



## Not rational, but not brutally causal either: A response to Fodor on concept acquisition

*(Ni racional ni toscamente causal tampoco:  
una respuesta a Fodor sobre la adquisición de conceptos)*

Louise ANTONY\*  
University of Massachusetts

**ABSTRACT:** Jerry Fodor has argued that concept acquisition cannot be a psychological or “rational-causal” process, but can only be a “brute-causal” process of acquisition. This position generates the “doorknob → DOORKNOB” problem: why are concepts typically acquired on the basis of experience with items in their extensions? I argue that Fodor’s taxonomy of causal processes needs supplementation, and characterize a third type: what I call “intelligible-causal processes.” Armed with this new category I present what I regard as a better response than Fodor’s to the doorknob → DOORKNOB problem.

**KEYWORDS:** concept, nativism, acquisition, rational-causal, brute-causal, intelligible-causal, representation.

**RESUMEN:** Jerry Fodor ha argüido que la adquisición de conceptos no puede ser un proceso psicológico o «causal-racional», sino que solamente puede ser un proceso de adquisición «causal-tosco». Esta postura da lugar al problema «pomo → POMO»: ¿Por qué los conceptos normalmente se adquieren a partir de experiencias de objetos que pertenecen a sus extensiones? Arguyo que la taxonomía de Fodor de los procesos causales necesita ser complementada, y caracterizo un tercer tipo de procesos a los que denomino «procesos causales-inteligibles». Basándome en esta nueva categoría, ofrezco una respuesta al problema pomo → POMO que, a mi juicio, es mejor que la respuesta de Fodor.

**PALABRAS CLAVE:** concepto, innatismo, adquisición, causal-racional, causal-tosco, causal-inteligible, representación.

\* **Correspondence to:** Louise Antony. Department of Philosophy. University of Massachusetts Amherst. E-305 South College. 150 Hicks Way. Amherst, MA 01003-9274 – lantony@philos.umass.edu – https://orcid.org/0000-0003-3116-4895

**How to cite:** Antony, Louise. (2020). «Not rational, but not brutally causal either: A response to Fodor on concept acquisition»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 35(1), 45-57. (https://doi.org/10.1387/theoria.21031).

Received: 24 July, 2019; Final version: 21 December, 2019.

ISSN 0495-4548 - eISSN 2171-679X / © 2020 UPV/EHU



This article is distributed under the terms of the  
Creative Commons Attribution 4.0 International License

In *The Language of Thought*, Jerry Fodor argued—notoriously—that all (or almost all – a qualification I’ll ignore from hereon) of our concepts are *innate*. The argument went like this:

1. For a concept to be *learned*, the subject must project a hypothesis about the extension of the concept.
2. In order to project a hypothesis about the extension of a concept, the subject must already have the conceptual resources to represent the extension of that concept.
3. Having the conceptual resources to represent the extension of a concept is sufficient for possessing the concept.
4. Therefore, no concept can be learned.
5. Therefore, every concept we possess is unlearned, i.e., *innate*.<sup>1</sup>

This conclusion seemed to many to be crazy; Dennett dubbed the view “Mad Dog Nativism,” a sobriquet Fodor cheerfully accepted.

However, Fodor modified his stand in later work (Fodor, 1998, 2008). While he continued to maintain that no primitive concept could be *learned*, he more recently allowed that some concepts might be *acquired*. This is an important difference, especially in the computationalist tradition. Let me unpack it. Learning is a causal, computational process relating experience to a novel mental state such that, first, there is a *rational* relation between the experience and the content of the subsequent mental state, and second, the state-transitions between the experience and the end mental state mirror rational relations. “Rational relations” here is meant to include the steps of all reasoning processes—deductive inference, but also inductive and abductive reasoning and practical reasoning.

“Acquisition,” however, is non-specific. It covers *any* causal process that begins in experience and eventuates in a new mental content. To say that concepts can be acquired, therefore, is only to say that some experience caused the generation of the concept. There is no requirement that there be any rational—or even merely semantic—relation between the experience and the subsequently added concept. To say that concepts could be acquired, then, even if they couldn’t be learned, Fodor is explicitly allowing that the connection between experience and a concept it engenders could be completely fortuitous—a concept could arise “through surgical implantation; or by swallowing a pill, or by hitting one’s head against a hard surface” (Fodor 2008, p. 135).

But if saying that concepts can be acquired puts no constraints on the way that they can be acquired, a puzzle arises. There seems to be a regularity that goes unexplained: concepts (empirical concepts, at any rate) seem to be acquired on the basis of certain particular kinds of experience, specifically, experiences of their instances. Why should this be? Fodor calls this the “doorknob → DOORKNOB” problem, and proposes to answer it by appeal to a particular kind of concept acquisition that he calls “triggering.”

Empirical concepts, he points out, are often associated with *stereotypes*. Stereotypes, we may suppose, are complex ideas consisting, perhaps, of visual images, general beliefs, and associated other ideas. Although concepts cannot be learned, Fodor says, stereotypes can.

<sup>1</sup> See (Fodor, 1975, Ch. 2, Sec. 3, “What the Private Language Must Be Like,” especially pp. 79-85). Fodor gives a brief summary of this argument in (Fodor, 2008, pp. 129-30).

They can be learned inductively as statistical generalizations over experience. Once a stereotype has been learned, a procedure Fodor calls “locking” takes over—a new concept-type is “locked” onto the kind of which the stereotype *is* the stereotype. Schematically, triggering looks like this:

1. Experience → stereotype formation (P1)
2. P1 → locking = concept attainment (P2)

Fodor allows that the first step, P1, is a learning process, and hence a rational-causal process.<sup>2</sup> However, P2, Fodor tell us, is “some reliable, but not intentional (and hence, *a fortiori*, not inferential) neurological process” (Fodor, 2008, p. 151). Fodor goes on to tell us a little more about P2. It is, he says

...no sort of inference at all. It’s, as it were, a *subintentional* and *subcomputational* process; it’s a kind of thing that our sort of brain tissue just does. Psychology gets you from the initial state to P2; then neurology takes over and gets you the rest of the way to concept attainment... Intentional explanation can’t, in any event, go all the way down on anybody’s story. Sooner or later neurology has to take over. (Fodor 2008, p. 152)

In saying that the process of triggering is “not intentional,” Fodor is saying that the states involved do not possess *intentional content*. For Fodor, this is another way of saying that these states do not fall within the domain of psychology. Psychology, for him, just is the science of intentional states; the posits of psychology are structures and processes defined over states with intentional content. Intentional explanation is explanation in terms of states with intentional content. Given these assumptions, Fodor is committed to there being the following two types of causal process:

1. *Rational-causal (Intentional, Psychological)*
2. *Brute-causal*

This simple taxonomy of causal processes has been more or less presumed in the literature on mental processes, and is shared by both advocates and opponents of the idea that psychological processes involve classical computation over structured representations. But I’m about to argue that this taxonomy is *too* simple—that it needs to be complicated.

In the first place, it’s useful to distinguish between two varieties of rational-causal processes. The distinction is not an architectural one; rather it is semantic. All rational-causal processes mirror rational inferences, and could therefore with justice be called *rule-following* processes. But in some inferences, there is an explicit representation of a rule, while in others there is not. Fodor did not find reason to discuss the notion of rule-following, beyond answering critics of LOT who claimed that he could not reconstruct the intuitive dif-

---

<sup>2</sup> I do not see how Fodor can say this without an account of how brute-causal interactions between the environment and our sense organs become or generate states with intentional content. Once experience is conceptualized, it’s possible to see the inductive construction of a stereotype as an instance of learning. But what about that first step? It’s the ‘doorknob → DOORKNOB’ problem, or, if you will, the ‘round → ROUND’ problem all over again. That is, we need an account of why and how experiences of round things reliably trigger deployment of the concept ROUND.

ference between processes that *followed* rules and processes that merely conformed to rules. His answer appears in a footnote in *The Language of Thought*, and is short enough to be quoted in full:

A [...] chestnut of a question that is supposed to embarrass information flow psychologists goes like this: 'If you are willing to attribute regularities in the *behavior* of organisms to rules that they unconsciously follow, why don't you say (e.g.) that the planets 'follow' Kepler's laws in pursuit of their orbits about the sun?' [...]

It should now be clear how this sort of question is to be dealt with. What distinguishes what organisms do from what the planets do is that a *representation of the rules they follow constitutes one of the causal determinants of their behavior*. So far as we know, however, this is not true of the planets. (Fodor, 1975, p. 74n.)

This characterization certainly serves to characterize the causal processes involved when someone consciously and explicitly follows a rule, as when a beginning chess player recites the rule of castling to himself before making her move. Some un- or sub-conscious processes may also fit this characterization of rational-causal processes as well, perhaps those constituted by automated processes involving learned rules. But Fodor ought to have recognized that, by his own account, there must be processes that do not represent any rules, yet should still be counted as rule-following processes. Any primitive computational process must be of this sort. As Lewis Carroll so wittily pointed out, the rule that an inference is following is not fungible with a premise that expresses that rule—otherwise, an infinite regress ensues.

But if we cannot base the following/conforming-to distinction on whether or not the process involves the representation of a rule, how can we draw the distinction? What is the essence of a rational-causal process?

I suggest that rational-causal processes are distinguished from other kinds of processes in that they exhibit *strong mirroring*. A causal process exhibits strong mirroring just in case that process

- a) involves a set of physical elements and physical states involving those elements, such that the elements at each stage of the process reflect, under intentional characterization, the premises or conclusion of a valid or cogent inference, and
- b) the properties of the physical elements to which the physical mechanism is causally sensitive mirror the logical or semantic properties of the premises and conclusion in virtue of which the mirrored inference is sound or cogent.

According to this definition, a primitive processor, like a logic gate, counts as a rational-causal process because it strongly mirrors a primitive logical rule, even though it need not have any state that is interpretable as a representation of any logical rule. The physical states over which the causal process is defined will mirror primitive logical operations, like conjunction formation, because the physical properties of the states involved in the causal process will mirror the logically relevant features of propositions or truth-values. Thus primitive processors can be built out of many different kinds of physical material, and can involve marks on paper, electronic impulses, or neuronal signals.<sup>3</sup> Complex processors can,

<sup>3</sup> Joseph Levine and I once built a binary adder out of our daughter's fifth-grade class, following the design provided by David Macaulay in his excellent book *The Way Things Work Now* (2016 edition).

of course, be built out of primitive processors by linking the primitive processors in strategic ways.

I thus want to disagree with Fodor about what is necessary for a device's being a rule-following device—it is strong mirroring that's necessary, and not there being an explicit representation of a rule somewhere in the causal etiology of the end state. We can, if we like, distinguish between those rational-causal processes in which some rule is explicitly represented and those in which there is not—in both cases, it will be the process's exhibiting strong mirroring that makes it rational-causal. Whether or not there is a rule among the premises of the inference being performed is a matter that may be of interest to the theorist, but it is of no significance to the machine.

An example might help. Suppose I am in the process of learning to play chess, and I'm considering whether or not to castle. I might rehearse to myself the conditions on castling, note that I have violated one of these conditions, and then conclude that I may not castle. The reasoning would involve an inference that looks something like this:

1. If one has moved one's king, one may not legally castle.
2. I have moved my king.
3. Therefore, I may not legally castle.

Premise (1) articulates (part of) the rule that I, the chess player, am following. The neurological machine that is realizing my thought process (AKA "my brain") is also following a rule, but not the rule of castling. It is following the rule of *modus ponens*. This rule is nowhere represented in the inference itself; rather the causal processes occurring in my brain strongly mirror an inference that obeys that logical rule.

A final preliminary: I should say something about the idea of "representation" that I am presuming. First of all, I am taking "representation" to be interdefined with "intentional content:" a representation or representational state is an entity or state that has intentional content. Secondly, I follow Fodor and others in holding that there are two different ways in which an entity or state can come to have intentional content: the content can be *derivative*,<sup>4</sup> in which case it is conferred by some intentional agent or group of agents, or the content can be *original*. The intentionality of the representational states in digital computers is—at least for now—derivative, while the intentionality of naturally-existing minds must be original. Naturally-existing minds are where the intentionality buck stops. I am therefore committed to there being a naturalistic explanatory reduction of intentionality—that is, to there being an account of how materials that lack intentional content can, in a *constitutive* sense, give rise to material that possesses intentional content.<sup>5</sup> This is too bad, because none of the proposed

---

<sup>4</sup> It is, to say the least, controversial where original intentionality comes from, and it may be misleading to speak of original intentionality as something that an entity or state can "come to have", as if the entity or state lacks original intentionality, and then gains it. Some, but by no means all philosophers of mind who believe in original intentionality think this way. Fodor does not think this way; the possession of original intentionality is something that is coeval with the complicated nomological connections between mind and world that Fodor thinks constitute original intentionality.

<sup>5</sup> Strictly speaking, the thesis that there exists a naturalistic basis for intentionality is independent of there being an *account* of this basis. But skeptics about the first would be silenced if we could produce the second. Psychology is subject to doubts that don't seem to plague physics, even though no one can produce a theory that unifies gravity and quantum mechanics.

naturalistic accounts of intentionality seem to work. What I do think I have an account of, however, is how, in a *causal* sense, a state that lacks intentional content can give rise to a state that possesses intentional content. That's a bit of progress.

Understanding "rational-causal" in the way I've suggested allows us to differentiate rational-causal processes from brute-causal processes, and hence to distinguish learning processes from processes of mere acquisition in the way Fodor wants. So consider the following pair of cases:

- A. The parent of a toddler points to a doorknob and says to the toddler "That's a doorknob." That causes the toddler to token the concept DOORKNOB.
- B. A toddler bumps their head on the top bunk of their bed. That jiggles the brain of the toddler in such a way as to cause the toddler to token the concept DOORKNOB.

Case A is a paradigmatic case of concept *learning*. There is a rational causal process (or so we may suppose) that links the toddler's hearing and understanding the parent's (reliable) assertion that the thing they are looking at is a thing called a "doorknob," and generates (as we may suppose) a concept—a mental lexeme—dedicated to representing things like it, i.e., doorknobs.<sup>6</sup> Case B is a good example of what Fodor would disparagingly call "mere acquisition"—there is no reason under the sun why a bump on the head should trigger the formation of the concept DOORKNOB—the connection between experience and concept is utterly fortuitous.

But it's important to see that the crucial work here is being done by the requirement that there be strong mirroring, and not just by the requirement that there be a rational relation between the intentional content of the cause and the intentional content of the effect. The states of a brute-causal process may well lack original intentionality, but even if we were to impose upon them some derivative representational contents, so as to make it appear that the process is a rational one, the causal transitions involving them will not mirror any inferential process. For we must remember that strong mirroring requires more than there being some interpretation of states such that the causal sequence of those states *looks like* a rational inference when viewed under intentional interpretation; it must be that the physical properties of the state to which the process is sensitive themselves mirror the logical/intentional properties of the inference.<sup>7</sup>

To see this, consider another case where we would want to make a distinction between two types of acquisition.

- C. A toddler bumps their head on the top bunk of their bed. Perception of the bump causes a (first) tokening of the concept BUMP.
- D. A toddler bumps their head on the top bunk of their bed. That jiggles the brain of the toddler in such a way as to cause a (first) tokening of the concept BUMP.

Case (C) is an example of what we imagine to be the usual case of concept acquisition through experience. Case (D), on the other hand, is anomalous. Yet the relation between

<sup>6</sup> Many details are here suppressed. See (Margolis, 1999) for a plausible story that fills these in.

<sup>7</sup> Thus we cannot, as John Searle insisted we could, simply "treat" any old state of affairs as the realization of a computational device (Searle, 1980). The conditions for strong mirroring—satisfaction of which I would take to be necessary for something's really being a computational device/being—cannot be met by stipulation.

the “content” of the experience and the content of the concept acquired because of the experience is the same in both cases. If we can speak of the *experience* as having content at all, the content of the experience is the same in both cases. The difference, if there is one, lies in the *way* in which the new concept arises, given the experience. Intuitively, we want to say that in Case (C), the concept is somehow *based on* the experience that caused it, while this is not so in Case (D).

Now we have to agree with Fodor that not even properly based concept acquisition can be a rational-causal process. (I want to say the same thing about properly based perceptual *belief*, which is why I’m doubtful about the P1 step in Fodor’s account of triggering, I’ll say more about this below.) The reason goes back to Fodor’s original argument for Mad Dog Nativism, and involves our taking seriously the idea bruited above that we can meaningfully speak of “the content” of an experience. For the toddler (or anyone) to *learn* that a certain kind of experience is called BUMP is for them to draw a rational inference from an experiential state with intentional content to a state that involves the tokening of a concept that characterizes that sort of experience. But what could the intentional content of this experience *be*? If the experience itself involves the deployment of the concept BUMP, then one would have to already possess the concept that is supposedly being acquired.<sup>8</sup>

To distinguish the first case from the second, then, we need to say something about *how* the causal processes eventuated in the acquisition of a concept, and why one way (the way it happened in Case (C)) is epistemically good, and the other way (Case (D)) is not. Here I think it will be helpful to bring in some points from the epistemological debate about internalism and externalism about perceptual justification.

Externalists about perceptual knowledge hold that, as long as the right sort of relation holds between a perceptual belief and the state of affairs the belief represents, the belief counts as knowledge. Crucially, the externalist requires nothing of the subject—they do not have to be aware of or understand *what* the relation is between their belief and the state of affairs it represents, much less understand how that relation is forged or maintained. A common form of externalism is *reliabilism*, the view that knowledge consists in reliably formed belief. Normal perception is held to be a reliable process *par excellence*: certain states of affairs reliably give rise to certain sorts of perceptual experience, which in turn, reliably give rise to perceptual beliefs that such states of affairs obtain. Internalists object—they maintain that there are conditions internal to the subject that must be satisfied in order for the subject to be properly said to know. Many internalists say that these conditions must involve some sort of conscious awareness on the part of the subject—in the case of perception, some kind of rational recognition that an experience of a certain sort warrants a belief with a certain content.

The internalist’s objection to externalism is that without there being some kind of rational connection between perceptual state and subsequent perceptual belief, the citing of a perceptual encounter falls far short of justification—it is rather, as some philosophers have

<sup>8</sup> Of course, in my parade case of concept *learning*, I posited, in the toddler, both a full-blown intentional state (the state of the toddler’s thinking <MY PARENT SAID THAT THAT WAS CALLED A “BUMP”>) and conscious awareness of an inference (<MY PARENT IS PRETTY RELIABLE ABOUT WHAT THINGS ARE CALLED, SO I’LL HERETOFORE THINK OF THINGS LIKE THAT AS BUMPS>). But concept acquisition—triggering in the good cases—doesn’t involve either causation by a fully intentional state, or causal processes that strongly mirror rational inferences.

put it, a “mere exculpation” (McDowell, 1994, p. 20). My gloss on this somewhat puzzling statement is this. Given natural law and antecedent circumstances, the subject *couldn't help* but form the belief, any more than a window can help but break when hit by an errant baseball. The perceptual experience can explain *causally* why the subsequent belief occurs, but causes do not in general *justify* their effects—it's a kind of category mistake to think that they do.

Although I want to agree with the externalists that perceptual justification doesn't—probably can't, on pain of infinite regress—require any awareness of the circumstances that justify perceptual belief, I also think that the internalists have a point. In the terms of our current discussion, I want to say that brute-causal processes can only provide exculpation, when what we want is a type of process that somehow constitutes justification. Otherwise, we won't be able to distinguish the *perceptually based* belief from beliefs fortuitously caused by a perceptual experience. Ditto, *mutatis mutandis*, for properly based concept acquisition from acquisition that merely happens to involve perception. My suggestion is that we can do this if we distinguish causal processes that *utilize information* from those that do not. This means adding a third category to our taxonomy of causal processes—a category I'll call *intelligible-causal* processes.

To explain what I have in mind here, I want first to introduce the notion of a *Dretskean* representation. The root notion here is the notion of carrying information, where “information” is understood in the information-theoretic sense that Fred Dretske appealed to in developing his theory of mind and knowledge (Dretske, 1981, 1988). In the strict sense, a signal *s* carries the information that *p* just in case the probability of *p*'s being the case, given the occurrence of *s*, is equal to 1. But this is an idealization. In reality, there is virtually no signal that carries information in this strict sense, unless we allow that informational content is always massively disjunctive. To get a usable notion of informational content, we can lower the probability threshold, effectively setting aside extremely unlikely causes of everyday events. Thus we can say that the pattern of rings on a tree stump “carries information” about the age of the tree it was once part of, despite the fact that the pattern could have been caused by things other than the tree's having lived a certain number of years. It could have been caused, for example, by an extremely skilled special-effects artist or by even by mere chance. Still, neither of these is very likely, particularly if we expand the modulus of stimulation to include witnessing the felling of the tree. So let's set the probability threshold somewhat lower than but still pretty close to 1. In this somewhat relaxed sense, then, I claim that perceptual states are Dretskean representations, and carry information about their causes.<sup>9</sup>

Let me, by contrast, characterize what I call *Gricean* representations. Gricean representations are so-called because they possess what Paul Grice called “non-natural meaning” (Grice, 1957). (Dretskean representations possess what Grice called “natural” meaning.) For Gricean representation *R* has non-natural meaning *m* only if it's not the case that an occurrence of *R* entails that *m*. Thus, the bus conductor's ringing the bell three times *non-*

<sup>9</sup> I don't invoke this relaxed sense of “information” because I shrink from the consequence that otherwise I'd have to say that the informational content of a perceptual state is disjunctive, making all perceptual states veridical, and perceptual illusion impossible. I'm prepared to bite that bullet. See (Antony, 2011).



*naturally* means that (as it might be) that the bus is approaching the Finchley Road stop, because it's possible for the conductor to ring the bell three times without the bus's approaching the Finchley Road stop. And it is—the conductor may have made a mistake, or may be deliberately trying to deceive the passengers. Thus, it's criterial of a thing's possessing Gricean meaning (in my sense) that it *not* carry information about its causes. That is, it must be possible for a Gricean representation to occur in the absence of any of its causes. It will turn out that *concepts* are Gricean representations.

I want to apply this distinction, between Dretskean and Gricean representations, to the distinction, as Fodor draws it (Fodor, 2007) between *non-conceptual* and *conceptual* representation. The basic idea is that non-conceptual representation is picture-like—“iconic”—and conceptual representation is language-like—“discursive.” Fodor argues that perceptual representation, at least in its early stages, is non-conceptual, and hence iconic, and that representation in thought is conceptual, and hence discursive. I want to say—and Fodor absolutely did not agree with me about this [personal conversation]—non-conceptual representations are Dretskean, and that only conceptual representations are Gricean. The task of explaining concept acquisition is then the task of explaining how Dretskean representations with a certain informational content give rise to Gricean representations which render that content in a conceptual form, in a non-fortuitous way.

Dretskean representations often *encode* some of the information that they carry. By this I mean simply that there is often some regular correspondence between, on the one hand, physical features of the information-bearing state, and on the other, aspects of the information that the state carries. In the case of the tree stump, for example, every set of alternating dark and light circles on the stump corresponds to a single year of growth. When an information-bearing state thus encodes information that it carries, an observer who knows the code can extract the information from the physical character of the state. Dretskean representations are, in this sense, *interpretable*. This fact does not make the informational content of a Dretskean representation observer-relative, or non-objective in any way. And the fact that a Dretskean representation *is* interpretable is, similarly, fully objective. That information *can* be extracted from an information-bearing state remains true whether or not it is actually extracted.

Since encoding is a matter of the presence of certain physical features, we can imagine physical devices that are sensitive to the particular physical features of the Dretskean representation that realize the encoding. Consider our tree stump again. Given that the pattern of light and dark rings effects the encoding of the age of the tree when it was felled, if we can find a physical device that is sensitive to patterns of light and dark, we can use it to calculate the age of the tree. And of course, we *can* find such a device—the human eye! If we are enterprising and clever, we might be able to build an artificial device that could, without our attending to it, calculate the age of the tree. Perhaps—and I'm no doubt betraying here my ignorance of many things—we could rig up a tiny wheeled cart equipped with an extremely sensitive thermometer, which is in turn hooked up to a register. The idea is that as the cart is wheeled over the surface of the tree stump, the thermometer detects the very slight temperature gradients of the dark and light rings; every other time a gradient is crossed, the register clicks over. The cart will stop at the center of the stump. At the end of the exercise, the register will display a number of clicks corresponding to the number of rings, which in turn corresponds to the number of years the tree is alive.

I'm sure something like this could work; at least pretend with me that it does. What would we have if it did? We would have a computational device that, when applied to the right sort of stimulus (i.e., an information-bearing state of a specific kind), generates a representation of the age of a tree. But our device does not do this by strongly mirroring any inference. The intermediate states of the tree-ring counter do not correspond to the premises of any argument, nor are the state changes in the tree-ring counter sensitive to properties that encode steps in any inference.<sup>10</sup> Still, it is no *accident* that there is a reliable correspondence between the age of a tree, and the number of clicks on the register of a tree-ring counter that is run over the tree's stump. We, the inventor-observer, built the little thing in order to exploit what *we* can recognize as an evidential relation between the number of rings on the stump and the age of the tree.

Because of the sensitivity of the tree-ring counter to the physical parameters that encode information about the age of the tree, I want to say that the device *utilized* the information inherent in the pattern of tree rings, in a very real sense of "utilized." We could say, looking at the whole setup from the outside, that the register count is *based on* information about the age of the tree.

With this example as an illustration, let me now characterize the third kind of causal process, the one I want to add to the standard taxonomy: I call it *intelligible causation*. The causal relation, in my example, is intelligible in the following sense: the causal relation between the tree stump surface and the end state of the register is such that the information encoded in the tree rings, *if specified discursively*, would provide an evidential basis for the content we ascribe to the register count. But it's not that we simply choose to regard the cause and effect states in this way—unlike the cases of brain jiggling discussed earlier—it is not an accident that the tree stump pattern causes the tree-ringer counter to register the number that it does. It is not fortuitous that this causal process produced the result that it did.

Intelligible-causal processes, then are processes that resemble rational-causal processes in that they involve input-output relations that can be usefully characterized in intentional terms, but they differ from rational-causal processes in that they do not strongly mirror any valid or cogent inference from the intentional content ascribed to the input state to the intentional content ascribed to the end state. They resemble brute-causal processes, then, in that they do not strong mirror an inference, but they differ from brute-causal processes in

<sup>10</sup> It may be true that the intermediate stages of the operating device mirror lines in a *counting* sequence. But counting is not inferring, and the mirroring of a counting sequence is not a classical inference. When we count, we do not *infer* the next number in the sequence from the previous number—"7" does not implicitly *follow from* "6." We can imagine an inference that goes like this: "1) there are six things. 2) if there is one more thing added, then there are seven things. 3) One more is added. Therefore, (4), there are seven things." But our tree-ring counter is clearly not strongly mirroring *that* inference, since it does not represent either premise (2) or premise (3). Neither does it *follow* the rule given in premise (2), because the counter is only sensitive to local information about the particular rings it is hovering over, and thus does implicitly represent any specific numbers. Ah—you may think—but maybe the rule that the device is following is more general than the rule I've given; maybe it's something like this: "if there are  $n$  things, and one more is added, then then are  $n + 1$  things." But our device can't be following that rule, because in order to do so, it would have to have the capacity to perform arbitrary acts of adding one. And it can't do that.

that the fact that the connection between the input and output states considered under intentional interpretation is not fortuitous. It is not an accident that the causal process sustains intentional characterization. Since they are neither rational-causal nor brute processes, I say that intelligible-causal processes *weakly mirror* valid or cogent inferences. Weak mirroring obtains if and only if:

1. the input to the process, *construed under appropriate intentional interpretation*, bears a rational relation to the output, similarly construed. An intentional interpretation is appropriate iff
  - a) the input state is a Dretskean representation (and thus carries information)
  - b) the intentional interpretation of the Dretskean representation gives one correct specification of (part or all of) the informational content of the Dretskean representation.
2. The causal process that links input and output exploits the physical properties of the input that encode the information carried by the Dretskean representation.

Three notes: First, the fanciful (or “ridiculous”) example of an intelligible-causal process that I described above involves an artifact. The tree-ring counter was designed to exploit the relation between the number of rings on the stump and the age of the tree it came from. But this should not make you think that intelligible-causal processes require deliberate agency or design. Such processes can—and do—arise naturally. Second, the definition of “weak mirroring” says nothing about the status of the output state. I want to allow that that can be itself a Dretskean representation, but that it can also be a Gricean representation. Either way, the content of the output state will constrain the “appropriate intentional interpretation of the process.” If the output state is Dretskean, the interpretation that makes the process intelligible will have to be a correct specification of part or all of the informational content. Typically, of course, the informational content will be some part of the informational content of the input state. If the output state is Gricean, the intentional content of that state will be determined by the relation that determines Gricean content. In the case of the tree-ring counter, the content of the output state will be derived content—a number—which it has because we interpret it as expressing a number. But here again, on the presumption that *original* intentionality is constituted by some natural, non-intentional relation, the content of the output state might be fixed independently of the intentions, interpretations, or interests of the observer. “Appropriate intentional interpretation” must then respect the objective content of the Gricean representation that is the output of the intelligible-causal process. Third, the entire process is *not* psychological. Psychological processes (and Fodor would agree with this) relate only things that both have genuinely intentional content—in my parlance, they relate only Gricean vehicles to Gricean vehicles.<sup>11</sup>

So, at long last, here’s what I have to say about concept “triggering.” When an empirical concept is acquired on the basis on its instances, what happens is that (a) a perceptual

---

<sup>11</sup> I’ve only given atomic concepts as examples of Gricean vehicles, but I am presuming, as is standard, that atomic concepts, if composed properly, can combine to form complex Gricean vehicles that have propositional content.

encounter with an instance of the concept produces a Dretskean representation that carries information to the effect that a thing of a certain kind, *K*, is before you; and then (b) that this input *intelligibly causes* the tokening of a *Gricean* representation that has as its content ‘BEING *K*’<sup>12</sup>. There are two distinguishable ways, then, that a toddler’s bumping his head could result in his acquiring the concept BUMP. On one way, it’s simply an accident—a brute-causal process—that the toddler tokens a concept that (because of satisfying the naturalistic conditions sufficient for representing bumps) means BUMP. On the other way, the head bump initiates an intelligible-causal process that eventuates in a tokening of the concept BUMP. Now we have the needed distinctions, without having to suppose that concept *learning* occurs, but also without having to stigmatize normal concept acquisition as “merely brutally-causal.”

Recall the two-part account of triggering that Fodor offers in *LOT 2*:

1. Experience → stereotype formation (P1)
2. P1 → locking = concept attainment (P2)

What I want to say now is that Fodor might be wrong (depending on how the empirical details turn out) about the character of both P1 and P2. Neither is psychological/rational—causal, and neither is brute-causal either. Both are intelligible-causal. P1 might turn out to be an intelligible-causal process that extracts statistical regularities from perceptual encounters with, for example, horses, and builds a Dretskean representation that encodes these regularities.<sup>13</sup> Then, a separate process takes this constructed representation and generates a Gricean representation – a concept.

SPECIAL BONUS MATERIAL!!! I think that what I have to say about the difference between Dretskean and Gricean representation nicely complements Fodor’s way of drawing the non-conceptual/conceptual distinction, and actually explains the structural distinction between representations that satisfy the picture principle (that is, that every part of a picture represents a part of what’s represented) and those that don’t (i.e., representations that fail to respect the picture principle because they have a canonical representation). I don’t have space here to explain all that here, but details are forthcoming.

## REFERENCES

- Antony, L. (2011). The openness of illusions. *Philosophical Issues* 21(1), 25-44.
- Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: The MIT Press.
- Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: The MIT Press.
- Fodor, J. A. (1975). *The language of thought*. New York: Thomas Y. Crowell Company.
- Fodor, J. A. (1998). *Concepts*. Oxford: Oxford University Press.
- Fodor, J. A. (2007). The revenge of the given. In B. P. McLaughlin & J. Cohen (Eds.). *Contemporary debates in philosophy of mind* (pp. 105-116). Oxford: Blackwell.
- Fodor, J. A. (2008). *LOT 2*. Oxford: Clarendon/Oxford University Press.
- Grice, H. P. (1957). Meaning. *Philosophical Review* 66 (3): 377-388.

<sup>12</sup> This expression should be in corner quotes, but I don’t have the font to do it.

<sup>13</sup> Adam Lowet, Chaz Firestone and Brian Scholl have proposed an acquisition mechanism that has this character. See (Lowet et al, 2018).

- Lowet, A., Firestone, C., & Scholl, B. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics* 80(5): 1278-1289
- Macaulay, D. (2016). *The way things work now* (revised edition). Boston, MA: Houghton Mifflin Harcourt.
- Margolis, E. (1999). How to acquire a concept. In E. Margolis & S. Laurence (Eds.). *Concepts: Core readings* (pp. 549-567). Cambridge, MA: Blackwell.
- McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences* 3(3): 417-457.

**LOUISE ANTONY** is Professor of Philosophy at the University of Massachusetts Amherst. She is the author of many articles in the philosophy of mind, philosophy of cognitive science, feminist theory, and philosophy of religion. She recently edited *Philosophers Without Gods: Meditations on Atheism and the Secular Life* (Oxford University Press). A volume of her essays, *Only Natural*, will be published soon by OUP. Antony has served as President of the Eastern Division of the American Philosophical Association, and as President of the Society for Philosophy and Psychology. She was the co-founder and co-director of the Mentoring Program for Early-Career Women in Philosophy.

**ADDRESS:** Department of Philosophy. University of Massachusetts Amherst. E-305 South College. 150 Hicks Way. Amherst, MA 01003-9274. lantony@philos.umass.edu