On generating information

Wojciech Mamak
Logic and Cognitive Science Section
Polish Academy of Sciences

[DRAFT; PLEASE DO NOT CITE]

Outline

1. Background and motivations
2. Generation - vs. faux-generation (fixing the explananda)
3. Generativity in neural nets
4. Saved by maths? Interpolation vs extrapolation
5. Computational photography as a method of cases
6. Towards the definition
7. Potential objections
8. Conclusions + why defining generation matters.

1. Background and motivations.

Information-talk is omnipresent in cognitive science and philosophy thereof. It is uncontroversial to say that cognition is in the business of information-processsing and trading in informational states is minimally a necessary condition for the epistemic success of any successful agent. Simultaneously, discussion on actual *operations* on informational states does not feature often in philosophy of cognitive science literature, which is surprising. Rarely do philosophers try to disambiguate and classify different types of informational processing, neither do they talk in detail about the implementational differences between those.

Granted, certain terms appear as descriptions of differing informational operations in philosophy of neuroscience lingo - selection, filtering, storage, integration, amplification and attenuation of information are all examples of such. However, there does not seem to be clear understanding of differences between both in principle and in vivo, but more importantly - in ways in which they map onto higher-level explanatory posits for understanding cognition. Although, admittedly, here have been some recent notable examples in the literature. For example, (Fazekas & Nanay 2021) analyze attentional processing in terms of disambiguated informational operations - selection and amplification, arguing that the latter is responsible for implementing attention, even at sub-personal level. Similarly, Marvan et al. (2021) present a theory, where amplification receives a clear neural-level description - understood as a summation of inputs near apical integration zone - and mapping to its explanatory target, in this case - to context-sensitivity.

The aim of this chapter is to follow a similar path and offer a limited response to yet another informational mechanism in the class of such overlooked problems - information *generation*. For generativity is perhaps an even more peculiar omission on in recent philosophical discussions on the import of information-theoretical concepts for cognitive science, taking into account the sheer size of buzz around *generative AI* - a purportedly game-changing flavour of artificial intelligence. And, if that is generativity, which licenses the alleged breakthrough, one would expect an avalanche of discussions, of how we do (and should) understand the central concept of generation. After all, do we really grasp, what is being meant by generativity?

My answer to that is ‘no’. That in itself should justify the importance of the task at hand. But, the problem of generativity is not only in vogue, neither it is essentially new, I would argue. On the contrary, the problem seems to be more of an old philosophical chestnut. How’s new knowledge possible? Where do ideas and true statements come from? How come are they not entailed in the premises that have been used to infer them? That was as central and taxing of a problem for Plato (xx), as it was for British empiricists (xx) and as is now for Cameron Buckner (2018; whose philosophical strategy I will replicate here).

Terminological cloaks certainly change, and important nuances (I am intentionally *abstracting* from) linger, but the main question remains unchanged: how on Earth is abstracting or generalizing possible? And knowledge, ideas, and statements - while wildly different entities - are all tentatively made of information… right? So, where does new information come from? How’s it created? How’s it *generated*?

Summing up, the goal of this chapter is to provide a working definition of what is information-generation for a cognitive (esp. neural) system to contribute to the old chestnut.

Nevertheless, the intellectual connections do not only go that far back in time. A host of more current theoretical discussions and empirical findings will serve as important support blocks for the proposal presented here. Let me then first briefly discuss these, before we move forward.

First is the Helmholtzian idea of Inference as 'filling-in' the gaps. Famously, in 19th century psychology, the discovery of the underdetermination of stimuli - realization that different stimuli can project onto the retina in the same way (many-to-one mapping) brought about the need to explain perception as a hidden inference. For if there is no way to disambiguate the target solely on the basis of distal stimulation, surely some additional processing steps have to be involved, bringing background knowledge to the fore. For our purposes, we can say, that information available in the environment is not sufficient and has to come from somewhere else.

Note here that this revolutionary (at the time) idea of constructive perception is at odds with 'traditional', 'passive' accounts of perception, such as the ring-and-wax theory (xx). Nonetheless, these can be easily recast closer to our current purposes. For these theories all information needed for a cognizer is already present in the signal. The job of effective perceptual system is to filter-out the relevant parts of the input. In other words, perception is 'just' making sense of the informational bombardment and no internal information has to contribute, neither there is place for any internal generativity. This strand of theories remains very influential, mostly thanks to Gibsonian (xx) ecological perception, which gave it its most potent current formulation. But also, Natural Scene Statistics program and (at least according to some theorists) Bayesian perception framework (Orlandi 2014) are all essentially non-generative theories.

But the divide remains a lively one, and Neo-Helmholtzians galore, for example under the 'New Look' umbrella (Block 2014). Usually these types of theories are cashed out in terms of gist and reconstruction. Mental models, schema, prototype-based or fuzzy trace theories (Reyna and Brainerd xx) posit an internally stored skeleton of a percept (in memory), which contains the rough sketch of the desired output. Fully-formed percept is the product of this scaffolding (internal input), subsequently completed by the 'details' coming from 'actual' perceptual signal, i.e. from the objects themselves.

Please note that the roles played by 'direct' perception, understood as 'recording' of signals and the internally generated components can be switched, depending on the task at hand. When we talk of Helmholtzian optics, the rough sketch comes from the environment (underdetermined and potentially ambiguous stimuli) and internally-generated information is understood here as ‘fill-in’. But for predictive processing theories, the roles are reversed. The internal model is the gist that is stored and the details (i.e. remaining information) are added by the signal from the environment (to check against through error signal).

Regardless of that order, true internal information is constituted by whether it’s detached from currency sensory input. Be it recall, counterfactual reasoning, imagination or any other paradigmatic ‘constructive’ task, detached and internally stored representations are the contributing factors as soures of ‘filling-in’ information to be integrated, whereas any direct pipelines would limit the information-processing to selection and filtering. Perhaps it is useful to think of this idea of infernatial completion through the separation of inferential and descriptive statistics. The latter are concerned with utilizing what’s already entailed in the data, while inferential statistics requires making certain additional background assumptions that are brought to the table in order to go beyond what’s only ‘given’, or in other words - are not included in the data per se and need to be filled in.

There are three other ideas offered in the literature recently that pertinent to the account presented. I will discuss them very briefly, just to suggest some potentially interesting (and potentially illuminating) connections and avenues for further research.

First is Jake Quilty-Dunn’s work on concepts as generative pointers (2020). Quilty-Dunn tries to revive Fodorian atomism through reframing the debate on the cognitive import of concepts by interpreting them as engines or kernels of novel information. Concepts as thinking tools, he argues, are formed not only for the purpose of information-storage possible for integration or concatenation, but mostly to generate hypothesis about states-of-facts that they participate in. Clearly, it is a proposal similar to the idea of internal generativity as a core capacity of cognition.

Secondly, some work in consciousness studies seem to stem from similar intuitions. Especially, Kanai and colleagues’ (2019) paper tries to characterize conscious-processing by reserving it to processes that involve true generativity. According to their view, consciousness should be reserved only for processes that necessarily involve information-generation as its functional basis.

Thirdly, recent hotly contested debates in AI research community on compression (see Ch. 5) are also important for our current purposes. After all, the problem of compression and decompression as basic properties of effective representation hinge on the understanding, where does *new information* come from within the pipeline. Critics such as Ted Chang and Emily Bender seem think that current LLMs, regardless of their size and prowess, are only able to synthesize and filter information fed to them previously (be it in training, pre-training or tuning). At no point, according to them, there a step that can be fairly characterized as truly novel. All that language models do are regurgitating bits of the sea of information that they already possessed. Certain applications might be as impressive as they routinely are now, but the major limitation lingers firmly in place. If that is indeed the case, generative AI would not be generative, at all. No pressure.

We managed to quickly sketch both the diachronic and more current genealogy of this endeavour. We also touched on the motivations and importance of pinning down *real* generativity and providing it with adequate framework. Now we turn to the problem itself.

2. Generation - vs. faux-generation (fixing the explananda)

The first step for moving towards finding satisfactory definition of generativity ought to establishing the target. For some may say the problem is trivial, as generativity is inherently a very vast phenomenon, which occurs wherever new information is created, and as such should be defined very loosely. Why then not simply say that the only criterion for generativity is producing a new piece of information. This would seem offer a clear and simple solution to the problem.

Such a resolution might be elegant, but would not be very useful, though. It is not hard to see why. If we consider two transmitters of information, one that spits out truly random numbers, and the other one repeatedly sends only 0s, we are inclined to say that the former is more generative than the latter. Of course, the difference in this rather silly example, is immediately understandable with only basic Shannonian theory - the measure of entropy is what licenses us to posit the former coding scheme is maximally informative, whereas repetitive transmission is negligibly informational. What is important for us now is just observing that merely churning new strings that eat up (memory, channel) space is prima facie different from truly new information being produced. The simple solution trivializes the problem, rendering it uninteresting.

Adding available degrees of freedom (symbols at hand) to our 0s-only-sender would not help, either. If we think of neuronal architectures as trading-in information, simply adding more neurons does not translate into the informational gain. It does endow the system with additional power to potentially hold and process more information, granted. But, neurogenesis itself is not yet the informational brain gain. Same goes for establishing new connections between neurons (or between their parts) and for the strengthening of these. Even though, these are widely postulated as mechanisms of learning (e.g. Kennedy 2016; Castello-Waldow et al. 2020) - a paradigmatic case of informational gain, at least intuitively.

Of course, I am not denying here that these mechanisms are not plausibly necessary for true generativity. So far, I am merely pointing out the fact, that it's not enough, if we want to differentiate between actual and faux-generativity.
Again, transforming just what's - in a sense - 'already in there' does not qualify as generative, either. Otherwise, the bar is on the ground. Every new output would be 'generated'.

It is not particularly difficult to come up with a slew of routine computational processes that we would be hardpressed to categorize as generative, even though they produce 'new' input. Copying is the most obvious one. If I have a file (or a picture) and I make a backup or xerox the physical object, a new token is generated. However, regardless of whether it is identical (as in the case of a computer file, in principle) or only ever slightly degraded by noise, there is information transfer (as Dretske would agree), surely, but there is no real generativity. And we can think of more complex procedures that would not clear the bar either. Translation (in the geometric) sense is a transform that ends with a different (shifted) output on a rudimentary level. So do lossy (irreversible) transforms, which perhaps helps get the point across. The final product seems to be already entailed in the input. Importantly, there does not seem to be a major difference whether we consider something like downsizing to - for instance - data merging. In both cases, we end up with less information than we 'had' before the transform and - again - we merely transformed, not created any new data.

The cases, however, can get murkier. And indeed these murkier waters we want to probe to look for a satisfactory place for a delineation. Let's then consider procedural generation (proc-gen). It is a computational technique explicitly designed to produce 'new' (potentially arbitrarily many) inputs on the basis of a much smaller set of generation rules. It can provide an abundance of riches in terms with very simple toolbox, as any person that has sacrificed thousands of hours to a fiendishly and deceptively rogue-like game would be happy to confirm. However, proc-gen is famously rigid and gimmicky. Mentioned gamers were often unhappy when the scenarios become feeling too scripted and predictable pointing out to the fact that procedural generation is but a trick designed to mimic real creativity or novelty - albeit a very clever one.

How are these faux-generative processes different from truly generative ones we hunt down? How does new information emerge that is not already entailed in the available or accessible in the data? We need a *contentful* understanding of generativity. However this upgraded understanding of generativity has to avoid the pitfalls of overmentalization. We want naturalistic explanations with biological and artificial systems to serve as proof-of-concepts or models of the real generativity being in principle achievable without resorting to human performance as the only source of novel informational input, as certain phenomenological trends within philosophy tended to do, borrowing the idea of 'generativity' from Husserl's philosophy (xx).

If we are serious about looking for true generativity under the personal-level description, then perhaps there are already concepts and tools in information engineering already in place that could help understand where could new information come from. Information-theory is an obvious candidate for this type of insight.

Usually, in the literature - both technical and philosophical building on information-theory more focus was devoted to the notion of informational loss. Famously, Dennett (xx) argued for the existence of real patterns on the principles of algorithmic information theory (AIT; Solomonoff & Chaitin xx; Kolmogorov xx), where the former are defined over the limits of lossless compression. Similarly, in practical application, the veridical and efficient transmission of signal has more to do with the loss of information (and ways to limit, decrease or circumvent it) than information net gain. The existence of a source is usually assumed as an entry point for problem-solving (even if the properties of the source are not known and the problem is indeed tackling this uncertainty). Anyway, the aetiology of information is rarely the point of interest. Nevertheless, information-theory is equipped with a notion of 'information gain' (Cover & Thomas 2006) and it is hardly a niche concept. It is, for instance, extensively used in constructing decision-trees in data science. Information gain is a measure of the quality of attributes that could be used for parsing a given dataset. A feature that maximizes information gain is treated as a strong candidate for a good decision criterion.

Unfortunately, the 'informativeness' here is purely a technical one. Information gain in this context equals just the Kullback-Leibler divergence of a particular variable. It is a measure of how much information about a given signal could be obtained by learning the value of another variable. Importantly, then, information gain is a *relational* property stemming from the differences between a true and estimated (used) distributions. The word gain comes from the fact that substituting the distribution closer to the real one yields an increase of informativeness from the one less similar to the true distribution (the KL measure is hence quantification of this increase; Burham & Anderson 2002). Information gain does not tell us much how to differentiate 'real' generativity from an ostensible one.

3. Generativity in neural nets

A second obvious place to search for ready answers are 'generative models'. The word has become commonplace in the machine learning community already in the 1990s

(Schmidhuber xx), but has really exploded onto the scene with the emergence of deep neural networks (DNNs) and then transformers. Philosophical literature has not turned a blind eye to this fashionable concept. Especially, predictive processing framework has adopted the idea of 'generative models' as a core concept in their project of rewriting the explanatory project in philosophy of cognition, starting with Andy Clark's work on Rao and Ballard's model (Clark xx; Rao and Ballard 1991).

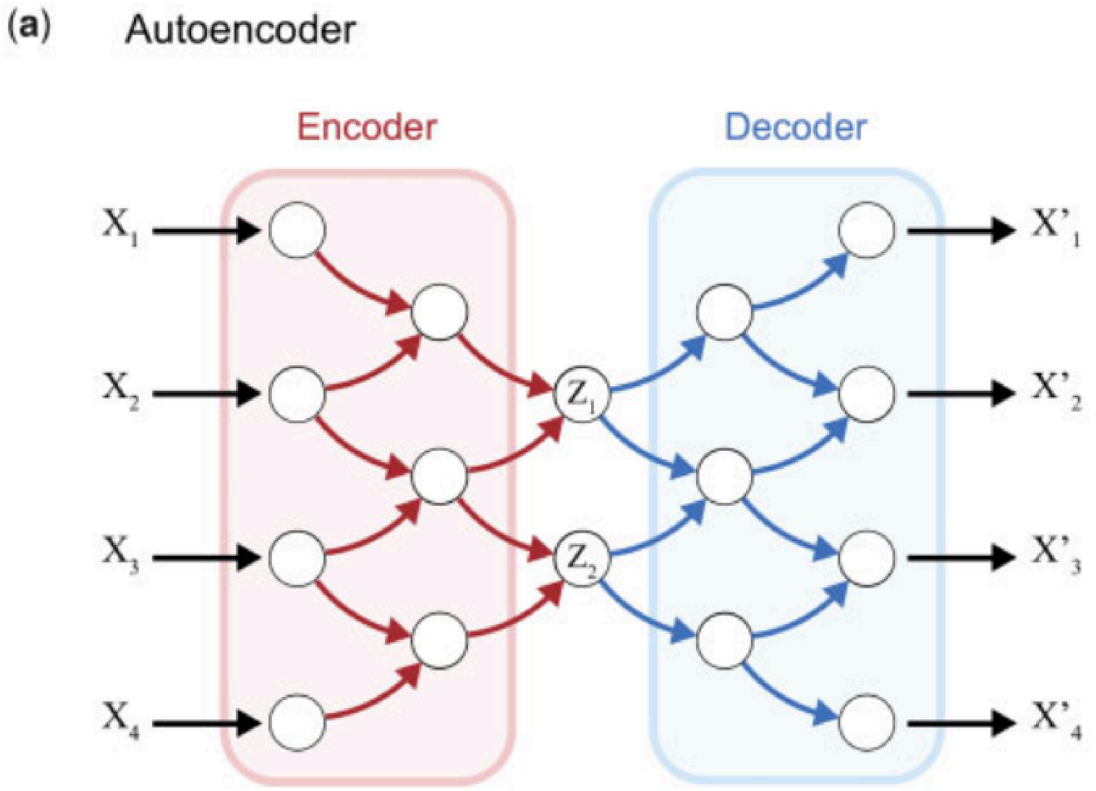
Generative model are usually defined as statistical model of the process of how sensory data are generated from a set of hidden causes (xx). They are functionally different from more classical discriminative models. While the goal of the latter is to learn the boundary of the classes, generative models are aimed at learning true distributions of the data in order to produce new samples on the basis of the training data.

Technically speaking, the beginnings of generative models can be traced even back to the 1950s with the emergence of Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs). Yet, these methods, in parts due to technical (computational) limitations, did not cause a major shift in artificial intelligent systems design in a way that we associate with modern generative systems. A partial breakthrough came in 1980s with the discovery of recurrent networks (RNNs; xx 1982), which allowed for more convincing and variational generation previously unseen inputs. An even more impressive results were achieved with the introduction of LTSM (Long Short-Term Memory) in 1991 (Hochreiter 1991), which significantly improved the performance of the models with longer strings.

However, up to the 2010s the models were mostly deployed for analyzing numerical data, sometimes texts. They struggled with other modes of data, including ones that were most sought after from a market perspective - like speech or images. That has largley changed with the introduction of variational autoencoders in 2013 by Kingma and Welling (2013). "VAEs opened the floodgates to deep generative modeling by making models easier to scale [...] Much of what we think of today as generative AI started here." (Akash Srivastava; quoted in (Martineau 2023)).

Autoencoders were revolutionary, partly because they consisted of two parts, working somewhat in opposite directions. The first one is an encoder, which is tasked with mapping unlabeled data into a compressed representation. These representations is stored in the space called the latent space, which can be conceptualized as an internal map of the model created on the basis of the training input. Regular autoencoders keep single entries, that is values of particular variables. Variational autoencoders' improvement was to store not values, but entire distributions within the latent space. That allows for the second part of the pipeline to achieve much more interesting properties.

That second part is the decoder, which intuitively does the encoders work in reverse. They invert the mapping, linking the latent space representations with the original type of input (now as output). For 'regular' autoencoders that meant reconstructing the original message and that's why they were used for denoising, such as reconstructing corrupted files, detecting errors in speech or deblurring of images. However, the aforementioned critical feature of VAEs that store distributions, not points enabled them to sample from this space. The effect was that the outputs were alike to the original message, but not identical - tokens of the same type, but not copied, or merely decompressed. Whereas more traditional coding schemes achieved transmission and processing of original data, VAEs paved the way for outputting variations of the original data.

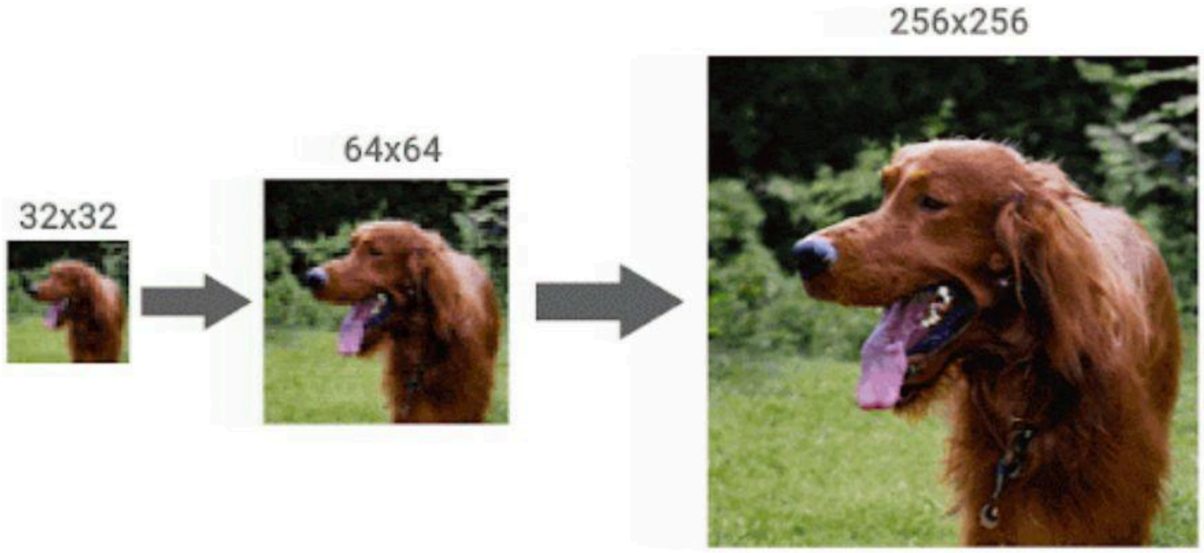


Fig, X: In autoencoders, the encoder part (shown in red) compresses sensory information to compact representations in a latent space. This representation is decoded into sensory data format. The decoder (shown in blue) can be used for counterfactual information generation using a seed chosen from the latent space. The variables z_1 and z_2 represent the latent variables. [after Kanai et al. 2019]

On the face of it, then, generative component happening in the decoder part of the autoencoder is realized through the decompression from the latent space, which is the

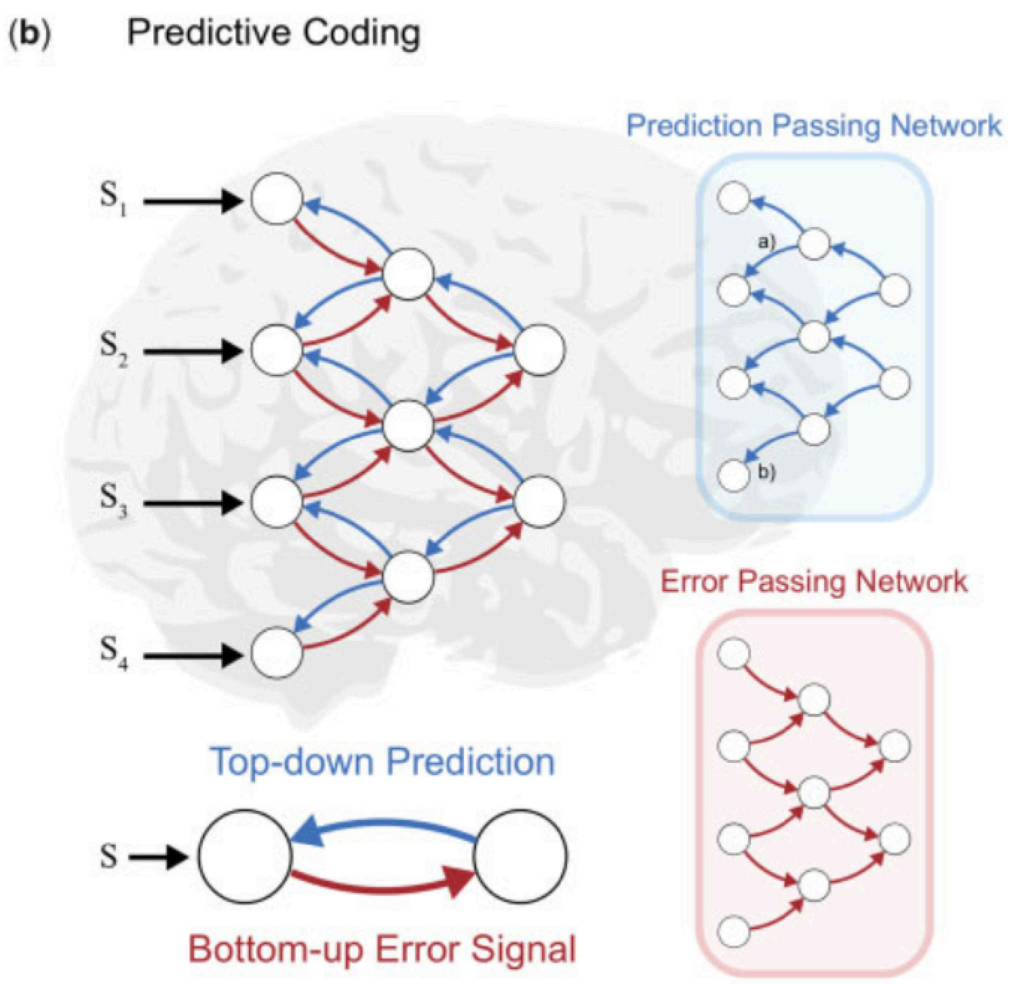
seat of ‘representations’ (in the most technical, machine learning sense of the word, where it is satisfied merely by the stand-in relation) for the model.
But it’s not a regular decompression as it as merely an expansion of what’s already there. To understand the difference, think of zipfiles then compare them to the technique called upscaling.

Diagram X: Upscaling technique. Same image of increasing resolution is inferentially upgraded in order to preserve the high fidelity to details even on a bigger scale. The feat usually not possible using ‘direct’ image processing.



Traditional compression, whilst an art in itself, is somewhat limited. Well known theorems, such as Zipf’s Law are bounds to the levels of efficacy. Also, the lossy compression is not reversible. That is immediately clear to anybody who tried to magnify a jpeg and was disappointed to find out that the bigger the picture, the more quality of the detail goes astray. The reason for that is that there is simply not enough information in the file to decide which pixel values should be taken between the known ones - if we extend the image, the empty spaces appear between them and decision gimmicks implemented in the jpeg protocol do lead to the impression of blurriness and deterioration of details, This problem is impossible to overcome without resorting to the background knowledge of images in general (at least to ones we tend to call 'natural', according to the Natural Scene Statistics camp; certain power laws (autocorrelation, power spectra) apply to images that are, well, worldly). Operating solely on the file itself only gets you so far. Discovery of these constancies in the signal led to the invention of protocols that enables reliable and noticeable differences in the ability of image storage protocols to 'fill in' the unknown details (values) of pixels that are not directly stored within a file. Using histograms of known distributions, methodologies developed by Bill Geisler and colleagues (Natural Scene Statistics researchers) enabled them to 'upscale' images with remarkably higher levels of precision. But these protocols are not mere decompressions, they utilize known dependencies and regularities in the signal to make informed inferences about the missing information and fill in the blanks on the basis of this content.

Coming back to the autoencoders. When one studies the Diagram X, it quickly becomes apparent why these sorts of advances in artificial models building excited proponents of predictive processing, bayesianism and advocated of generative models in cognition in general. After all, the encoder-decoder pipeline, resemble, quite strikingly, the picture presented by these constructivist approaches in cognitive science. For a PP proponent, cognitive architectures hinge on top-down inferences. In more radical versions, where all perception is prediction, opportunistic guesswork (Clark 2015), or 'controlled hallucination' (Seth xx), 'hypotheses' (in their very particular understanding of the term) or hyperparameters are responsible for generating predictions for the lower level. Bottom-up processing is then readily modelled by the encoder in the VAE pipeline, whereas top-down processing is modelled by the decoder. It is instructive to compare the previous diagram with the PP version of encoding-decoding duo:



(b) Predictive Coding

The colors hold their meaning across both diagrams. We can see that the encoding part's counterpart (red) is the 'error-passing network' in PP lingo, which is roughly the 'traditional' bottom-up direct perceptual input being proapagated up the network. It it responsible for fixing the latent space through learning process, which results in the formation of the internal model, that tries to approximate the structure of the environment (totality of inputs). Hence, the formation of the internal model is nothing more than the process of representation learning.

On the second hand, the 'blue' part of the schema is the reverse process - the decoding/generating side. On PP reading, this process equals to producing sensory events on the basis of information stored on the hierarchically higher levels of the system. Respectively, for VAEs this amounts to variational decompression from the latent space onto the original input space after sampling a seed from the latent space.

> Proponents of VAE-PP marriage put it succinctly: ‘We argue that information generation can be seen as production of sensory representations using internal models. This is achieved by reversing the process of representation learning, which is projection from sensory inputs to internal models.’ (Kanai et al. 2019:4)

Also the semantic pointer research program (Thagard & Stewart 2014; Eliasmith 2013) adheres to similar principles, when discussing how concepts work within their framework. For instance, Blouw et al.(2015) write:

In its most basic form, a semantic pointer can be thought of as a compressed representation that captures summary information about a particular domain. Typically, such representations derive from perceptual inputs. An image of an object in one’s visual field, for instance, will initially be encoded as a pattern of activity in a very large population of neurons. Through transformations of the sort described above, however, further layers of neural populations produce increasingly abstract statistical summaries of the origina visual input (see Fig. 1). Eventually, a highly compressed representation of the input can be produced. [...] The reason compressed representations of this sort are called semantic pointers is because they retain semantic information about the states they represent by virtue of being non-arbitrarily related to these states through the compression process. The reason why the representations are referred to as pointers is because they can be used to “point to” or regenerate representations at lower levels in the compression network (Hinton & Salakhutdinov, 2006).

We see here both crucial points of the picture. Concepts are conceived as highly compressed bundles of the original input, generalized and abstracted and stored within internal models (that is pretty standard even for ‘traditional’, i.e. purely bottom-up accounts of perception). But they are also functionally responsible for pointing to lower-level addresses in perceptual memory (the metaphor is here very literal, semantic pointers are explicitly inspired by memory addresses in computing machines) in order to regenerate information that they encode for. For this reason, they are able to guide categorization as they bundle and organize all relevant information pertinent to the task requiring this concept. It is important to note here that on this reading concepts are unstructured, contrary to e.g. coherentists (Sellarsian; xx) accounts, where concepts are highly interconnected and embedded in inferential relations that constrain them.

A very similar account of conceptual role in generative processing has been offered by Quilty-Dunn’s paper ‘Polysemy and thought: Toward a generative theory of concepts’

(2020), in which it is argued that 'regular' pointer architectures (Gallistel & King 2010) are insufficient for explaining away polysemous phenomena. The crux of the idea is to salvage compactness of concepts as compositionally-friendly and multipurpose tools of thinking without sacrificing the ability to explain how they are still able to refer to separate chunks of information they refer to (richness). It could suggest the kinship to 'mental files' (Recanati xx) type of theories, with the difference that the associationist commitments of these theories are rejected.

Interetingly, these needn't even have to be 'bank' type of scenarios where the same label refers to strictly different bodies of information, here - money vaults and river coasts (among others). Separate - well - banks of information could be subtler and cater to different functions of the same object for example. The example Quilty-Dunn gives are door, which can be summoned in cognition as 'door-qua-apertures' and 'door-qua-barriers' invoking both different functional roles, but also separate bodies of information.

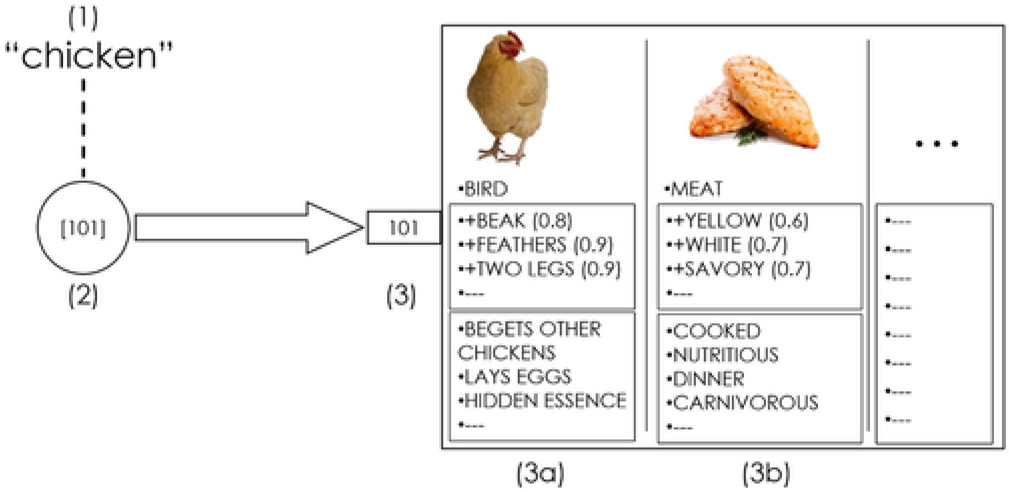


Diagram X: Yet another example of the same label that could point to separate arrays of information stored at the lower level. Chicken-qua-bird is stored at a different memory address than chicken-qua-meat, even though they are served by the same concept 'chicken'. Chicken is thus a *generative* pointer insofar it could generate one or another, contingent on the task at hand.

To borrow terminology, concepts in this framework are not merely activated, but *deployed*. Calling the concept equals to modulating the relevant denotation, not simply rigidly referring to it. As far as I am skeptical of whether this move truly enables escaping the troubles of original conceptual atomism of Fodor and followers (e.g. Lawrence & Margolis 1999) and it does seem to me that Quilty-Dunn's solution just potentially shifts the problem of disambiguation the level down without explaining how endowing concepts with diverse quasi-associative does not infringe on their compositionality, this problem should not concern us here. What is important for our purposes is to establish that the generative nature of concepts in both semantic pointer architectures and predictive processing strands of theories, both follow closely the principles postulated in artificial nets engineering dubbing themselves generative. This suggests that there is some shared understanding of generativity that exceeds mere creation or production of new information. However, it is usually not spelled out explicitly there.

4. Interpolation / Extrapolation boundary

However, there is an intresting conceptual debate going on machine learning community that tentatively could help provide a more clear-cut way to pry apart genuinely-generative from faux-generative processes. What he have in mind is the interpolation / extrapolation boundary.

If we go back to the upscaling example from part 3, we recall that what enabled the 'supercharging' of images was the inference-backed ability to 'fill in' the in-between pixels. Idea is simple - when we want to guess the value of the pixel in question (which is the blank spot after extending the image - we take what know: the value of the pixel preceding and value following and try to estimate what the *middle* value should be. We do that on the basis of prior knowledge about the internal relations between pixels in natural images.

Now we can generalize this thinking about inferring the 'middle' values between known data points and take it to apply to different modalities of data and also higher-dimensional datasets. If we consider learning algorithms, then the known data points in the previous examples would correspond to the points in the training regime. They are the known values, the already-seen input. Now, what we want from our systems (predictive ones, at least) is to be able to - well - generalize reliably from these known sets of points onto the unknown.

Now, when we think of known data-points, we can delineate a boundary around them - its geometric closure. When we consider the simplest 2-dimensional case, where points lie on X-Y axis, a most common type of that could be the line of linear regression. Now, these generalizations can occur in two ways.

First is called the interpolation regime and it entails the types of cases we discussed previously - when the value to be predicted lies between known datapoints. When this condition is not met and the value in question lies outside the geometric closure of know data, then one enters the extrapolation regime. This difference is very straightforward on the face of it and as such is routinely taught in virtually every major data science course. Diagram below illustrates the difference in the simplest terms:

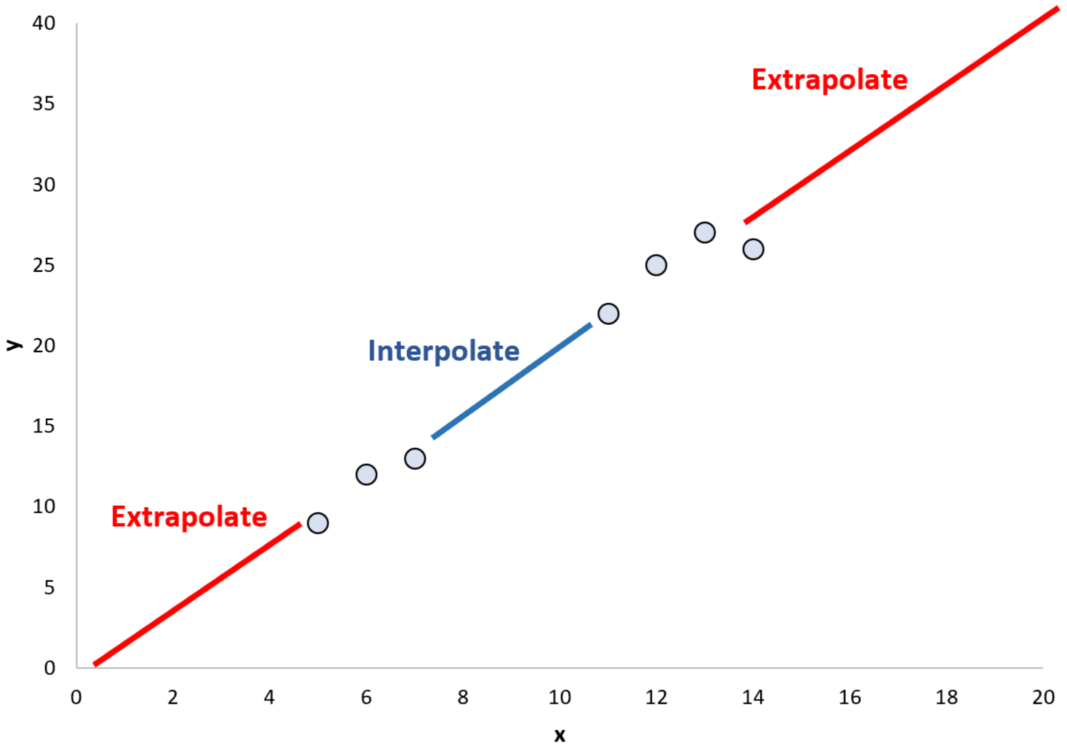


Diagram X: Difference between the interpolation (blue) and extrapolation (red) regimes. Former is reserved for inferences beteween known data points (blue dots) on the regression line, whereas the domain of the latter lies outside of it.

As said, this seems hardly controversial, even more higher-dimensional spaces. Interpretation is quite clear. Interpolation is the prediction of data points that lie on the manifold of the training data, while extrapolation is the prediction of data points that lie outside that manifold

We have talked mostly about interpolation so far, but the extrapolative abilities are the one most sought after in information technology and engineering. The reason for that is simple - the extrapolation is plausibly connected to generalization. In fact, this has been explicitly discussed after initial major successes of ANNs in the 2010s. Whilst very impressive in a variety of use-cases, critics (and some proponents; eg. Chollet 2023) were worried about the abilities for more flexible, generalized behaviour of these models. According to them, what the models only do is essentially only ‘curve-fitting’ (Marcus 2018) - getting very good at assessing the distributions of variables in the data, but only within it’s training regimes. Just to be clear, this line of reasoning was not discovered only after the carnival of deep nets in 2010s, but much earlier. For example, Haley & Solloway (1992) and Barnard & Wessels (1992) warned as far back as the early 1990s of the limitations of systems that do not extrapolate well beyond their training data.

And the real challenge - the critique went - is to escape outside, that is to truly predict, not only fill-in. That's why extrapolation was widely agreed upon to be an adequate operationalization of the ability of models to generalize and abstract well. It also seems to capture well the idea important for our purpose of pinning down genuine generativity, as it appears to give a precise mathematical formulation to the difference between true generation and mere unfurling or expansions of already known information.

However, recent literature suggests that the picture is trickier, which spells potential trouble for us, if we want to delegate explanatory work to this mathematical concepts. The reason for that is that the intuitions, built upon lower-dimensional cases, seem to fall apart for higher dimensional spaces. This is exacerbated by the fact that sophisticated cognition surely has to tackle high-dimensional input spaces well in order to be efficient without giving up on the richness of the environment.

But why things would fall apart for high-dim spaces? Shouldn't same principles of delineating along the closure of training data hold for all dimensionalities. In principle, yes, but the problem is subtler and practical. Balestriero et al. (2021) in recent study demonstrate that for spaces, which dimensionalities exceed around one hundred (R>100), true interpolation is very unlikely to be achieved in practice. One way to think of it is via the aforementioned curse of dimensionality - in highly complex spaces points are significantly more scattered. Any new given point to be predicted is very unlikely to lie inside the closure of data.

There is perhaps a more intuitive way to think about it, somewhat Monte Carlo style. If we think of our manifold defined by known data as multidimensional target and we try to throw darts at it, the entire space that we shoot at is so vastly bigger than the target, it is very unlikely we will hit anywhere of the target landing zone. This unlikelihood explodes with the number of dimensions, which explains why elegant and seemingly reasonably intuitions from simpler manifold seem to break apart, when things got fairly large.

This is a significant difficulty, but regardless of its mathematical roots, is in reality a practical obstacle, more than in principle one. It's not that it's impossible to interpolate in high-dim, just very unlikely. But problems do not stop here. After all, we are not eventually concerned with mathematical properties of interpolation or extrapolation. We are concerned with their potential explanatory work. And here even more philosophical questions surface. As usually, these stem from the fact that math is not the territory and things can get tricky when we want to find correspondences between formalisms and their realizations.

First, it is not obvious which type of geometric closure should be used for setting up the boundary between two regimes. Is it the linear hull of the training data? Is it its convex hull (as Balestriero et al. claim)? Or is it some other geometric closure of a set (Milliere, personal communication)? We can call it the 'boundary problem'

Yet, there is a further conundrum: when we consider neural nets, in what space should we be looking for that manifold? Should we look in the input space? Or should we look in the neural space? And if it's the latter, which layer of the network should we be looking at (ibidem)? This could be dubbed the 'manifold problem'.

Bonnasse-Gahot (2022) takes both the boundary problem and the manifold problem in the attempt to criticize Balestriero and colleague's results. He zooms on the penultimate layer in a neural network, that is the last hidden layer (one before the categorization decision). He utilizes an autoencoder to uncover what he calls 'the intrinsic space' of the layer, that is, the space that underlies the neural activity at this level. According to him, the dimensionality of this space is actually low. As a result, test data on this level lies well within the geometric closure defined by Balestriero (the convex hull of training data), thus satisfying the conditions of interpolation.

Interestingly, similar types of results have been reported in neuroscience literature, trying to unearth the mathematical descriptions of spaces underlying neural activities on certain levels of processing. These studies look for counterparts of what Bonasse-Gahot calls 'intrinsic spaces', that is latent 'representations' of significantly reduced dimensionality (Archer et al.,2014; Gallego et al., 2017; Jazayeri and Ostojic, 2021; Chung and Abbott, 2021)

In conjunction, it suggests that by refocusing the search, one arrives at a different conclusion about interpolation/extrapolation debate. In fact, lower dimensionality is positively correlated with the performance of the model. Such interpretation goes directly against the worries about the practical impossibility of interpolating, since it undermines the assumption of the fatal ramifications of dimensional explosion of input data. If one focuses on the hidden layer.

This has two potentially important consequences. First, it demonstrates that by probing the manifold problem, we can not only make the assumptions relying on mathematical basis for delineation more explicit, but could also help diffuse some worries. Secondly, however, it undermines the solidity of the boundary problem. For, if Bonnasse-Gahot is right (and the discussion is still ongoing), his results while alleviating the manifold problem worries, seem to suggest that the interpolation/extrapolation difference is less important than we (following both the consensus and Balestriero) suspected.

The rationale for that is the fact that what turns out to be more positively correlated with good generalization was not the inclusion in the convex hull, but the proximity to the training data. In other words, statistical measures of distance within the test space are more predictive of performance accuracy. Locality/globality (Chollet, 2021) is thus a more important feature than extrapolation/interpolation distrinction, as defined by the geometric closure. For a point laying outside the hull, but closer to the training data is on average (at least for some networks), more positively correlated with accurate prediction than the point that is scored higher by distance, but is entailed by the 'interpolation area' (i.e. is located inside the boundary)

Altogether, proximity, not closure principle does sound intuitive. We are in general more inclined to ascribe more credence to predictions closer to the already-known cases (xx). Expressing it through the terms of statistical distance bolsters the intuition. However, there may be some deeper philosophical conclusion here, too. Tentatively, it demonstrates how it can be helpful to make background assumptions explicit when making claims about the interpolation/extrapolation distinction, that is, making it clear that it is conditional on some specific mathematical definition of the delineating constraint (boundary problem) and the search of its instantiations (manifold problem). It shows that mere mathematical arguments are insufficient for answering problems such as defining genuine generativity and, as such, they cannot be treated in isolation from broader theoretical considerations. If we agree and concede that resorting to pure mathematics cannot salvage a philosophical project (again), we are ready to plunge again into assessing candidate cases of generative processes.

5. Computational photography as the method of cases

We have so far touched upon the simple cases such as copying and slightly more advanced ones like upscaling. Also, sophisticated, high-level and famously data-hungry computational pipelines such as generative neural networks were considered. In some way, we have probed candidates for generative processes from both the bottom and the top. Now, having conceded that purely mathematical resolutions are insufficient for our purposes, I want to revert to discussing cases, but now aiming for more controversial, middle-ground cases in the hope that we may stumble upon the level of generativity where difference between faux-generativity and genuine generativity could be plausibly established.

For that purpose, it could be instructive to refer to computational photography as a sandbox to create thought experiments in order to adjudicate cases through philosophical intuitions. Then we can move towards setting up the relevant definition that could account for the differences between said cases.

Modern computational photography is an interesting place to look for these cases for a host of reasons. First, it is somewhat classical of philosophy of perception to look for photography for analogies and intuition pumps for theorizing about vision (Draaisma xx), but this is the least important reason. Secondly, shifts from traditional, optical photography towards the computational one mimic well the differences discussed in the beginning of the chapter, that concerned the activity and constructivism of image or percept formation; they also relate directly to questions of informational richness of the environment vs. the need for informational completion 'from within'.

This last difference becomes very apparent right from the beginning, when the definition of computational photography is attempted. Mark Levoy defines it as "computational imaging techniques that enhance or extend the capabilities of digital photography in which the output is an ordinary photograph, but one that could not have been taken by a traditional camera'. Let's note here two things - 'enhancement' and the counterfactual. Computational photography begins whenever traditional optical photography cannot produce a given output, because it needs an additional boost.

And modern photography - even the amateur kind, hidden inside most commong smartphones - is filled to the brim with clever tricks and enhancements. Importantly, we can roughly divide them into hardware and software enchancements. Plenoptics would be an example of the former kind. The difference here stems from the fact that many modern photorecording devices are now plenoptical cameras (also known as light field cameras or Lytros -coming from the first commercially introduces devices). They are capable of capturing the intensity of light in a scene, and also the locations and directions of light rays that end up as parts of the picture. Tt's achieved through putting additional lenses between the sensor or using a multi-camera setup within one device This contrasts with conventional cameras, which record only light intensity at various wavelengths. In an important sense, these camerae are able to catch a richer informational intake than the traditional cameras do.

The situation is somewhat different for software-based enhancements. Another technique called 'stacking' may be instructive here. Contrary to everyday opinions, modern camera do not start shooting at the moment we press the shutter button. Smartphone cameras pre-shoot even before the decision have been made. Several times in rapid barrages, in fact. These proto-shoots are kept in a small memory buffer that is overwritten every few seconds. Kept entirely away from the eyes of the users, this secret life of phones feature enable them to do the said ‘stacking’. For it turns out that using the input from snapshots just directly preceding the actual ‘target’ picture (one after the shutter-stop decision by the photographer), makes possible to boost the quality through recombination of features, resolving blurriness and stabilizing the picture. In some sense, then, it is similar to plenoptics case - the camera simply has more information in its possession. In another, important sense, though, these cases differ significantly. In the stacking case, what happens is the recombination of several images, separated by time differences (albeit very short). Whereas, in the plenoptics case, what happens is a more straightforward enhancement, akin to better illumination or longer exposition period .

Nonetheless, in neither case the device resorts to information that is not available in real time. Nor it actively utilized any previous, stored information. And that is what we are looking for as a marker of generativity. Even the use of the buffer to ‘stack’, or recombine (partly) the pictures resembles more the compositional photography of Galton (xx) than any genuinely generative practice. Whenever, there is the usage of information not stored or recorded directly, then additional information has to come from somewhere and generativity is automatically one of prime candidates for that. Let’s turn to more sophisticated computational photography techniques, then.

Enter the infamous ‘moon case’. Somewhen in 2019, a Chinese tech journalist Wang Yue on Zhihu web portal (a Far-Eastern counterpart of Quora) has published his suspicions about the newly-debuted P30 Huawei smartphone, boasting a novel feature called ‘the Moon Mode’. Moon Mode promised providing high-fidelity, realistic pictures of the lunar surface without any additional utensils, such as magnifying lenses. The promised was supposedly guaranteed and powered by - of course - simply ‘AI’. The blogger run some tests, including blurring the initial input or pointing the camera at surfaces that roughly resemble the lunar surface (such as ping-pong ball illuminated in a certain fashion). The camera was still spitting out high-quality representations of the moon (Yue 2019). What’s more, however, was that further investigation revealed that the ready representations of the moon were drawn from a database and simply inserted (with some post-processing on the edges to iron out the bluntness of such a maneuver and obfuscate the real process from the user. Although, this view has been challenged (Zhang 2022), arguing that indeed an AI model was leveraged to put details in place, instead of simple overlaying (or copypasting, to be more blunt), this need not concern us here too much. The uproar about the possibility of texture insert or overlaying suggests that there is an important difference to be drawn between that and ‘real’ generative completion or between AI enhancement and mere ‘superimposed alteration’ (Zhang, 2022).

That would not be perhaps too different from the cases discussed so far, but the lunar controversy did not stop. Fast-forward to 2022, one of Huawei’s chief competitors -

Samsung, unveiled a new functionality called 'Scene Optimizer' (sic!) in its Galaxy S20 series phones. Accusations resurfaced in an almost identical fashion. User-led investigations and tests led to accusations that the company is covertly adding features and details that are 'not really there', where one points the camera to. Equally important, the selectivity of the processing raised suspicions. If the system is selective only to objects similar to the moon-shield, then the likelihood of using a sophisticated inference-based engine instead of a simple gimmick seemed more plausible and rendered the company's marketing statements of advanced AI-empowerement as unfounded (it critics were to believe).

Samsung replied quickly in the attempt to quash the controversy. Arguments they used were intresting and also helpful for understanding the types of processing in cutting edge computational photography, so I will quote them at length and comment.

> 'From 10x to 100x zoom, image quality is boosted by powerful Super Resolution AI. At one push of the shutter, up to 20 frames are captured and processed at instantaneous speeds. Advanced AI then evaluates and corrects thousands of fine details to produce detailed images even at high magnification levels. And when shooting at high magnifications, Zoom Lock uses intelligent software to set the image in place so you can shoot with minimal shake.
>
> When taking a photo with the Galaxy S21 cameras and Scene Optimizer is activated, once AI recognizes the object/scene it will work through every step of processing. AI will first start by detecting the scene/image at the preview stage by testing it from an AI model trained on hundreds of thousands images. Once the camera detects and identifies the image as a certain scene, for example, the Moon, then offers a detail enhancing function by reducing blurs and noises. Additionally in low light/high zoom situations, our Super Resolution processing is happening (I.e., multi-frames/multi-exposures are captured > A reference frame is selected > Alignment and Registration of multi-frame/multi-exposures > Solution Output). The actual photo will typically be higher quality than the camera preview. This is due to the additional AI-based multi-image processing that occurs as the photo is captured.
>
> For example, when taking photos of an object, **3 key elements are taken into place. Object detection (when Scene Optimizer is enabled), powerful AI processing and multiple frames enhancement. Each one of these features plays a critical role in order to deliver quality photos**. When combined, these features generate the proper balance between a natural

> look and detail. **The process starts by identifying an object based on a realistic human eye view, then multi-frame fusion and upscaling adds on by generating a higher level of detail to the subject, finally leveraged by AI deep learning solution it uses contextual assumption to process and piece together all the information** to delivering a high quality result.
>
> No image overlaying or texture effects are applied when taking a photo, because that w**ould cause similar objects to share the same texture patterns if an object detection were to be confused** by the Scene Optimizer' (Samsung 2022; emphases are mine)

There are a few issues to comment on here, so let's unpack it. At the end we see there is a flat rejection of the ping pong balls types of scenarios that we had seen in the Huawei case. In some way, the critics worries were accurate - the system is designed to fulfill a very narrow purpose. It is not a domain-general, flexible system. Yet, even then, a complex domain of information has to be involved in the processing taken holistically (including training), for the system on object recognition as the first step in the pipeline.

Furthermore, within processing there is a clear division between three steps of increasing sophistication, but indeed the types of processing we tried to pry apart and delineate. Object-recognition, multi-frame combination (type of stacking) and inference-based post-processing to fill in the details not directly available in the input are separate informational operations. It is the latter that endows computational photography its generative status, according to the engineers responsible for building these systems.

Does it render digital photography partly or mostly generative now? Not necessarily, for as we mentioned, Moon Mode or Scene Optimizer were limited, targeted systems and they are not exactly commonplace yet (hence the marketing spin on their novelty). But these types of systems proliferate. Google has recently introduced the 'Best Take' (xx) feature in their products, which enables users to replace the unwanted face expressions of people in the frame, using both the stored tokens of the given person's face, but also algorithmic post-processing in order to achieve seamless marquetry. One takes an almost perfect group snap, everyone bar a single person is smiling and looking great - the software gives the option to inlay a happy face onto the grimacing dissenter.

These are but gimmicks, limited tricks, a critic could say. That is warranted, these particular systems are concerned with just the gloss, one could be tempted to classify them as post-processing. But, it's not hard to imagine the toy examples where the entire or at least major part of the picture is caused by these types of generative engines. Recent successes of diffusion-based architectures (xx) and later transformers, too xx) proved it quite forcefully. Creating the entire picture using the text-to-image tools is easy and is everywhere. Technology is surely already there. What about philosophy, then? Let's play catchup and generate (sic!) more cases to probe our intuitions one more. Then we can finally attempt definitional remedies.

In order to probe the intuitions and search for differentiating factors for ascribing true generativity, let's consider a thought experiment.

> [Albanian Citizen] We present the computational pipeline (e.g. a neural network) a series of tasks that are prompted differently in each case (A-E). In all these five cases, however, regardless of how the instruction is formulated, the system is asked to produce a picture that best realizes the query. Now, let's assume that the system outputs the very same picture in all these cases (100% pixel values equivalence). Say that our end product looks like that:

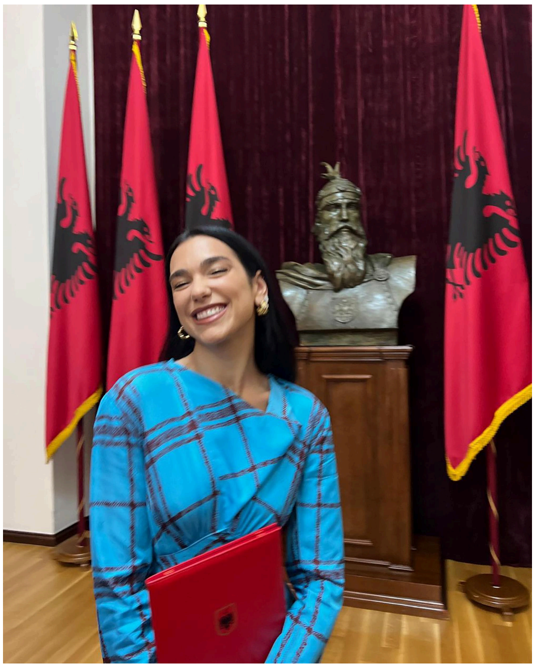


Fig X: The final generated output for all cases of [Albanian Citizen]: A-E

> For all five cases, a different query (and/or) computational process are used for arriving at the output above.

- Case A - imagine we use a simple a search engine and query 'Albanian citizen'
  - The search engine that looks up the database and brings back the relevant target from the search space (e.g. the internet) that suits the above description best.
- Case B - imagine now, instead of a traditional search engine, we instead inquire a large language general-purpose model and prompt it to'generate an 'Albanian citizen'
  - The model then creates the picture from scratch in a sense that it does not load any stored pictorial representation of this picture. It does indeed, plausibly use some sort of internal representation (again - in the

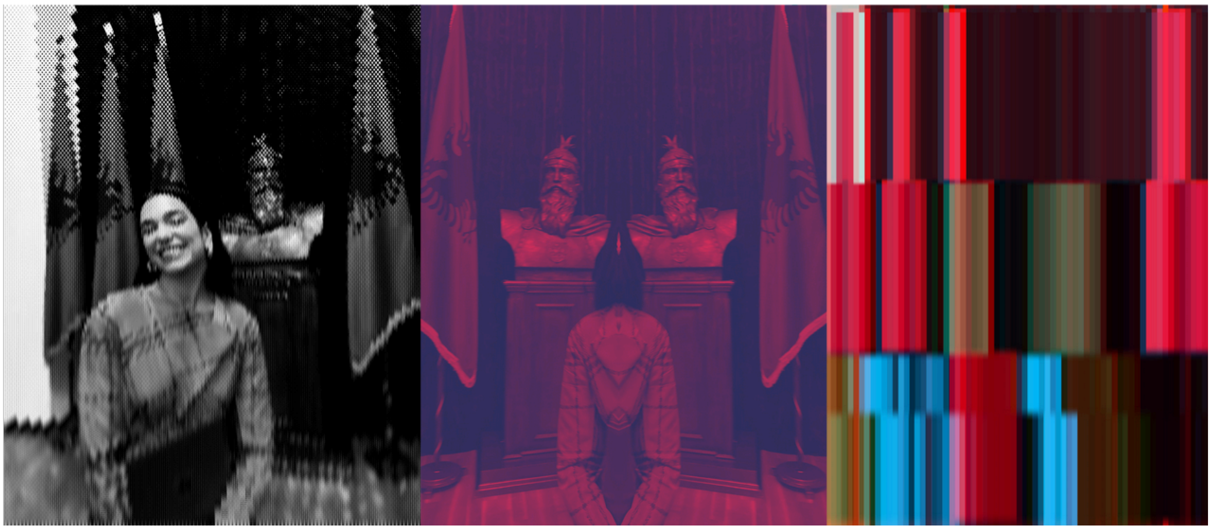
    deflated, technical sense), but much of more general nature; it is the effect of relevant learning on general terms such as ‘Albania’, ‘Citizenship’ or person (but we just assume, the entire process is largely blackboxed).

    - Thus, the result is partly spurious, in a sense that we have not given a direct instruction to generate precisely Dua Lipa in such setting. Yet, the result is not fully spurious or accidental, because of the more general representation learned through a signifcant influence of an arguably the most famous and online-present living Albanian right now in the training set. Therefore, although the process is optimized for more generalized task, the information about Dua Lipa gained during training or learning of the model, would still count as boosting performance (increasing the likelihood of generating the picture of Dua Lipa, not some other Albanian - real or hallucinated).

- Case C - now we consider a specialized network trained on of Dua Lipa, to generate new fictitious tokens of the Albanian in various situations and prompt it to ‘imagine Dua Lipa as getting citizenship’
    - In this case, the model uses the trained representation of an actual person photos to come up with an ‘unseen’ (not yet existing) picture of that person in a new, scenario according to the input instruction.

Now let’s assume that parts of the input are pictorial as well.

Fig. X: Cases - D-light (limited distortion); D-hard (more lossy distortion with more features lost); E (almost a seed), respectively left to right.



- Cases D (light and hard - left and center, respectively) - are cases of trying to get the model to attempt reversing deformation of the original picture. Imagine we applied a sequence of distorting transforms that have irrevocably deformed the picture giving us the pictures seen above on the left and center (they differ in this example only but the level of distortion, in the center case more original information we had lost). Now, using the distorted version, we want to go back to the original form. However, unfortunately, we do not store the history of performed transformations. Neither can we reverse the procedures, because the information was lost and is not expandable from the current format. We try to salvage what we can and as such revert to a clever algorithm (a neural network) and feed it the distorted picture as a gist and ask it to make its best informed guess about the original source. It exceeds expectations and achieves a full accuracy. It is somewhat unlikely, granted, but still not purely accidental - the gist plus known statistical dependencies for natural scenes give the well-trained model enough head start to be able to reconstruct the original arbitrarily closely (in principle).
- Case E - finally, let's imagine a case similar to D. Again, we use the 'blob' (right) as a seed for the model to decompress from its latent model space from. However, with a crucial difference that we assume here that the blob is not causally connected to the original. Note here that, in the example, I did produce the blob through heavy distortion, so there is still a causal link to the original. But for the sake of the illustration, let's imagine it is a more or less random blob. Now what we do is ask the model to try to predict (guess?) from this non-causal; we ask: 'how would this blob look if it was Dua Lipa' and it returns the same output as in all previous cases.

So, in which sense are all these cases different, then? Well, for A the answer is straightforward. It’s a non-generative case, but the output is not recreated, just retrieved. Information is transmitted from the source (the picture’s original address) and not created or recreated. But for B-E the interpretation is more difficult. All of them are tentatively generative, because not all relevant information is present and accessible to the system before it processes the query. As it is known, the loss of information is not reversible solely on the basis of the input. Neither it is the case of ‘unfurling’, where quantitatively more information is produced from a shorter formula according to some rigid rules (as in proc-gen).

But the final outcome is informationally richer also for our tentatively generative cases. Hence, the foundational question for this project is: where does this missing information come from? Obviously, there are two main candidates, First option - the lacking information comes from the prompt. Surely, prompt is responsible for the nature of transformation to be performed by the system. After all, prompts are instructions that should semantically contribute to what type of outcome is the operator able to expect at the end of the pipeline. But first, as we remember from VAEs example, decompressing from latent space can be triggered by a random seed, bringing about a token different from any original input, pertaining to the same kind, but not directly instructed by additional information. This suggests that the import of the information of the prompt is not requisite for at least some versions of generative processing. But even for prompted cases, and these are all remaining cases in our example (B-E), prompt is unlikely to account for the entire informational completion. Information storage from within the model has to contribute, too.

Cases B and C intuitively differ because of the scope of the model, breadth of information that the model has to have internalized and be able to generalize from. Also, the flexibility of them differs - the first one is a general purpose LLM, whereas the second is a domain-specific, fine-tuned type of model. Plausibly, the type of inference performed, differs, too. Recall the discussion of the importance of locality/globality as a criterion for true extrapolative generalizability. Here we see two systems of significantly different scope, but both relying on informational completion based on the highly-compressed internal model. Both produce new pictures adequately in accord to instructions, regardless of their varied abilities and purposes (the B model is plausibly much more powerful and flexible).. This suggests that locality/globality might not constitute a defining feature of generative models.

Case D-light may spring associations with aforementioned cases of denoising. The seed is recognizably close to the original (and in this case - the target). Still, not enough information has been salvaged for a non-completionist, traditional denoising. It becomes even cleared for the more extreme D-case [D-hard]. There, the level of deformation makes the original content barely recognizable (if at all). The task of the model might be more accurately described as ‘reconstruction’ here than mere denoising. Still, because of the causal link between the original picture and its distortion, the informational operation amount to the necessity of performing the recreation of the original signal. The lesson here is simple - regardless of ideal pixel-value identity, the aetiology matters.

The importance of the causal structure becomes more apparent in comparison to case E. As we remember, the blob there is purely uncorrelated with the target picture. There is no reversal to speak of. Although the blob certainly enters the informational picture as a source of information ('inspiration' to cook up such-an-such Albanian citizen that somewhat resembles it), it surely does not suffice to produce the output. As there is no causal lineage to follow, neither there is sufficient prompt input, the responsibility lies on the internal model decompression and its learned regularities that it employs. Case E is another strong candidate for genuine generativity.

But are these cases on a spectrum? Or it's just a family of distinct processes? When it is merely transformed and when generativity kicks in?

6. Towards the definition

On the basis of the analysis conducted so far, we are better positioned to tackle ontology of generativty and try proposing a definition that would better serve the theoretical desiderata we laid out whilst conserving some intuitions about genuine generativity (as compared to the faux-generativity). By that, one also attempts to demonstrate that only certain architectures achieve generativity (or should be ascribed that ability at least) and that whlist there are differences between generative pipelines, they are not necessarily constitutive for generativity itself (e.g. non-locality).

What was shown is that the brute generation view is not sufficient, neither are the mathematical methods for enforcing a generation/non-generation boundary and a content-based approach cannot be avoided. For the end-product is truly new (generative), not in virtue of just being outside the learned space (within the space of already known, informally) but only when there is a content net gain, too .

Thus, what *was* crucial across all candidates for genuine generativity, however, was that all pipelines able to generate genuinely acted upon actively stored information to systematically transform the signal in a regimented way according to these internally represented regularities. This stored information has to contentfully correspond to some regularities that enable that filling-in, which provides the missing informational import to infer from. We come back here to the idea we started with - inference as informational fill-in whenever there is not enough information present 'online' or ‘directly’ (same principle that guides gist+completion paradigms or fuzzy trace theories in memory research; Reyna & Brainerd 1995; xx; xx). In short, previously stored information is mobilized for generation of a new batch of information.

We can sum it up using a slightly worn out paraphrase - no generation without compressed information.

New information tokens are genuinely generated if and only if: previously stored information about underlying regularities that are explicitly represented for and accessed by the system are used in the virtue of their content

Note here that this condition might resonate with the discussion on the nature of inference itself, where in Boghossian’s groundworking paper [2014] one can find the ‘Taking Condition’ [for an ensuing discussion see e.g. Wright (2016) or Broome (xx); major contrary view can be found in (McHugh & Way 2016)]. [Taking] in its original form posits that for an inference to occur, it is not only required that the premises support the conclusion, but also that the conclusion is drawn *because* of that taking. In other words, in the virtue of their content and not only in accord with it. Boghossian originally invokes [Taking] to highlight the role of the agent performing the inference and her intentional usage of information entailed in the premises. For our purposes, the personal level description and intentional terminology are overmentalistic (and, as promised in the beginning of this chapter, a reductionist, naturalistic account of generativity is argued for here), so the analogy is only partial. Regardless, it conveys an important intuition about the causally active role of informational content of premises to make an inferential move. Another useful intuition pump could be the requirement of being causally active in the virtue of their content, imposed on representations by Ramsey (2007).

Let’s bundle up all the restrictions that were drawn for generativity and repack them into a definition.

[Generativity]: A genuinely generative process is a process that fulfills the following set of conditions:

T1: Produces quantitatively net gain batch of information [brute generation]

T2: This new batch of information is non-identical to parts of the training regime (used in the learning of the internal model), nor it is entailed in already known data. [variationality]

T3: But the new output is content-preserving (e.g. subsumed under the same type) [contentfulness]

T4: Production of the output tokens takes advantage of regularities that are actively stored in within the compressed internal model, which retains (mirrors) information about the underlying structure of the environment (through learning data) [internal structure]

T5: Generation is performed *in the virtue* of this stored informational content [taking]

T6: Internal informational import is used to ***fill-in*** the information lacking in the direct input (prompt, seed, stimulus) that is available online to the agent [inferential completion]

I argue that the view defended here avoid being too wide and trivializing the notion of generativity, as the naive brute generativity view would hold. Simultaneously, it avoids being overly restrictive, particularly with being agnostic about personal-level processing necessary for genuine generativity (creativity, novelty), sidestepping both the pitfalls of traditional philosophical insight of generativity as a phenomenological phenomenon, which declare a priori, which agents are able of performing generative processing, rather than assessing it fairly ex post, on the basis of empirical evidence. For, eventually, if we want to be serious about our naturalistic commitments, we as theorists should strive for providing fair definitional standards that all agents should be judged against in order to responsibly delineate which types of systems can achieve which types of cognitive processing.

Another upshot of this approach is that it provides a way to potentially escape the type of criticism in the mould of 'stochastic parrots' (Bender xx) or 'blurry JPGs' (Chiang xx) arguments that are too broad and vague to be considered truly damaging. If we want to truly investigate where the boundary lies and what type of system can clear the bar, we should provide more detailed description of specifications (functional or organizational) that could satisfy the critics. If any artificial information-processing systems are excluded solely on the basis of being artificial information-processors doing statistical inference, that equals throwing the baby out with the baby water (even if I agree with their assessment of some current state of affairs in LLMs, the critique is not even too scathing, but too indiscriminate). What we need as theorists are more granular operationalizations and more careful definitions, which I interned to provide here.

Thus, additionally, the account present demonstrates the importance of philosophical theorizing, showing that purely mathematical/formal solutions are not only encumbered by potentially problematic assumptions, but also insufficient for explanatory purchase in principle. Defining genuine generativity satisfactorily and prying it apart from faux-generativity has to be consistent with philosophical intuitions about the wide array of potential candidates for generativity, which is what the account presented here hopefully has achieved.

What I also want to defend very strongly is both the explanatory and pragmatic importance of having a workable definition of generativity. I flatly reject critiques that trying to set these boundaries is hopeless because of heterogeneity of generative systems (xx). Nor do I agree with statements that posing Socratic questions on the ontic status of computing processes are pseudoproblems (xx).

On the pragmatic side, ostensibly it makes a whole lot of sense to be able to discern information retrieval from generation. To assess on which side do recapitulation falls on. To have the basis for saying there is an understandable and applicable difference between loaded, processed and hallucinated images, videos and texts. Legal and practical ramifications are not very hard to see and are even more so with the flurry of recent lawsuits on behalf of artists (xx), journalists (WSJ xx), academics (xx) or even Reddit internauts (xx).

On the explanatory side, it seems to me that cognitive science cannot do without the ability of disambiguating between generative mental processes, such as imagining, planning or counterfactual thinking from non-generative ones like filtering, synthesizing, direct perception. Also, we want to be able to assess generativity within debate about particular mental natural kinds. Memory research is a prime example here. Being able to differentiate sharply between conditions for generative and non/faux-generative seems to me as crucial for judging between constructive, active, completionist (xx; xx; xx; xx) theories and passive, transmissive, retrieval-based theories (xx; xx; xx) without degenerating into verbal disputes (Chalmers 2011). The same goes for debates about inferentialism in perception or predictive coding or generative models of cognition in general. One can disagree whether purely information-based picture suffices to set up and police these boundaries, and that is fair. However, even if it is the case, it is impossible to go further with this explanatory segmentation (e.g. onto the questions on veridicality, causal structure, functional analysis) without assessing how much can be done on informational level first. Also, the information-based analysis is inescapable, inasmuch, virtually all approaches in (philosophy of) cognitive subscribe to the idea of information-processing as at least a necessary, minimal condition for cognitive processing.

7. Conclusion

In this chapter, I argued for a novel account of generative processing. Whilst the problem, has largely (and surprisingly) avoided more detailed attention from philosophers. To ameliorate that I have suggested a new account of generative processing based on the idea of internal inferential completion. It draws heavily from both philosophical intuitions, but also recent empirical findings from both cognitive science and artificial modelling in neural networks research. I assessed some potential criticisms and provided tentative responses to them. Chapter is concluded with a short defense of both the reality and importance of defining generativity both for pragmatic purposes, but also for the central (and so far, unfortunately, mostly tacit) role it plays as an explanatory concept in cognitive science.