JNφ
phi

Blur and Knowledge from Falsehood: Neural Network Science and Neurophysiology Meets Epistemology

Jody Azzouni

Abstract

Convolutional neural networks (CNNs) trained on clear images fail catastrophically with degraded or blurry imagery. New results by Jang and Tong, and Pramod, Katti and Arun show visual object recognition is optimized by introducing peripheral blur. Optimizing recognition of objects this way empirically supports the significance of there being a hundred times less photoreceptors dedicated for peripheral vision than in the retina. These results refute a longstanding epistemic slogan: Knowledge of truths arises only from knowledge of truths. Blur-trained CNNs and humans recognize things in blurry, degraded and noisy environments—a dog, a radiator—that clear-image-trained CNNs don't. Blurring is *misinformation* about what is seen, so the human perceptual system recognizes objects by processes that start from falsehood. Peripheral blur—misinformation about what is seen—is essential to perceptual knowledge.

Key Words: convolutional neural nets, knowledge because of falsehood, knowledge from falsehood, peripheral blur, Gettier cases, Hilpinen cases

281

Introduction

Consider the task that human visual systems face with respect to object recognition: Balancing *selectivity* with *tolerance*. That is, they must optimize getting the right answer ("that's a dog") with being able to recognize that *the same answer* ("that's the same dog again") is often called for in different viewing conditions—different environments, angles of view, amount of noise, background and otherwise, etc.

As it turns out, humans are extraordinary at visual object recognition:

Corresponding author: Jody Azzouni
Address: Department of Philosophy- Miner Hall, Tufts University, Medford, MA 02155
e-mail ✉ jody.azzouni@tufts.edu

The apparent ease of our visual recognition abilities belies the computational magnitude of this feat: we effortlessly detect and classify objects from among tens of thousands of possibilities … and we do so within a fraction of a second … despite the tremendous variation in appearance that each object produces on our eyes … (DiCarlo et al., 2012, p.415).

The striking fact is that until very recently, convolutional neural networks (CNNs) trained on images, even enormous numbers of images, were unable to manage the reliability of the human capacity for visual object recognition. A second point is lurking here. Not only, that is, were the neural networks in question not as good as humans at recognizing objects in visual scenes, they also remained dramatically poor in certain circumstances at modeling human visual capacity at all:

> … recent studies have found that deep convolutional neural networks (CNNs) trained on tasks of object recognition provide the best current models of the visual system, allowing for reliable prediction of visual cortical responses in humans and neuronal responses in the macaque inferotemporal cortex. While these initial findings are highly promising, a mounting concern is that CNNs tend to catastrophically fail where humans do not, especially when presented with noisy, blurry or otherwise degraded visual stimuli. Such findings demonstrate that the computations and learned representations of these CNNs are not truly aligned with those of the human brain (Jang and Tong, 2024, p.1).

How did the recent work change this? The CNNs heretofore were trained on clear (and sharp) images. But that isn't the data that humans—the human eye—has access to. Instead, as is well-known, visual blur (Jang and Tong, 2024, p.1) "is pervasive in everyday human vision." For example, although the fovea can process stimuli with high spatial resolution, this ability drops off rapidly towards the periphery of the eye. Human visual practice, apparently as a result, involves multiple eye movements where what's in focus changes rapidly. Thus (Jang and Tong, 2024, p.2): "low-resolution vision and blur are prominent features of everyday vision. By contrast, the image datasets commonly used to train CNNs predominantly consist of clear, well-focused images."

The predominance of peripheral blur in human (and animal) vision is thought to be a metabolic "cost-saving" maneuver. "It is widely believed that this sampling scheme [retinas contain 100 times more photoreceptors compared to the periphery] saves on the metabolic cost of processing orders of magnitude more information that would result from full resolution scenes without affecting overall performance" (Pramod *et al.,* 2022, 1). It has, therefore, only been recently recognized that this isn't the case—that the pervasiveness of blur (in human vision) *optimizes* the capacity to recognize objects—

both with respect to humans and with respect to CNNs. That peripheral blur does this has been empirically established by (essentially) training CNNs on images that are blurred *just in* the respect that the imagistic data the human eye processes are blurred. The striking results are threefold:

> Such blur-trained CNNs outperform ones trained on clear images.

> Such blur-trained CNNs outperform blur-trained ones with either more blur or less blur than what's provided by the human eye. The structure of the human eye apparently supplies exactly the appropriate amount of peripheral blur to optimize the recognition of objects (Pramod et al., 2022). Evolution has done right by us (at least in this case).

> The object-recognition capacity of blur-trained CNNs closely tracks human capacities. For example, they don't "catastrophically fail" in noisy, blurry or degraded visual environments (Pramod *et al.*, 2022; Jang and Tong, 2024).

Exactly what is it about peripheral blur that enables networks "trained on images foveated according to the human peripheral blur function [to give] the best performance"? (Pramod et al., 2022, p.7) One must speculate to some extent because the results, as of now, only tell us that CNNs when trained on certain kinds of images perform optimally (or not). One possible factor is that if the peripheries of images are blurred out, this lowers the chances of false positives (Pramod et al., 2022, p.11)—a peripherally blurred fire hydrant in the background is less likely to be identified as a person than a clear image of the same fire hydrant in the background.

Another possible factor is that by introducing blur into the images that CNNs are trained on, one increases the *tolerance* of those networks to recognizing objects as invariantly the same despite differing viewing conditions. Further, clear-image-trained CNNs can misidentify two (different kinds of) objects as the same because their textures are the same (polka-dotted dogs and cats); blur-trained CNNs don't do this as often as they seem to be shape-focused rather than texture-focused. Textures are blurred out relative to shape, which isn't distorted as much. Blur-trained neural networks can recognize the same object despite different viewing conditions, as opposed to clear-image-trained neural nets for the same reason: because the salient details that most successfully enable identification of an object (shape, most notably) aren't lost in a welter of other details that don't as successfully point towards two viewings of something being of the same object. The fact that blur-trained CNNs, which are more sensitive to shape than to texture, do better at object recognition suggests that shape, generally speaking, is a more significant cue for object identity than texture is. Blur-trained CNNs are sensitive to "significant aspects" of objects, such as the bottlecap on a bottle or the head of a

283

dog (Jang and Tong, 2024, p.5). Notice: significant aspects of objects are ones apparently marked by factors such as contours or detailed differences in the shape(s) of their parts. The general sensitivity of clear-image-trained CNNs to texture (and the corresponding less sensitivity of blur-trained CNNs) may also explain why blur-trained CNNs do better with respect to noisy visual environments.

I'll turn now to philosophical ramifications, but preliminarily to that, to a little history of 20th century analytic epistemology. Then I'll return to the empirical results described in Section 1, and bring them to bear on a new issue that's emerged in analytic epistemology in the last 20 years or so. To start: 20th and 21st century analytic epistemology is based firmly on the evaluation of examples ("thought experiments"), and the shared intuitions of those who consider those examples. A classic example is due to Gettier (1963); other examples, along the same lines, subsequently became known as "Gettier cases":

Ford Smith has good reasons to believe that Nogot owns a Ford and that Nogot works in Smith's office. From this Smith infers that someone in his office owns a Ford. Nogot doesn't own a Ford, although someone in Smith's office does: Havit works in Smith's office and owns a Ford, although Smith doesn't know this. Smith, thus, doesn't know that someone in his office owns a Ford, even though he's justified in believing someone does, and it's true. (This version of Gettier's original example is due to (Lehrer, 1965)).

284

If the reader (you) agrees with the description of **Ford**, what does it show? According to Gettier there was a longstanding definition of knowledge in place, **Tripartite**, dating back to Plato:

Tripartite An agent knows a proposition if and only if: that proposition is true, the agent believes that proposition, and the agent is justified in believing that proposition.

Ford is a counterexample to this definition. Smith believes Nogot owns a Ford, and Smith is justified in believing that Nogot owns a Ford. Nevertheless, he doesn't know this since Nogot *doesn't* own a Ford. (There are different versions of why Smith is justified in believing what isn't true. Perhaps Nogot sold his Ford just that morning, or it was just stolen, etc.) He then draws an inference from what he's justified in believing to something weaker: *Someone* in the office owns a Ford. Smith is justified in believing whatever he deductively infers from what he justifiably believes. Furthermore, the statement he's deduced is true because of Havit, although he doesn't realize this. The result is a statement—someone in the office owns a Ford—which Smith is justified in believing and which is true but which he doesn't know to be true.

Why doesn't Smith know that someone in the office owns a Ford? The obvious thought is that someone in the office owning a Ford is true by virtue of facts that are independent of the reasoning that

Smith has engaged in to draw his conclusion. In any case, for decades after (Gettier, 1963), a host of papers were published, all attempting to patch up the definition of knowledge (**Tripartite**) that **Ford** is a counterexample to. This is a fascinating bit of philosophical history, but not directly relevant to the concerns of this paper. The main point is that philosophers understood Gettier cases to always involve assumptions (often false) that speakers assumed and/or missteps in inference, and that this was key to explaining why Gettier cases didn't yield knowledge. Indeed, a candidate patch on **Tripartite** went as follows:

> An agent knows a proposition if that proposition is true, if the agent believes that proposition, if that agent's belief in the proposition is justified, *and* if the agent's reasoning isn't based on intermediate inferential missteps or false premises (Harman, 1973).

But there are Gettier cases, that is, counterexamples to **Tripartite**, which don't involve assumptions or inferences but arise directly, as it were, from sensory experience. Consider:

Barn: Henry is driving through an area where, unbeknownst to him, the inhabitants have erected a large number of fake barns facing the highway. Fake barns are single-sided facades painted to look like barns. There is, though, one real barn among all the fakes, and Henry *happens* to be looking right at it. Henry's looking at the barn causes him to believe that *there is a barn there* (the one that he's looking at). We don't think Henry *knows* that there is a barn he's looking at. Goldman (1976, 772-773). (This version is from (Bernecker 2022, 1626)).

285

Here, Henry simply points at something and says (or thinks) it's a barn; there's no inference or assumptions that Henry is basing his thought on. What seems to falsify Henry knowing there's a barn in front of him (although there is) is, again, a certain contingency that's in play here. Had Henry pointed moments sooner or later than he did, he would have been pointing at a fake barn. As I mentioned, **Barn** also violates **Tripartite**. Henry believes there's a barn in front of him and he's justified in believing this. (Who, after all, erects a lot of fake barns near highways? This isn't something anyone should expect.) But he doesn't know there's a barn that he's pointing at.

Nevertheless, philosophers have not taken **Barn** to be a Gettier case. And the lesson from Gettier cases, and the working assumption behind trying to repair **Tripartite**, has largely remained a strategy of supplementing **Tripartite** with extra conditions thought to exclude Gettier cases. The thought, that is, has remained that knowledge, in any case, should arise and can only arise from knowledge. Gettier cases, it's assumed, occur precisely because one can be justified in believing something true but the inferences in question aren't good because they don't start from something one knows. It's this

longstanding (and largely unquestioned) assumption that philosophers began to undercut in the 21ˢᵗ century and that the empirical results described in section 1 bear on. I turn in the next section to the examples (thought experiments) that philosophers gave that undercut the knowledge-only-from-knowledge requirement.

That knowledge can only arise from knowledge can be codified in the following principle:

Knowledge counter-closure: If someone believes a proposition solely on the basis of a set of premises and if that someone knows that proposition, then that someone knows those premises as well.

This definition is based on (Luzzi 2019, 11).

Here are two thought experiments, given, respectively, in 1988 and 2005, that seem to undercut **Knowledge counter-closure**: only the latter publication (Warfield 2005) drew philosophical attention to counterexamples to knowledge counter-closure—attention that continues to this day. Actually, an even earlier counterexample to knowledge counter-closure was published in 1964, but subsequently ignored, its significance apparently unnoticed. This was (Saunders and Champawat, 1964).

Temperature: A mother suspects her child has a temperature, and when she measures the temperature and looks at the thermometer, she takes it to read 40.0 degrees Celsius. If the thermometer is fairly accurate and the mother has reasonably good eyesight, we can say … that she knows that the child has a temperature …. (To say that the child has a temperature is just another way of saying that the temperature of the child is more than 37 degrees Celsius.) But the mother need not have perfect eyesight and the thermometer need not be completely accurate … this does not prevent her from knowing [this] … (Hilpinen, 1988, p.163).

Handout: Counting with some care the people present at Warfield's talk, Warfield reasons: "There are 52 people at my talk; therefore my 100 handout copies are sufficient." Warfield's premise is false. There are 53 people in attendance—he overlooked someone who came in late. And yet he knows his conclusion (Warfield, 2005, pp.407-408).

The history of the knowledge from falsehood literature echoes the over-a-half-century earlier history of Gettier puzzles (starting from Gettier, 1963). This is because both literatures (as with so much else in epistemology) were, are—and must be—example-driven. Consequently, there's an inertial tendency to focus on various elements of the specific *examples*, elements that turn out to be inessential—but can only be *seen as* inessential after further examples

without those elements are given. In the Gettier literature, cases of justified belief without knowledge were seen as *essentially* involving inferential missteps or false assumptions. Coupled with the thought that Gettier cases are counter-examples to **Tripartite**, as I mentioned, this invited solutions, along the lines of (Harman, 1973). The profession, as noted earlier, thus failed to see that fake barn cases (Goldman, 1976) are *also* Gettier cases. Bernecker (2022, 1626), for example, describes **Barn** with the jargon, "failed threat case" and says it's "Gettier-like." Rather, fake barn cases are exactly the same epistemic phenomena as the inferential Gettier cases, although they don't involve inference, or properly speaking, premises, but instead a judgment arising directly from sensory experience. This is what's crucial to their being *Gettier* cases: they involve truths justifiably believed by agents but despite this, they fail to be instances of *knowledge*.

Knowledge from falsehood cases (**Handout** and **Temperature**), the study of which is now over 20 years old, have been similarly restricted in their recognized scope. (Warfield 2005, 405), for example, writes:

> … one apparently [can know] something despite the existence of something non-ideal in the epistemic pedigree of the belief in question (examples: faulty reasoning, false premise, problematic testimony).

Notable: There is no mention of the possibility of knowledge from falsehood cases, along the lines of **Barn**—arising, that is, entirely from the senses, without assumptions or inferences. Assuming that knowledge from falsehood cases are restricted in this way has had two unfortunate effects. The first effect is a widespread tendency to explain away the cases, relying on the presence of inference and mistaken assumptions. For example, one claims about **Handout**, that the operative assumption that the speaker draws their conclusion on the basis of isn't that there are 52 people at the talk but rather that there are *about* 52 people at the talk. Then the basis on which the speaker draws the conclusion that they have enough handouts is something true, not something false. Thus this case—and others that are handled in the same way—aren't seen as genuine cases of knowledge from falsehood. (Luzzi, 2019) has a thorough discussion of the literature on this move, called the *proxy premise strategy*, and evaluates it.

The second effect is to give the impression that, in any case, such purported cases of knowledge from falsehood are rare or atypical. But, as it turns out, there are *many* cases of knowledge from falsehood that, like **Barn** in the Gettier tradition, arise directly from the knower's sensory experience without the intercession of assumptions or inference. And, ironically, Hilpinen's early **Temperature** indicates this, although this was overlooked by the subsequent literature. Consider:

287

Astigmatism: I see the color and the (shifting) shape of a rabbit as it hops by. I'm not wearing my glasses that correct my astigmatism, however, and so the rabbit's shape is distorted. Nevertheless, its appearance falls within the range of rabbit-shaped items. Thus, I think: "that's a rabbit," and I know this even though I do so via a false perception of its actual shape.

A striking fact about astigmatism: it's, generally, not phenomenologically accessible: removing and then replacing one's glasses doesn't make visible the distortions in shape that one's astigmatism induces. Thus, there's a striking similarity about the role of inference in both cases. In **Barn**, there's no good reason to think Henry is engaged in inferring from a false assumption. Harman (Harman, 1980), on the contrary, thinks Henry relies on the assumption that it's unlikely his belief is false. This is, in turn, unlikely if only because relying on such an assumption is an act of metacognition; and only under special circumstances do we think *about* the properties of our beliefs as opposed to the contents of our beliefs (what our beliefs are about). Seeing something one (immediately) takes to be a barn is (usually) an ordinary circumstance. Therefore, **Barn** is similar to **Astigmatism** in that there's no good reason to think I'm inferring anything relevant to my astigmatism, for example, "that the shape-distortion isn't significant enough to threaten my knowledge claim." This tight similarity invites the thought that we must find out what it is about our knowledge-gathering methods that explains when misapprehensions, missteps, and false assumptions lead to knowledge and when they don't—when they instead lead to Gettier cases.

Although **Astigmatism** makes clear that knowledge from falsehood can occur on a sensory basis just as Gettier cases can, it may not make clear how importantly widespread—almost everywhere—knowledge from falsehood phenomena are. So consider this next case:

Blurry: I see the color and the shifting shape of a rabbit as it hops by. My eyes (at this moment, because of allergies) are making everything I see *blurry*. Despite the fact that *what I see* isn't the actual shape of the rabbit (because *what I actually see* isn't anything's shape: a blurry smear that's rabbitish), nevertheless, when I think, "that's a rabbit," I *know* that it's a rabbit.

Thus, knowledge from falsehood, at least with respect to sensory information, is absolutely ubiquitous.

Two caveats about **Blurry**. First, the word "see" is treacherous here, as philosophers know. The point is that one can be asked to describe what one "actually sees" (by an optometrist, say): One would understand what was being asked for, and would give a description of

288

something blurry, not the short answer, "a rabbit," unless one were joking.

Second, the blurriness in **Blurry** shouldn't be confused with peripheral blur, as discussed in Section 1. Peripheral blur is (notoriously) phenomenologically invisible to us. We don't detect—we don't see—the drop-off in acuity in the periphery (as compared to the fovea) except by the use of clever devices like Dennett's card trick (Dennett, 1991, pp.53-54):

> Take a deck of playing cards and remove a card face down, so that you do not yet know which it is. Hold it out at the left or the right periphery of your visual field and turn its face to you, being careful to keep looking straight ahead …. You will find that you cannot tell even if it is red or black or a face card.

If we compare Gettier cases with knowledge from falsehood cases, a puzzle emerges. In both cases, there is a false basis on which an agent believes something. But in the Gettier cases, they fail to know what they believe although in the knowledge from falsehood cases, they do know what they believe. What is it about our knowledge-gathering methods that explains when misapprehensions, missteps, and false assumptions lead to knowledge and when they don't? Those who treat *reliability* as what's central to knowledge have an answer to this. I'll discuss an example first, and then generalize on its basis.

Consider **Handout** again. Describe the method of counting that the speaker uses to be one that allows them to overlook audience members who come in late. Call this alternative method *qcounting*: the speaker is qcounting (not counting) audience members. Qcounting, as is easily seen, isn't as reliable as counting, but nevertheless, it's pretty reliable. It's reliable enough, that is, to yield knowledge when it gets the right answer. Contrast this with **Ford**. To draw an inference to something logically weaker than what one started from, when starting from something that's false, is *not* a particularly reliable method for getting to the truth: sometimes doing so yields something true and sometimes it doesn't (unless one independently makes sure to infer only logical truths).

So, if we provide a necessary condition on knowledge, roughly, this way: *Someone knows a proposition if that someone has gained access to that proposition via a sufficiently reliable method*, then, notice, this necessary condition is ecumenical enough to handle knowledge from falsehood. For even if one starts with false assumptions, or engages in inferential missteps, if the process of doing so is sufficiently reliable, then this necessary condition is satisfied. The striking fact is that reliability, as a necessary condition on knowledge, *doesn't* require, for example, impeccable deductive methods. We can have somewhat sloppy methods, but as long as such methods aren't *too* sloppy, *they'll do for establishing knowledge.*

Let's return to **Handout**. We tend to think of Warfield using one method, counting, but making a mistake. As suggested two paragraphs back, a more philosophically illuminating way to describe what's happening is to say that Warfield (unknowingly) has stumbled into using a *different* method, qcounting, which simply isn't as good (isn't as reliable) as counting. One qcounts the members of an audience, instead of counting them, if one overlooks those audience members who arrive late. Looked at this way, any method of knowledge gathering that has been modified by the introduction of one or more mistakes can instead be seen as the use of a *different* knowledge-gathering method.

In these terms, we can describe certain methods we use to learn about the world as more or less *resilient*. Resilient methods are open to being applied mistakenly while still giving us a high possibility of right answers. Methods that aren't resilient don't do this. What follows is an illustration:

Start with a given method for determining whether a proposition is true or not—for example, Warfield's counting the number of people present at a talk to determine if he has enough handouts. Let's understand the method as one of counting without making any mistakes. Consider all those alternative methods that involve "operator-failures" (fumbles, miscountings, etc.), what are normally regarded as mistaken applications of a particular method (as applications, that is, of that method that involve mistakes). These alternative methods fall within the resilience threshold of the original method if such methods are *reliable enough* to yield knowledge when the answers they yield are true. Otherwise, if their reliability is too low, they fall outside the resilience threshold of the original method. Another way to put the point: A misapplied method yields knowledge despite the misapplication *only if* the misapplied method—taken instead as a distinct method of knowledge gathering, in and of itself— is sufficiently reliable.

Let's apply this notion of a resilience threshold to a Gettier case and to a knowledge-from-falsehood case to see how it illuminates why in one case, a fumble nevertheless yields knowledge and in the other case it doesn't.

First consider **Ford** from Section 1. If we try to understand Smith as relying on an alternative method, it can only be something to the effect of: *Infer from something that can be false something that, strictly speaking, is logically weaker.*

But, as noted before, *that's* not a reliable knowledge-gathering method under any description because sometimes doing so yields something true and sometimes it doesn't (unless one independently makes sure to infer only logical truths).

Now reconsider **Handout**. The description of the case introduces (implicitly) the alternative method that Warfield is using,

and it's something to the effect: *Identify rapidly each occupied seat in the auditorium to determine if there are enough handouts, and don't stop to check if anyone new has shown up during the count procedure.* The original method, by contrast, is a method we understand as: *Count all and only the individuals in the auditorium to determine if there are enough handouts.* But in the circumstances where the question of, "Do I have a sufficient number of handouts?" is relative to the possession of a *large* number of handouts, the alternative method is sufficiently reliable.

There are a large number of alternative methods that Warfield could actually be using when making various "mistakes"—that is, there are a large number of ways that Warfield could have fumbled his count. Most of them are sufficiently reliable, given the number of handouts that Warfield has. This is why it's reasonable to speak here of there being a resilience *threshold.* It's also important to note that **Handout** admits versions where Warfield gets the correct answer via mistakes that cancel each other. For example, he could double-count someone, but then an audience member that he'd counted leaves before the beginning of the talk, and he doesn't realize this. He still knows in such cases, if the number of handouts is large enough.

There's a striking contrast here, encapsulated in the difference between **Ford** and **Handout**, that shows up in the distinction between formal methods of proving results in mathematics and the informal methods that professional mathematicians, as well as everyone else, except logicians, employ. Formal methods are the ones studied in logic books, such as (Shoenfield, 1967). "Informal rigorous" mathematics— a mixture of technical mathematics and ordinary language—is the form of mathematics published in professional journals. I discuss this important distinction further in (Azzouni, 2024). Recall the earlier observation that, although there's a resilience threshold around any method of counting (correctly) the number of people in an auditorium—with respect to the question of whether there are a sufficient number of handouts—there doesn't seem to be a corresponding resilience threshold around the drawing of an inference in **Ford**. This corresponds to an unnoticed contrast between many informal mathematical methods of proof and certain formal ones: formal methods are often frail in the sense that errors are fatal; an inferential misstep leaves one with no hope of having gotten anything even *close* to the right result (let alone the right result itself) except by a sheer accident. Many informal mathematical methods, however, have resilience thresholds. One (empirical) indication of this is the notorious fact that published proofs (in professional mathematical journals) often have numerous errors, despite the theorems purportedly being proven nevertheless being right. These errors are usually ironed out only in the secondary textbook literature. Nevertheless the original authors are given credit for discovering (and knowing) the results, and not the textbook authors who actually give *valid* inference pattern of the proofs for the first time.

291

Differences in the resilience thresholds of knowledge-gathering methods arise from the details of the methods themselves. This is worth illustrating with some easy examples. Standard counting errors, for example, leave us close to, or at, the right answer as **Handout** indicates. I skip someone when counting heads or count someone twice: errors like this leave me adjacent to, or at (if errors cancel), the right answer. The same is true of our short-hand methods of adding, e.g.,

$$1^17^16$$
$$+\ \underline{2\ 3\ 4}$$
$$4\ 1\ 0$$

Here we sum the numerals in each column, and "carry" numerals to the next column (when appropriate), where we include them in the sums of the numbers in those columns. Carrying mistakes always leave us close to the right answer. This is *not* true of all informal methods used in mathematics; it's not true, for example of the standard short-cut method of multiplication that's still (I believe) taught to grammar-school children:

$$176$$
$$x\ \underline{234}$$
$$704$$
$$5280$$
$$\underline{35200}$$
$$41,184$$

This is because, in this case, the correct final summands to be added require remembering to suffix-place a sequence of zeros corresponding to the location of the column involved. Do this incorrectly, and the resulting answer can be very far off in magnitude from the right answer. It, thus, shouldn't be assumed that informal mathematical methods are *invariably* more resilient than formal ones; it really does depend on the details of the methods themselves, as well as on what question is being answered (what particular piece of purported knowledge is being pursued).

Here's something that can said in general about knowledge-gathering methods, mistakes, and resilience thresholds: Any method, as a whole, generates the contours of its resilience threshold. That is, encapsulated in the method (as it's used in particular circumstances) is the range of mistakes that are possible using that method. The

specific (kinds of) mistakes, when they're incorporated as part of an alternative method for knowledge gathering, themselves fall within or outside the resilience threshold, depending on how reliable those alternative methods are. Consider the short-cut method of multiplication again. This involves the suffixing of zeros, mistakes about which generate, generally, answers quite far from the right ones. Thus such mistakes can be very likely to generate answers to questions that are wrong (e.g., to the question: Is the number of sound bites under 100,000 or over 20,000?). However, this method of multiplication also allows mistakes that only involve carrying numerals to the next column: the methods that incorporate such "mistakes" may be acceptably reliable, given the question that's being posed.

We turn now to establishing that the results described in Section 1 constitute an example of knowledge from falsehood. Actually, something stronger will be established: the results show that human object recognition involves not only knowledge from falsehood, but, further, knowledge *because of* falsehood.

As I've advertised, what's on offer in this paper isn't merely knowledge from falsehood (or knowledge despite falsehood) but knowledge *because of* falsehood. That is, what's on offer is showing that there are cases where the knowledge in question is unavailable except through a route that involves falsehood. There is no falsehood-free approach to the knowledge in question. The primary example of this, in this paper, is human peripheral blur, although other examples will be alluded to at the end of the paper.

To get a grip on *why* human peripheral blur should be regarded as helping to provide knowledge because of *falsehood*, rather than, say, knowledge because of ignorance (or something like that), let's start by asking what peripheral blur is. Here I can be understood as asking not just a specific question about peripheral blur, but, in fact, one about the blur of all sorts that shows up in human vision—so the kind of blurriness arising in **Blurry** is also relevant. An easy (but wrong) answer is that it's *lack of information* about that part of the visual scene that's been blurred out. On this mistaken view, to blur out (part of) an image is to eradicate information about the visual scene itself corresponding to that (part of) the image.

Although this is certainly neurophysiologically correct in the case of peripheral blur (recall section 1), this isn't phenomenologically correct. On the contrary, as far as the viewer is concerned, blur of all sorts isn't less information—it's *misinformation* about the parts of the visual scene in question. What's blurred, one way or the other, always has *sensory content*—although the specifics of that content aren't always easy to give; peripheral-blur content is easy to *visualize*, although not describe in words. Indeed, see the pictures that accompany Jang and Tong (2024) and Pramod et al. (2022) to see what peripherally-blurred items *look like*.

Regardless, what's key to seeing that blur is misinformation rather than absence of information is that a blurry something is (or can be seen as) conveying something out there that looks *exactly like the blurred image presents it as looking*. That is, imagine that someone can't see a texture pattern on a wall because their watery eyes are making it appear blurry. That very same (watery eyes) texture pattern could be painted on the wall so that someone whose eyes aren't watery would see exactly the same thing.

Caveat. By "sensory content," in the above paragraph, all I mean is this: there is something it looks like. It doesn't look like nothing or like an absence. See the points that follow. This caveat is to officially note that I'm not committing myself to a representational view of vision, or anything like that.

In any case, a general argument can be given that misinformation rather than absence of information is *always* presented to us phenomenologically, when, nevertheless, the sensory experience we're having is due to an absence of sensory information. Question: What could absence of sensory information possibly look like? A hole in space? Well, no. Consider the blind spot. This, neurophysiologically, is information that's *literally* missing. But that's not how it phenomenologically presents; we haven't an impression of something missing. What we see is a smooth surface of a certain sort. That's *misinformation* not absence of information.

A little thought shows that we never experience absence of information. And this is empirically demonstrated by audition with the fact that successively degraded audio recordings of the same thing sound progressively as if there's more and more background *noise*. (They don't sound like something is missing: they sound like noise is running interference with what we can hear.) Similarly, the loss of information in the ear (age-related hearing loss, for example) doesn't sound like an absence of sound; it sounds like, well, intrusive sounds (tinnitus, for example). We also can experience having trouble hearing something: it's not loud or clear enough. But that isn't to experience the presence (as it were) of the absence of information, but instead something that simply isn't loud or clear enough to hear. (A quiet sound isn't experienced as the absence of loudness.)

This completes the case that the role of peripheral blur in human object recognition distorts or falsifies what would otherwise be seen, and that, therefore, what we have is at least a case of knowledge from falsehood.

A stronger claim, however, can be made about this case. This is, as advertised, that the knowledge in question isn't simply from falsehood but *because of* the falsehoods induced—in this case, the sensory falsehoods that peripheral blur introduces. The striking empirical fact that's been recently established is that CNN object recognition is *optimized* by the introduction of peripheral blur that

corresponds to how the human eye sees images—that is (as established in the earlier paragraphs), by the introduction of *misinformation* about the relevant visual environment. It isn't that *less* information optimizes human object-recognition capacity it's that *wrong* information does this. By the considerations of the last few paragraphs, what we have is out-and-out falsehood instead of, say, ignorance. Couple this with the fact noted above that the introduction of peripheral blur optimizes the performance of neural-net visual recognition, and in fact (because clear-trained CNNs catastrophically fail in certain situations although humans and blur-trained CNNs do not), we have knowledge *because of* falsehood. The misinformation induced by peripheral blur is essential, in certain circumstances, to knowing what one is seeing. No knowledge is possible, in these cases, without falsehoods, without distortions.

A last thought

Knowledge from falsehood has introduced something of a philosophical shockwave into the epistemic literature. But next door, in philosophy of mathematics and philosophy of science, knowledge from falsehood and indeed, knowledge because of falsehood are well-known truisms of mathematical/scientific practice. These truisms don't seem to have been officially described this way, in that literature, in terms of their implications for *knowledge*; that is, dramatic implications about how knowledge works have not been drawn in philosophy of mathematics and philosophy of science.

295

The central scientific phenomenon I'll use as an illustration is this: theories, especially complex contemporary ones in physics (quantum mechanics, general relativity) are often intractable when applied to certain domains. What must be done instead is to introduce a falsified user-friendly alternative theory. There are many ways to do this (there are many sorts of idealizations); but many of them share two important characteristics. First, they invariably involve the introduction of falsehood, either in the description of the phenomena the theory is being applied to (simplifications of the geometry of the objects under study, e.g., molecules, or other sorts of physical structures, such as magnetized materials). Second, many of them (although not all) are essential to deriving results that otherwise, as I said, are unavailable because of intractability—either because:

> The phenomena in question are too complex to describe. Sometimes, with physical structures, we aren't in a position to describe them. We haven't the tools to survey what they're *actually like*. Other times, the mathematical complexity explodes if we don't "idealize," simplify shapes for example—knowingly falsify the parameters in question.

> Or,

> because the theories in question are themselves mathematically intractable. Quantum-mechanical theories, for example, are

notoriously intractable. The three-body problem illustrates when mathematical intractability shows up in Newtonian mechanics. Quantum theories are much, much worse.

Nevertheless, the results derived are *true*: Knowledge *because of* falsehood, and not merely knowledge from falsehood (or despite falsehood).

Some final citational points to round out the history of this set of interrelated topics:

For an early discussion of idealizations, from the perspective of their role in avoiding intractability, see (Azzouni 2000). For a survey of idealizations in the sciences (nowadays described as models), see Frigg and Hartmann (2020).

Hilpinen (Hilpinen, 1988, 164) noticed the presence of knowledge from falsehood in the sciences, and cites (Franklin, 1988) as making the point that "incorrect experimental outcomes need not result in incorrect (or unjustified) theory choices." The example he gives: Millikan's mistaken value of the basic unit of elementary charge didn't weaken support for charge quantization. Oddly, Hilpinen suggests that the derived true results tend to be "vague." Not so, I think, no more than in **Handout**. One may use Newtonian physics to calculate the trajectory of a missile instead of general relativity. One yields a false description of the trajectory of the missile (or, if you wish, an approximately true description)—but the exact statement (the missile will hit the target) is nevertheless true, and not vague.

Lastly, it would hardly be a surprise if the other senses didn't operate similarly. In my view, they do. Take audition. There is a lot of evidence that our perception of speech involves something analogous to foveal/peripheral modifications of what is heard. In particular, misleading presentation of sound is involved in our ability to better understand speech, and segment it into phonemes and words. See, for example, (Khalighinejad et al., 2019).

Acknowledgments

References

Azzouni J. Knowledge and reference in empirical science. Routledge; 2000.

Azzouni J. The algorithmic-device view of informal rigorous mathematical proof. In: Sriraman B, editor. Handbook of the history and philosophy of mathematical practice. Volume 3. Springer; 2024. p. 2179-2260. (Appeared online in 2020.)

Bernecker S. Knowledge from falsehood and truth-closeness. Philosophia. 2022;50:1623-38.

Dennett DC. Consciousness explained. Little, Brown and Company; 1991.

DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? Neuron. 2012;72:415-34.

Franklin A. Experiment, theory choice, and the Duhem-Quine problem. In: Batens D, van Bendegem J, editors. Theory and experiment: Proceedings of the 6th joint international conference of history and philosophy of science. Dordrecht; 1988.

Frigg R, Hartmann S. Models in science. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy [Internet]. Spring 2020 Edition. Available from: https://plato.stanford.edu/archives/spr2020/entries/models-science/.

Gettier E. Is justified true belief knowledge? Analysis. 1963;23:121-3.

Goldman A. Discrimination and perceptual knowledge. J Philos. 1976;73(20):771-91.

Harman G. Thought. Princeton University Press; 1973.

Harman G. Reasoning and evidence one does not possess. Midwest Stud Philos. 1980;5:163-82.

Hilpinen R. Knowledge and conditionals. Philos Perspect. 1988;2:157-82.

Jang H, Tong F. Improved modeling of human vision by incorporating robustness to blur in convolutional neural networks. Nat Commun. 2024. doi:10.1038/s41467-024-45679-0.

Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N. Adaptation of the human auditory cortex to changing background noise. Nat Commun. 2019. doi:10.1038/s41467-019-10611-4.

Lehrer K. Knowledge, truth and evidence. Analysis. 1965;25:168-75.

Luzzi F. Knowledge from non-knowledge. Cambridge University Press; 2019.

Pramod RT, Katti H, Arun SP. Human peripheral blur is optimal for object recognition. Vision Res. 2022;2000:108083. doi:10.1016/j.visres.2022.108083.

Saunders JT, Champawat N. Mr. Clark's definition of 'knowledge'. Analysis. 1964;25(1):8-9.

Shoenfield JR. Mathematical logic. Addison-Wesley Publishing Company; 1967.

Warfield TA. Knowledge from falsehood. Philos Perspect. 2005;19:405-16.

297