respect, or the same entity as recurring on multiple occasions, both of which require abstracting away from differences between those distinct entities or occasions (Camp 2015). Further, we do regularly, and rightly, criticize

representations (and representers) for neglecting relevant features. Sometimes the criticism is for ignoring features that are diagnostic relative to the representation's own presupposed taxonomy; but sometimes we challenge the taxonomy itself, either for assuming that certain kinds of features tend to cluster together or are 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. While speakers can obviously mislead or be misinterpreted, a representation itself can only be false if it positively represents a state of affairs as obtaining which does not obtain.[4] Moreover, because assessment for truth can only take place against the background of a presupposed taxonomy, the very assumption of a taxonomy cannot itself be grounds for falsity, but rather for inappropriateness of some other variety.

Insofar as abstraction does not itself introduce falsity, it differs 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, for instance as a way to isolate key causal factors (Elliott-Graves and Weisberg 2014). But where abstraction engages in "mere omission" (Thomson-Jones 2005), by remaining silent about features that are 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. While some idealizations are straightforward, idealizing in one respect often affects the values of other, related features; insofar as idealizing agents are often not aware of such interdependent effects, they need to actively probe for unrecognized falsities that matter relative to the operative taxonomy and explanatory purposes.

The contrast between abstraction and idealization both emerges clearly and brings out the contrast between 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 *F*s in general. But the use of 'telling instances' as models combines uneasily with idealization, insofar as idealization imaginatively constructs an entity that differs from the actual target subject, and so inevitably shifts attention away from probing the sample itself. By contrast, the use of 'telling facts' as frames is fully compatible with

[4] Speakers are especially likely to exploit, and insist on, the difference between active misrepresentation and mere non-representation in strategic conversational contexts (Camp 2017). Assessing falsity is more complex in extended conversations, where representations are embedded within entailed structures of presupposition and relevance (Roberts 2012, Stokke 2016). To the extent that scientific theories (as opposed to inquiry) also exhibit discourse structure, the distinction between semantic falsity and pragmatic implication likewise becomes more complex.

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 employing contrasting models of tool use based on starkly differentiated 'male' and 'female' roles; or the two theories might each invoke highly idealized "agent-based models" that compute the long-term dynamic effects of repeated interactions between individuals, where those individuals are defined by just a few traits defined in terms of gender roles.

3.3: Fact and Fiction

If idealization, unlike abstraction, does introduce a form of known falsity, does this make all idealizations fictions? Here too, I think we should resist assimilation. The falsification introduced by idealization is still like abstraction in ignoring (purportedly) irrelevant complexities of the target subject. 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 the subject's assumed reality, and 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). 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 helps us to understand how a perpetual 'heat engine' could be physically or metaphysically possible while still being extremely unlikely (Stuart 2016, 27).

Unlike paradigmatic cases of idealization as ignoring or "imagining away," that is, Maxwell's thought experiment directs investigative attention toward a situation which is overtly counterfactual.

Much as with a just-so story, we are asked to imagine that this very situation, as described, is true; and we do so in order to highlight other features that follow directly from the framing proposition, but that are actually (purportedly) true. Assessing the fiction's aptness as a frame is thus a matter of determining two things: first, what is true within the fiction: what unstated implications follow from the propositions it literally and explicitly articulates, given its 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 vexed (Lewis 1978, Walton 1990, Byrne 1993), with different genres employing importantly different principles. However, these issues are considerably less pressing for a scientific fiction like Maxwell's: 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 within the 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 seek to identify relevant respects in which the frame and subject 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, indirect mapping employed by external frames, and of 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 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 how they are implemented within the frame and the target.

The difference between fiction and metaphor or analogy becomes more palpable if we contrast Maxwell's original thought experiment with its subsequent deployment as a metaphor. 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 that is (purportedly) shared by Maxwell's fictional situation and actual schools: of an entropy-reversing and therefore 'unnatural' mechanism which produces global effects by sorting individuals. But again, this common structure is implemented in very different ways in each case. And where Maxwell's fiction directs attention at the target phenomenon itself – the trajectory of distribution of heat in a closed volume – and asks us to imagine something literal but counterfactual about *it*, Bourdieu applies that structure to a very different domain.

A proponent of assimilating metaphor, analogy and fiction might point out that analogy, and to a lesser degree metaphor, present the frame and target 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. Given this, at a suitably high level of abstraction analogical and metaphorical frames do impute features that are actually possessed by the framing subject to the target – just as a just-so story imputes features possessed by the fictional subject to the actual target. The proponent of a unified fictionalist account of scientific models might propose that any difference between metaphor and fiction is one of the level at which common features are imputed, rather than a difference between, on the one hand, pretending that a nonfactual feature *f* really does apply in order to impute further features that would follow from *f*, and on the other hand identifying matches between merely similar features.

Unsurprisingly, I want to reject this analysis: I think it distorts the real representational import of analogy and metaphor, both in everyday discourse and in science. The claim made by a metaphor or analogy is not merely that the target is like the frame in this common, highly abstract respect, but rather that the target possesses a lower-level feature which is identified by means of instantiating this higher-order feature. For instance, Romeo claims, among other things, that Juliet is more beautiful than the other Veronese girls, not just that she is comparatively maximally intense in some respect or other.

In the context of scientific models, we might put the point by saying that metaphors and analogies do not typically function as abstract models, such as the Lotka-Volterra equations describing the effects of predator-prey dynamics on population distribution. Such abstract models prescind from messy detail in order to focus attention on high-level, general, purely structural features. By contrast, in metaphor and analogy, the shared high-level features warrant attention only instrumentally, as a means for identifying the more specific lower-level feature within the target. In a pedagogical context – for instance, when explaining electrical current by analogy to the flow of water through a pipe – the speaker will explicitly

identify, or ask listeners to identify for themselves, those lower-level instantiating features. But in a context of discovery, the investigators typically don't yet know what they are; they employ the possibility of a structural match between frame and target precisely as a principle for investigation. And in both pedagogical and investigative contexts, the structural match serves as a guide for focusing attention on the basic-level features.

So far, I've been emphasizing ways in which both metaphor and analogy differ from fiction; I've argued that they involve a qualitatively greater gap between representation and reality than fiction, idealization or abstraction do, because they rely 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. Gentner and Jeziorski (1993) argue that contemporary scientific practice valorizes analogies, such as Bohr's solar model, because they employ precise, consistent, systematic matches between complex, causally-connected systems of features. However, they also claim that this is distinctive to modern Western science. In particular, they argue 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. The birth of modern science, they claim, is due in significant part to this shift from promiscuous similarity to higher-order structural matching. The upshot is that metaphor in contemporary science serves primarily as a poor cousin to analogy, as encapsulated by George Polya (1954): "And remember, do not neglect vague analogies. But if you wish them respectable, try to clarify them."

I have largely followed Gentner in emphasizing the ways that metaphor approximates to analogy, especially in deploying selective, structural, indirect mappings. Further, Gentner and Jeziorski's priority claim about modern scientific practice is importantly right in several respects. First, metaphors in science, in contrast to literature, are typically more analogy-like, in emphasizing fewer, more consistent matches over richer, inconsistent ones – especially in the contexts of pedagogy and theoretical advocacy, which are the cases that Gentner and Jeziorski focus on almost exclusively. Further, it is widely agreed that at least one central aim of science is to develop a precise, articulate understanding of objects, properties, and their relations, and that to accomplish this, we need 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 modern scientific metaphors 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 structural relations while remaining fairly neutral about implementational mechanisms, leaving the connection between functional role and physical substrate to be forged after each side is understood better in their own terms – a strategy that Pylyshyn (1993, 551) calls the "principle of least commitment" or "principle of procrastination." And as we saw in §2.3, the selectivity and abstractness of analogies makes them especially well-equipped for identifying novel structural properties.

Thus, Polya's dictum about making vague analogies respectable by articulating precise structural relations is largely apt. However, 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 engages 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 a code model of genetic potential and action lacked clear, coherent characterizations of both target systems *and* 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 were mutually supporting, with each serving as a frame for 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 did the metaphor of complex machines as organisms orient research within systems analysis and cybernetics, in turn reciprocally influencing theories of biological development and cellular coordination.

In effect, each metaphor provided what Richard Boyd (1979/1993, 488) calls an "inductive open-endedness": they guided research by gesturing toward a 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) 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 and 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.

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 an agent's interests and goals. As a result, different scientists will often bring markedly different characterizations and perspectives on their subjects to the interpretive table, especially at the beginning of inquiry. Further, even given a fixed pair of characterizations of both target and frame, multiple plausible overall mappings between them will nearly always be available, which trade off preferences for systematicity against directness in matching, and for identifying new features and connections against preserving already known ones, in different but legitimate ways. Beyond this, frames do more than just interpret a fixed set of assumptions about their targets: they provide open-ended tools for assimilating new information and for generating hypotheses about 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 depends in deep, important ways on external factor, including on what alternative theories it is being compared to, and so what expressive and epistemic needs it distinctively addresses (Okruhlik 1994), as well as on its interaction with current technological opportunities and limitations (Fox Keller 1995).

Perhaps the best way 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): beginning with an intuitive, holistic, and open-ended, and therefore diffuse and relatively unarticulated mode of construing one subject in terms of something else, where one or both domains may be only minimally understood; moving through a process of articulating, probing, and refining the characterizations of one or both domains and 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 literally applicable in a few restricted respects, as has arguably happened with 'computation' and with gravitational 'waves'. 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 thought, or imputes features that are not in fact possessed. Or perhaps it was once genuinely fruitful, and continues to identify features which are both prominent and central, but 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 (Camp 2015).

3. Models and Frames

Much current philosophical discussion about scientific models has focused on their ontological status – in particular, on 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 share remarkably many features, and are used for many common epistemic purposes. One benefit of turning at least temporarily to an investigation of frames is that it helps to integrate 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 also commonalities and differences between the practice and evaluation of those strategies in science and in everyday cognition and communication.

Specifically, I have argued that frames are representational vehicles which provide an overarching interpretive principle or perspective. 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 actually implementing rather than merely representing these interpretive structures.

However, within this broad genus, different species of frames function quite differently. Frames themselves can be more or less articulated, 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 or 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 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 or most of its subsidiary features, directly into the target subject. Others, like Maxwell's Demon, assimilate that feature directly but only temporarily, in order to highlight or introduce subsidiary features that the target really would have if that 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 by an indirect mapping that is at least initially inchoate, potentially inconsistent, and 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

Belyaev D. K. (1978): "Destabilization as a factor in domestication," *The Journal of Heredity* 70:301-308.

Black, Max (1962): "Metaphor," in *Models and Metaphors: Studies in the Philosophy of Language* (Ithaca: Cornell University Press), 25-47.

Bourdieu, Pierre (1998): *Practical Reason: On the Theory of Action* (Cambridge UK: Polity Press).

Byrne, Alex (1993): "Truth in Fiction: The Story Continued," *Australasian Journal of Philosophy* 78:1, 24-35.

Bohr, Niels (1913): "On the Constitution of Atoms and Molecules, Part I," *Philosophical Magazine* 26:151, 1–24.

Boyd, Richard (1979/1993): "Metaphor and Theory Change: What is 'Metaphor' a Metaphor For?" in *Metaphor and Thought*, 2nd ed. A. Ortony (Cambridge: Cambridge University Press), 481-532.

Camp, Elisabeth (2015): "Logical Concepts and Associative Characterizations," in *The Conceptual Mind: New Directions in the Study of Concepts*, ed. E. Margolis and S. Laurence (Cambridge MA: MIT Press), 591-621.

——— (2013): "Slurring Perspectives," *Analytic Philosophy* 54:3 (September 2013), 330-349.

——— (2009): "Two Varieties of Literary Imagination: Metaphor, Fiction, and Thought Experiments," *Midwest Studies in Philosophy: Poetry and Philosophy* XXXIII, ed. Howard Wettstein (Oxford: Wiley-Blackwell), 107-130.

——— (2008): "Showing, Telling, and Seeing: Metaphor and 'Poetic' Language," *The Baltic International Yearbook of Cognition, Logic, and Communication,* vol. 3: *A Figure of Speech: Metaphor* [Online], 1-24.

——— (2006): "Metaphor and That Certain 'Je Ne Sais Quoi'," *Philosophical Studies* 129:1, 1-25.

——— (2003): *Saying and Seeing-As: The Linguistic Uses and Cognitive Effects of Metaphor*, Dissertation, University of California, Berkeley.

Currie, Gregory and Ian Ravenscroft (2003): *Recreative Minds: Imagination in Philosophy and Psychology* (Oxford: Oxford University Press).

Elgin, Catherine (2006): "From Knowledge to Understanding," *Epistemology Futures,* Stephen Hetherington, ed. (Oxford: Clarendon Press), 199-215.

Elliott-Graves, Alkistis and Michael Weisberg (2014): "Idealization," *Philosophy Compass*, 9: 176–185.

Fox Keller, Evelyn (1995): *Refiguring Life: Metaphors of Twentieth Century Biology* (New York: Columbia University Press).

Friend, Stacie (2008): "Imagining Fact and Fiction," in Kathleen Stock & Katherine Thomson-Jones (eds.), *New Waves in Aesthetics* (London: Palgrave Macmillan), 150-169

Frigg, Roman (2010): "Models and fiction," *Synthese* 172:251–268.

Markman, Arthur and Dedre Gentner (1993): "All differences are not created equal: A structural alignment view of similarity," *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, 682-686.

Gentner, Dedre and Michael Jeziorski (1993): "The shift from metaphor to analogy in Western science," in A. Ortony, ed. *Metaphor and Thought*, 2nd ed. (Cambridge UK: Cambridge University Press), 447-480.

Giere, Ronald (1988): *Explaining science: a cognitive approach* (Chicago: Chicago University Press).

Godfrey-Smith, Peter (2009a): "Models and fictions in science," *Philosophical Studies*, 143, 101–116.

——— (2009b): "Abstractions, Idealizations, and Evolutionary Biology," in A. Barberousse, M. Morange, and T. Pradeu (eds.), *Mapping the Future of Biology: Evolving Concepts and Theories*. (*Boston Studies in the Philosophy of Science*) (Dordrecht, Netherlands: Springer) 47-56.

——— (2006): "The strategy of model-based science," *Biol Philos* 21:725–740.

Hesse, Mary (1993): "Models, Metaphors and Truth," *Knowledge and Language, vol. 3: Metaphor and Knowledge*, ed. Ankersmit, F.R. and Mooij, J.J.A. (Dordrecht, Netherlands: Springer), 49-66.
Hills, David: "Aptness and Truth in Verbal Metaphor," *Philosophical Topics* 25:1 (Spring 1997), pp. 117-153.
Kuhn, Thomas (1962): *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press).
Lakoff, George and Mark Johnson (1980): *Metaphors We Live By* (Chicago: University of Chicago Press).
Lerner, Jennifer, Roxana Gonzalez, Deborah Small, and Baruch Fischhoff (2003): "Emotion and perceived risks of terrorism: A national field experiment," *Psychological Science* 14:2, 144–150.
Levin, Samuel (1988): *Metaphoric Worlds: Conceptions of a Romantic Nature* (New Haven: Yale University Press).
Levy, Arnon (2015): "Modeling without models," *Philosophical Studies* 172, 781–798.
Lewis, David (1978): "Truth in Fiction," *American Philosophical Quarterly* 15:1, 37-46.
Longino, Helen and Ruth Doell (1983): "Body, Bias, and Behavior: A Comparative Analysis of Reasoning in Two Areas of Biological Science" *Signs* 9, 206-27.
Maxwell, J. Clerk (1871): *Theory of Heat* (London: Longmans, Green).
Murphy, G., and D. Medin (1985): "The role of theories in conceptual coherence," *Psychological Review* 92: 289-316.
Nagaoka, H. (1904): "Kinetics of a System of Particles illustrating the Line and the Band Spectrum and the Phenomena of Radioactivity," Philosophical Magazine 6(7): 445-455.
Okruhlik, Kathleen (1994): "Gender and the Biological Sciences*," Canadian Journal of Philosophy*, Supp. volume 20, 21-42.
Peacocke, Christopher (1992): *A Study of Concepts* (Cambridge MA: MIT Press).
Pylyshyn, Zenon (1993): "Metaphorical imprecision", in *Metaphor and Thought*, 2nd ed., A. Ortony ed. (Cambridge: Cambridge University Press), 481–532.
Quine, Willard V. O. (1951): "Two Dogmas of Empiricism," *Philosophical Review* 60 (1):20–43.
Roberts, Craige (2012): "Information structure in discourse: Towards an integrated formal theory of pragmatics," *Semantics & Pragmatics* 5:6, 1–69.
Rosch, Eleanor (1978): "Principles of Classification," in *Cognition and categorization* ed. E. Rosch and B. Lloyd (Hillsdale, NJ: Lawrence Erlbaum), 27-48.
Rutherford, Ernest (1911): "The Scattering of α and β Particles by Matter and the Structure of the Atom," *Philosophical Magazine* 6:21.
Searle, John (1979): *Expression and Meaning*: *Studies in the Theory of Speech Acts* (Cambridge: Cambridge University Press).
Sloman, Steven, Bradley Love and Woo-Kyoung Ahn (1998): "Feature Centrality and Conceptual Coherence," *Cognitive Science* 22:2, 189-228.
Small, Deborah, Jennifer Lerner, and Baruch Fischhoff (2006): "Emotion Priming and Attributions for Terrorism: Americans' Reactions in a National Field Experiment," *Political Psychology* 27:2, 289-298.
Stokke, Andreas (2016): "Lying and Misleading in Discourse," *Philosophical Review* 125(1), 83–134.
Strevens, Michael (2008): *Depth: An Account of Scientific Explanation* (Cambridge MA: Harvard University Press).
Stuart, Michael (2016): "Taming theory with thought experiments: Understanding and scientific progress," *Studies in History and Philosophy of Science* 58: 24-33.
Thagard, Paul (1989): "Explanatory coherence," *Behavioral and Brain Sciences* 12:435–502.
Thomson, J.J. (1904): "On the Structure of the Atom: an Investigation of the Stability and Periods of Oscillation of a number of Corpuscles arranged at equal intervals around the Circumference of a Circle; with Application of the Results to the Theory of Atomic Structure" *Philosophical Magazine* 7:39, 237–265.
Thomson-Jones, M. (2005): "Idealization and Abstraction: A Framework," in M. Thomson-Jones and N. Cartwright (eds.), *Idealization XII: Correcting the Model* (Amsterdam: Rodopi), 173-217.

Tiedens, Larissa and Susan Linton (2001): "Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing," *Journal of Personality and Social Psychology,* 81(6):973-88.
Trut, Lyudmila (1999): "Early Canid Domestication: The Farm-Fox Experiment," *American Scientist* 87: 160-169.
Walton, Kendall (1993): "Metaphor and Prop-Oriented Make-Believe," *European Journal of Philosophy* 1(1): 39-57.
——— (1990): *Mimesis as Make-Believe*: *On the Foundations of the Representational Arts* (Oxford: Oxford University Press).
Weisberg, Michael (2012): *Simulation and Similarity: Using Models to Understand the World* (Oxford: Oxford University Press).
Wittgenstein, Ludwig (1953): *Philosophical Investigations*, trans. G.E.M. Anscombe (Oxford: Blackwell).
Woodfield, Andrew (1991): "Conceptions," *Mind* 100 (399):547-72.
Woodward, James (2003): *Making Things Happen: A Theory of Causal Explanation* (Oxford: Oxford University Press).
Yablo, Steven (2001): "Go Figure: A Path Through Fictionalism." *Midwest Studies in Philosophy* 25: 72–102.