Agent-oriented Metaphysics of the Manifest Image
or
The Evolution of Categories

Christopher von Bülow*

Doktorandenkolloquium Leon Horsten & Carolin Antos
Konstanz, February 4, 2022

Abstract

This is a brief description of my constructive project in metaphysics: a more contentful and informative alternative to the standard way analytical metaphysics is done, e.g., by David Lewis and D. M. Armstrong.

The ways in which agents can react to changing circumstances differ in their degree of flexibility and in the dimensions in which they are flexible. These different degrees and dimensions of flexibility correspond to different kinds of information processing (or different 'cognitive data types'), and these in turn correspond to different categories of (stipulated) entities. In each case, aspects of the world which are actually different in their physical details are treated as (manifestations of) the same entity. Thus entities are stipulated: first simple states of affairs, then variously parametrized states of affairs (e.g., graded, directional, localized), then stuffs and kinds of objects as special cases of localized states of affairs, then individual occurrences of stuffs and individual objects.

Localization depends on an area's being salient, by structurally contrasting with its surroundings and being roughly connected. I call such (vague) areas "pockets of contrasting structure". By tracking a *sequence* of these pockets where the boundary of contrast 'moves' relatively continuously and the relevant structure changes only slowly – a 'sequence of pockets with relatively stable structure' – and treating its members as manifestations (or 'stages') of a single entity, agents stipulate persisting things of various categories. The category depends on what sort of structure is considered relevant: for stuff occurrences it is their material or microstructure, for objects also their form or macrostructure, for intentional systems their beliefs and desires, whereas for places it is merely their spatial relations to landmarks in the environment.

Thus each category of thing has its own 'continuity criterion', which determines whether a sequence of pockets can be considered as corresponding to a thing of that category. Thereby it also determines at which (vague) points in time a thing begins and ends, which of several candidate continuations of a sequence, if any, is the right one, and what would count as its continuation if circumstances were (or had been) different.

*Christopher.von.Buelow@uni-konstanz.de; https://www.philosophie.uni-konstanz.de/fachbereich/personen/christopher-von-buelow/

1 Introduction

Hi, thanks to Leon and Carolin for letting me speak here, and thanks to you for coming; I hope it's going to be worth your while. I'm going to talk about metaphysics; in particular, I want to present a proposal for how to *do* metaphysics.

I propose doing metaphysics by asking why agents like us use the categories we use, and my answer is that much of it is due to evolution, more precisely, to the evolution of information-processing abilities: first comes the ability to react to simple states of affairs, then this ability is refined until agents can react to instantiations of kinds and then, even later, to individuals.

I will also argue that these states of affairs and kind-instances and individuals aren't perfectly real, that they are just stipulated by our brains. Nevertheless they are real enough for most purposes; they just don't satisfy the philosophical desire for precision.

2 Good and bad analytic metaphysics

The reason why I propose a new way of doing metaphysics is because I am dissatisfied with the usual way metaphysics is done in analytic philosophy. For me, paradigmatic representatives of the standard method are David Lewis and David Armstrong; and everything in the textbooks on analytic metaphysics I know is also in what I call the standard method.

So, that is what I oppose. A large part of my Ph.D. thesis is concerned with why I think that this general approach doesn't work – but that is not the topic today.

There is a relatively recent alternative to the standard method, proposed by James Ladyman and Don Ross in their book *Every Thing Must Go*.[1] They propose that metaphysicians should refer to science, but not to watered-down popularized versions, but to actual, state-of-the-art science, especially physics. To understand the subjects of the special sciences, they use Daniel Dennett's concept of 'real patterns'.[2]

And Dennett, with his idea of different stances – physical stance, design stance and intentional stance[3] – has also been an important inspiration for me.

So, this is where my sympathies lie; just to get you oriented for today.

3 Scientific vs. manifest image

Now, I understand Ladyman and Ross as saying that the true basic structure of the world – what metaphysics should be about – is described by fundamental physics (i.e., quantum field theory and the theory of relativity). (Actually, they do not say that, but what they really say was too difficult for me.) I assume that this physicalist picture of the world is what Wilfrid Sellars called "the scientific image". However, I don't know nearly enough about physics to say anything useful regarding the scientific image.

But I have some ideas about what Sellars called "the manifest image": which is just our everyday worldview. For example, we see material objects of various kinds

[1] James Ladyman and Don Ross with David Spurrett and John Collier: *Every Thing Must Go: Metaphysics Naturalized*, Oxford/New York: Oxford University Press 2007.

[2] Daniel C. Dennett: "Real Patterns", *Journal of Philosophy* 88 (1): 27–51 (January 1991).

[3] "Intentional Systems", *Journal of Philosophy* 68 (4): 87–106 (February 25, 1971).

having various properties and relations, and we also 'see' other kinds of entities. So, I want to propose a way of understanding these everyday categories, and of explaining why we have them.

I do not primarily aim to convince you of my ideas. The prime goal is to get my main ideas across at all. So, if you want to prove to me where I'm wrong, please wait until the discussion.

4 Our categories reflect types of information processing

Why do we have concepts like "material object"? I simply assume that these concepts do not just mirror the basic structure of the world. So I assume that, say, ordinary material objects are neither metaphysical building blocks of the world nor straightforward constructions from them. – I won't argue for this point here.

Assuming that the categories of the manifest image are *not* the actual basic categories of the world, then why do we see the world through these glasses? I suppose many philosophers would answer, "Because of our language." But this answer begs the question why our languages are built this way. Are our basic categories just accidents of grammar? I don't think so.

Also, I believe that many animals have the same basic categories we do; they 'parse' the world along much the same lines as we do, even though they don't use language: dogs also see the world as food and people and bones and trees and other dogs. Of course they don't see money or air pollution, but little children don't do that either.

So, this is another thing I am not going to argue for: that many nonlinguistic animals also conceive the world as carved into stuffs, like milk and water; into material objects having properties; and into states of affairs being the case or not (for example, door-open vs. door-closed). If this point about animal categories is right then it reinforces the claim that our basic categories are not due to language.

If we do not have these categories because of language then why do we have them? – Because they correspond to the types of information processing that are easiest to engineer and thus come earliest in evolution. This may sound a bit cryptic, but I can't explain it very well in the abstract. An abstract explanation would probably be incomprehensible anyway. Therefore I'll try to illustrate it using a series of examples.

5 Information processing serves agency

So, I suggest that categories have to do with information processing. In my view, information and information processing are primarily about how to act. The connection may not be so obvious, because we associate information processing with computers, and computers usually just produce symbols from other symbols, not actions from perceptions. But maybe we can agree that the beginning of information processing was with agents having to find out what to do when, depending on their environment.

So let's look at some simple agents and the different ways in which they process information. For this purpose, I have brought with me a very simple agent. [I put a mousetrap on the table.]

Is a mousetrap an agent?? Yes – in the sense that it can do something on the right occasion.

But isn't that just cause-and-effect, without any thinking or rationality behind it? Yes, but for me it is sufficient for 'agency' that a mousetrap has a function, a purpose. Because then its behavior can be interpreted as an attempt to fulfill that purpose – as an action.

This weak concept of agency may not be your one true concept of agency, but it is the one I use here: if a thing can be fruitfully considered from Daniel Dennett's design stance and as being designed to do something sometimes, then I view it as an agent, and I view its function-serving or goal-directed behavior as its actions.

6 Re agency: Dennett's stances (capture real patterns)

Briefly, what are Dennett's stances? They are his way of explaining certain metaphysically contentious phenomena. The idea is that certain phenomena, like the beliefs and desires of organisms, and the purposes of artifacts and organs, can neither be reduced to something else (as in physicalism) nor are they metaphysical additions to the physical world (as in dualism). Rather, they are much like useful fictions (as in instrumentalism), but nevertheless they capture real patterns.

What are 'real patterns'? An example: Is this tabletop rectangular? For many practical purposes it is all right and even very useful to treat the tabletop as rectangular. But it is not *perfectly, 100 % exactly* rectangular (down to the microscopic level); so, in that perfectionist sense, it is not rectangular. Its rectangularity is not perfectly real, but real enough for government work. It's a real pattern.

Back to the stances. How do Dennett's stances work? Let's look just at the design stance; it's a bit simpler than the intentional stance. The idea is that you decide (or your brain decides) to interpret a given physical thing or system as something that has a function, that is to say, as something that has been designed for something, is there for something. Then you can say, "Let's assume that the whole system is there to do this, and this part is there for doing that," and so on. You can then use these ascriptions to predict what the system is going to do under various circumstances – assuming that it is well designed.

If you have hit upon a good interpretation, then it will be much easier to predict and manipulate the system than it would be just relying on its physical properties. (That would be the physical stance.) If you have hit upon a good interpretation, your function ascriptions have captured a real pattern: the system and its parts really have those functions (in a good-enough sense) – even though there may be other function ascriptions that work just as well.

Compare this with the tabletop: it really is rectangular, in an imperfect sense, but for some purposes it may be useful to consider its shape in more detail, where it isn't rectangular anymore.

The intentional stance works analogously, only now you interpret the system as an agent and ascribe to it certain beliefs and desires, and work under the supposition that this agent is more or less rational.

In some sense, the intentional stance is a special case of the design stance: If the functional system you are dealing with can *do* something – that is, if it can change states to adapt to different circumstances – then you can treat those states as representing the system's beliefs, and you can treat its function or purpose as its main goal or desire. That's why I can treat a mousetrap (a designed system) as a

primitive agent (an intentional system): it *wants* to kill mice, and when something triggers the trip [I hold a pen close to the armed mousetrap] it *believes* there is a mouse there and *acts* accordingly [I let the trap snap shut]: it clamps shut in order to kill that mouse.

So, even simple mechanisms are agents in my sense, and of course persons and other organisms are agents, too.

7 Categories correspond to types of representing variables

Back to information processing and agency. Another example of a very simple agent is an automatic door opener controlled by a motion detector: If it detects movement in its range it believes that someone wants to pass and accordingly opens the door. If there is no more movement in its range, it believes that no one wants to pass and therefore closes the door again. And so on.

The mousetrap and the door opener instantiate a very simple kind of information processing: their behavior is just 'either–or'; the only variation is between doing one thing or doing the other: open or close; slam shut, or wait and do nothing. The only information they take from the environment is whether to do the one or the other.

Now think about a thermostat in a room. As long as it's warm enough, the thermostat does nothing. When it gets too cold, however, the thermostat switches on the heating. If all it can do is switch on or off the heating, then it would be just like the mousetrap and the door opener: do one thing or the other. But suppose the thermostat can turn on the heating power to different degrees. This would of course depend on how cold the room is. So there is a gradual variation in its behavior, which depends on gradual variation in the *reason* for the behavior.

While the door opener's reaction is 'either–or,' the thermostat's reaction is 'more or less' – or nothing, when the room is warm enough. So the thermostat's behavior is slightly more flexible than the door opener's; it instantiates a slightly more complex kind of information processing.

We can also describe this in terms of data types. In all three cases (the mousetrap, the door opener and the thermostat), there is a kind of functional variation [in a double sense: the behavior is a function of the input, and the variation is crucial to the way the agent functions] between the input and the behavior. Some of the variation in the input, in the stimuli, is transformed into appropriate variation in the reaction. So there is some (designed) variability in what goes on in between.

If we had to write a computer program to control the agent's behavior – that is, to do this sort of information processing – we would make the output depend on a *variable*, which represents the variation in the input in order to transform it into corresponding variation in the reaction. [The variable *represents* the reason for acting by adapting the reaction to it.]

If we view the functional variation between input and reaction as mediated by a variable, then the thermostat would need a more complex variable than the door opener or the mousetrap. The variation in the latter is merely binary, boolean: yes or no, open or close. But the variation in the *thermostat's* behavior is a matter of *degree*, it is like an interval in the real numbers.

In computer science, these different kinds of variables would be called different "data types". Different kinds of computer variables can take different sorts of values:

variables of type "integer" can take integer numbers, variables of type "real" can take real numbers, variables of type "boolean" can take the values "true" and "false". "Integer", "real" and "boolean" are simple data types, types of information that a computer program can work with.

There are also more complicated data types: a variable might hold *arrays* of numbers, that is, tuples or matrices of numbers, or it might hold *sets* of numbers, or it might hold a 'record', a short collection, of certain heterogeneous data, for example, a person's name, their address and their birthday. And then you can combine lots of similar records in a still more complex data type.

So, data types can be of various more or less simple types, or they can be arbitrarily complex combinations of other data types. The variable that describes the behavior of the mousetrap and the door opener is clearly of data type "boolean", whereas the data type of the thermostat's variable is rather something like "real", for example, a number between zero and one.

Now, at long last, some of the metaphysics you have all been waiting for. I think these variables controlling the behavior of agents do not just represent abstract truth values "yes" and "no", or abstract degrees like 0.6; they represent something in the world. Depending on their data type, they represent things of different categories. [To represent something, the behavior must discriminate it, must take it into account, must translate it into a (systematic/designed) difference in the output.]

8 Simple and parametrized states of affairs

In the case of the boolean variables, that is, for the mousetrap and the door opener, it seems pretty clear to me that what is represented here are not mice or moving people but rather states of affairs: "Mouse present" versus "No mouse", in the one case, "Someone wants to pass" versus "Nobody wants to pass", in the other. An example from our own lives would be reflexes like the knee-jerk reflex or the eye-blink reflex. That's our bodies reacting to perceived dangers.

Now, what is represented by the *thermostat's* variable? Not different temperatures, I'd say, but rather different degrees of too-cold-ness – besides the boring alternative "It's warm enough."

So, we could say that the thermostat's variable represents a range of *simple* states of affairs, say, "One degree too cold": turn the heating on to the lowest setting; "Two degrees too cold": turn the heating to the second lowest setting; and so on. But it is much more perspicuous if we say that it represents what I call a "*graded* state of affairs": what's going on in the thermostat's environment is treated as a state of affairs either not obtaining at all ("It's warm enough") or obtaining to various different degrees ("It's too cold to this or that degree"). This idea works well because the reactions corresponding to these degrees can similarly be described as one *type* of behavior which can assume different degrees or intensities.

Everyday-life examples would be being hungry or being tired. These are graded states of affairs too: we are more or less hungry, and our behavior changes accordingly.

Now, the different degrees to which the thermostat recognizes too-cold-ness, or the different degrees of being hungry or tired, are a kind of *parameter* of those states of affairs. And there are also other kinds of parameters for states of affairs which agents can react to: Imagine a very simple organism which is in search of certain nutrients and which has primitive sensory capabilities which allow it to sense the

direction where those nutrients are best to be found. (I'm thinking of a chemical sense akin to smell or taste, which allows the organism to detect a concentration gradient which is used to choose a direction in which to turn.) So, the state of affairs recognized and reacted to would be something like "Sugar in that direction." What is recognized here is again not a simple, either–or state of affairs but rather a state of affairs with a *direction* as parameter: a 'directional' state of affairs.

Of course, we could again treat all of these simple states of affairs – "Sugar in this direction", "Sugar in that direction", ... – we could treat these as separate states of affairs, reacted to separately – but it would again be very unilluminating.

So, we have two kinds of states of affairs with parameters now: graded ones and directional ones.

When agents get more sophisticated, they can gain the capability of recognizing still other kinds of parametrized states of affairs: they may be able to recognize, not only the direction where something relevant is, but also a distance. This could be called a "vectored" state of affairs. [The verb actually exists.] An example of a corresponding agent would be any mechanism that shoots a ballistic projectile after locating its target. It doesn't care about how big or small the target is or what other properties it has, it cares only about the place it aims to hit. So, the corresponding vectored state of affairs would be "Enemy this far off in that direction."

Agents which are even more flexible could recognize *areas* where a state of affairs is prominent, that is, particular regions with specific forms [vague, not precise], for example, an area where there is food or a tree or a river. These I call "localized" states of affairs. An example would be "Table here", "Table there", ... [Imagine that you leave the room and someone switches a few of the tables. Would it make any difference to you? Only if one table were somehow special – treated as an individual.]

So, these are different sorts of parametrized states of affairs. I suggest that these ways of coupling reactions to stimuli are types of information processing which amount to treating certain configurations in the world as one state of affairs, either simple or parametrized in some way.

9 Locating things via pockets of contrasting structure

Now, with localized states of affairs, we have already more or less arrived at kinds of things [... and recognition of kinds of things is a step towards recognition of particular things].

Why is representation of localized states of affairs a primitive form of the representation of kinds of things (or of stuffs)? – I'm not perfectly sure. Maybe there are localized states of affairs that cannot be considered as stuffs or kinds of things.[4] But the most primitive way of representing kinds of things would surely be as localized states of affairs, for instance, "Tiger here", "Water there", "Ants over there".[5]

(We could also talk about stuffs, but for brevity's sake I concentrate on things. After kinds of things, we will later come to particular things, individuals.)

[4]Yes, for example, "There's something going on there (but I don't know what, yet)", a state of affairs that warrants attention and perhaps careful exploration.

[5]Why are these states of affairs? Because the agent does not treat them as *particular* instances of X: He doesn't take in, or store, or use any information that is specifically about this particular instance. He treats the instance just as a generic instance of X and forgets about the specific place and features of it as soon as he is done with it.

Now, consider what an agent needs to locate a state of affairs like "Water over there" or "Tiger here, there, and over there". The agent needs something to fix her attention on a particular area, and to direct her actions at that area. Otherwise the agent cannot treat the stimuli from there in a special way, namely, determine what kind is instantiated there in the first place, and then track that instance and direct the appropriate behavior towards it. (That behavior might be drinking from the water instance or hiding or fleeing from the tiger instance.)

What would an area have to be like to allow a simple agent to thus focus on it? What would make an area salient in this sense? At first I thought I had to characterize such areas as "homogeneous" and "cohesive" or something like that, but now I think the crucial points are that they are roughly *connected* and that their physical structure *contrasts* with that of their surroundings. Connectedness plus contrast is what potentially makes them salient for an agent's sense organs. [The area must be perceptible and manipulable as one by the agent, therefore it must be connected; and it must stick out to the senses, therefore contrasting. Size (relative to the agent) matters: if it's too small, it cannot be perceived; if it's too large, maybe it cannot really be manipulated – or not even perceived? (E.g., a lizard on Hawaii doesn't see the island group it is on.)]

I call these connected, contrasting areas *pockets of* (*contrasting*) *structure*. They need not be crisp, that is, they may be vague. In fact, I think vagueness will be the rule, not the exception. [For examples, look at the Game of Life world, e.g., under `https://copy.sh/life/`.]

10 Tracking things via sequences of pockets

If we look at one of these areas over time, it may happen that its contrast to its surroundings *decays* (fades) immediately. In that case, there is nothing worth looking at for longer.

But it may also happen that this particular area *retains* its contrast with its environment, even though the physical structure inside (and outside) will probably change more or less strongly in its specific details. So we have a temporal sequence of areas, where each area is the same and this area furthermore remains a pocket of structure over time, even if it's not exactly the same structure at each moment.[6] In that case, looking at this time sequence of pockets is much like looking at a stationary persisting object that changes somewhat over time.

So, that's what I'd say is really, objectively there when we perceive an object persisting over time: a sequence of pockets of contrasting structure.

Whether it's worth it for an agent to treat this area as an object[7] depends on various factors: How long does the contrast persist? (If too briefly, it's not worth it.) Does its structure change radically, unpredictably, during that time? (If so, it's not worth it, because there's no consistent reaction preparable.) Is the structure not only relatively stable but also *relevant* to the agent, that is, possibly useful or possibly harmful? Then it may be worth it.

[6] Why a (discrete) sequence instead of a 'continuous' function from points in time to areas? (a) Because the latter is difficult to express in ordinary language, (b) because I don't require true continuity anyway.

[7] The phrase "worth it" needs explaining: Will natural selection tend to install a capability for dealing with it? Should a designer try to furnish a robot with such a capability? Should an intelligent agent learn about this area and prepare ways of dealing with it?

One thing that may also happen if we look at such an area over time, is that the boundary of contrast *moves*, so that, if we want to keep track of that special sort of contrast, we have to adjust the area we look at over time. Thus we get a time sequence of areas, or better, a time sequence of pockets of contrasting structure, where the areas or locations change over time [draw on blackboard?] – but not too much or too unpredictably, because otherwise we couldn't follow/track them. So, looking at such a relatively continuous sequence of pockets would be like looking at a moving something.

But whether this something could be treated as a more specific kind of thing would still depend on how much and how unpredictably the structure inside these pockets changes. If it changes very radically and unpredictably over time, then we would have a sequence of areas where something weird is going on without there being a consistent way of dealing with it.

If, however, the structure in these pockets changes only slowly, or at least predictably, then there may be a way for agents to deal with it, to either exploit or avoid this structure in some way (at least before it has changed too much for that way of reacting to be appropriate). In that case, I call such a sequence of pockets a *sequence of pockets with relatively stable structure*.

11 Stipulating entities by adopting stances towards pockets

Now, when you look at a pocket of contrasting structure, or a sequence of pockets, your brain has to decide what it cares about in this pocket. What you are looking at is primarily only a certain rough area with a certain physical structure at a certain moment in time. Which of its features make it relevant to you?

You might be interested in that pocket because of the stuff or *material* [microstructure] that is instantiated there, say, water or sugar or meat ... In that case you wouldn't care much about the shape it's in. If the shape or form changes a lot, you could still treat it as the same occurrence of water or sugar, i.e., an *occurrence of a stuff*.

Then again, you might also be interested in the specific shape or *form* of that pocket [macrostructure]. For example, suppose you have an icicle: do you want it just for its water content, or do you want it for poking holes into the snow? In the first case, you only care about its material and treat it metaphysically as an occurrence of a stuff: "this ice" or "this water"; in the second case, you also care about its form and treat it metaphysically as a *material object*: "this icicle". If the material object melts, it ceases to exist; but if the stuff melts, it is still the same occurrence of water (although not an occurrence of ice anymore).

In other cases you might care about what a certain stuff or thing *is there for*, considering it as being designed for some function. In that case, you would treat it as a kind of artifact. (It might be a biological 'artifact', for example, an organ, a trait or a behavior.)

Or you might consider a certain thing as an *agent* and be interested in its goals and its beliefs. Or you might consider the thing as a *person* and care also about that person's memories and opinions and things like that.

Or, for something completely different, you might be interested in that area because you have calculated that a meteorite will strike there, or because of what

can be seen from there. In that case, you will not care at all about how that area is structured; you will only care about that fixed area itself, about its mere location. Then you would treat it as a *place*.

So, depending on what your interests are at a given moment, you (or your brain) will adopt one or another metaphysical stance toward a given pocket: you decide to treat the pocket as an entity of this or that category. Thus you *stipulate* an entity of that category in the given area. You do not *recognize* a perfectly and objectively real entity there, nor do you *create* an entity; you just put on a certain pair of glasses toward a certain pocket, or rather toward a sequence of pockets. In a way, the stipulated entity is a fiction, not perfectly real. But the fiction can be real enough to be very useful in dealing with the environment; it can be a *real pattern*.

How does your brain stipulate an entity? By setting up a variable to represent that entity, a variable of the right data type for storing the ascribed properties or parameters, [... for interpreting stimuli 'from' the entity in terms of the right sorts of properties] and connecting this variable to the right internal model for entities of that kind to translate its information into appropriate behaviors. [Also tracking and perhaps reidentification of that entity over time.]

12 States of affairs vs. particulars/individuals

If an agent has ways of recognizing and dealing with a sequence of pockets of relatively stable structure, then it depends on the agent's sophistication whether she can treat this sequence of pockets merely as an instance of a *kind* of thing, that is, as a certain sort of localized state of affairs, or rather as a *particular* thing, an individual.

Treating a sequence of pockets as a localized state of affairs means keeping track of its shifting location and possibly of other parameters. For example, a simple agent might just notice, "Tiger here and Tiger there", and follow the movements of the tiger instances, but not take in any additional information about them, whereas a somewhat more sophisticated agent might also take in information about size or age or possible disabilities (parameters of "tiger") and update this information as the tiger instances change, and adapt her behavior accordingly. [We could also treat the parametrized state of affairs "Tiger in state so-and-so" as several more specific states of affairs. As with the thermostat, not doing so is more convenient and perspicuous.]

By contrast, what does not happen with mere states of affairs is that the agent stores information about *past* parameters, parameters which have changed since then. [Although one could perhaps consider information about the past as just another sort of parameter of a current state of affairs.] In a parametrized state of affairs, the agent tracks, and takes into account, only the current parameters.

Information about past parameters or states or properties comes into play only when pockets of relatively stable structure are treated as *particular* things. Treating such a pocket, or rather sequence of pockets, as a particular thing would involve storing and using what we might call "historical information".

When would that be valuable for an agent? Apparently when such information about past states of a pocket is (a) presumably still valid now but (b) is not easily obtainable from the present exposure to that pocket. That is, the pocket's structure hopefully hasn't changed that much, but the agent can't learn it by just taking a look at the pocket.

Imagine you encounter a tiger. Then if you don't know anything more about that tiger, you simply treat it as an instance of tigerhood: you hide or run away

– whatever seems best. But if you are able to *reidentify* this tiger as one that you have encountered before, then you might remember something useful from that past encounter, for example, that this particular tiger is blind, or that it has an aversion to loud noises. That could help you in dealing with it more successfully.

That's an instance of what I call "inference across time". In order to do this, it is not enough to just store historical information about past sequences of pockets: you also have to be able to recognize when to retrieve and apply it. That is, you have to be able to recognize a current pocket of relatively stable structure as the continuation of some past sequence of pockets that you have information about. So you need to have an internal variable representing the corresponding entity, which contains (a) whatever information you have about past pockets in the sequence (which implies how best to deal with them) and (b) information for how to recognize continuations of that sequence.

Furthermore, you presumably need some heuristics concerning which properties of the past members of the sequence will probably also apply to the *present* member, which will *not* apply, and which have to be modified appropriately to allow for the time elapsed. For example, if you encountered the tiger as a kitten you would not assume years later that it still is the same size.

So, to treat a sequence of pockets of relatively stable structure as a particular thing, as an individual, you must

1. store information about past members of the sequence,
2. you must have a method for recognizing continuations of that sequence,
3. and you must have a method of adjusting your expectations about what past information still holds.

13 Identity over time as continuity of pocket sequences

I have said that, depending on what aspects of a pocket an agent cares about, he will take different metaphysical stances towards that pocket: he will treat it either, say, as an occurrence of a stuff or as a material object; either as a mere instance of a kind or as an individual; perhaps as a designed entity or even as an agent.

Now, depending on what metaphysical stance you adopt towards a pocket of contrasting structure, different sequences of pockets count as continuations of that pocket. It's worth it to consider the pockets in a sequence as manifestations, say, of *the same material object*, as long as their form and their material do not change too much from one moment to the next [or from one encounter to the next . . .]. Because as long as these parameters or properties do not change too much, your historical information will still be somewhat accurate and thus useful. (I imagine standing next to a thing and watching it undergo small changes: "Well, this is still the same as that, and this is the same as that, and so on, so they are all the same thing" – no matter how far they change over time.)

This holds for all stances, but the properties that mustn't change too fast are different from stance to stance, so different stances have different 'continuity criteria'. I have given an example already with the icicle: considered merely as water, it survives melting, that is, loss of form; but considered as a material object, it ceases to exist when it melts. For material objects, both *material* and *form* are relevant. Analogously, for an agent, an intentional system, its *goals* and *beliefs* shouldn't

change too radically if it is to be considered as the same agent. Over long timespans, things may change radically in the relevant regards [no essences!], but in the short term, the changes should be small. Or at least that's my impression of how our ordinary concepts work. I suppose it must be similar for arbitrary agents, but I do not yet have a rationale for that idea.

This, however, is only an answer to the question whether some sequence of pockets is a continuation of another. Sometimes there arises the question which of various acceptable continuations is the *right* one. This happens in putative cases of fission and, going backwards in time, in cases of fusion. Briefly, in cases where there are two or more acceptable continuations, I think we have to look whether there is one continuation that is distinctly more similar to the original than the others. If so, then that is the unique right continuation. Otherwise there is no right continuation: the object has split into multiple objects each of which is similar enough to the original to merit identification with it; but since they cannot be identified with each other, they cannot both be identified with the original. And identifying one but not the other would be arbitrary.

[On temporary inexistence, e. g., a watch taken apart or a dried-out river: "a sufficient kernel of structure must be preserved in the meantime" – whatever that means exactly.]

So, if you have fixed a stance, and thus fixed a set of relevant sorts of properties and thus a continuity criterion for pocket sequences (and a category of entities), then you also have a criterion for whether something now is the same thing as something at another time: they must be both in the same sequence of pockets.

If you want to find out at which point in time some present thing's existence *begins* or *ends*, you have to look how far the corresponding pocket sequence continues into the past or into the future. [Although these are going to be vague points, not precise ones.]

And if you want to know what is *possible* for some present object, you have to investigate whether its pocket sequence could have taken a different turn somewhere in the past that would have taken it there: if so-and-so had happened in the past, would *that* be the right continuation of this pocket's antecedents?