Metacognitive Mechanisms for Novelty Processing: Lessons for AI

Giedrius T. Burachas[*], Scott Grigsby[†], William Ferguson[‡], Jeffrey Krichmar[§], Rajesh Rao[**]

Abstract

Novelty is central for survival of biological and design of artificial agents. On one hand, cognitive and neuro- sciences accumulated large corpus of experimental data addressing diverse mechanisms of novelty detection, response and adaptation. Increasing evidence supporting the Predictive Coding Theory[5] suggests an approach for integrating these diverse empirical findings of novelty research into coherent framework. On the other hand, AI and deep-learning-based machine learning systems in particular, have been mostly developed under the closed world assumption: Their performance is routinely tested using data that is in-distribution relative to training data, which resulted in fragility of these systems in face of open-world novelty. We propose an integrated approach to novelty processing in biological and AI systems, review supporting neurocognitive research and sketch a roadmap for designing novelty-aware AI systems based on Predictive Coding Theory.

1. Introduction

For successful operation in stochastic partially observable open world settings, natural and artificial agents have to be equipped with ability to detect, respond and adapt to novel stimuli and situations. Biological organisms possess genetically pre-programmed ability to detect, respond and adapt to novelty. Even honeybees with a one million-neuron brain can master zero-shot transfer tasks, while phylogenetically higher animal species use novelties as opportunities to learn rich models of the environment. Indeed, human brains are exquisitely attuned to detect all kinds of novelties that evoke a broad range of responses. These responses are subjectively experienced as being surprised, astonished, dazzled, puzzled, baffled, stumped and flabbergasted, to name just a few of the nuanced epistemic emotions associated with novelty.[1] The multitude of such responses reflects complex underlying (meta-)cognitive machinery that detects the novelties, controls attention, memory and learning and prepares an organism for adaptive responses. For example, surprise draws attention to and calls for reexamining percepts, recollections and spatial-temporal context; being puzzled or baffled call for revisiting one's understanding of the situation; being flabbergasted makes one realize substantial lack of some particular knowledge, which calls for extensive exploration and learning.

Such flexible, novelty-driven learning that updates and adapts internal models to novel circumstances is a hallmark of natural intelligence.[1] This is in contrast to current state-of-the-art machine learning (ML) based AI systems that tend to be fragile in face of novelty. Examples of such fragility to novelty in AI include in ML models, poor performance on samples drawn from

[*] SRI International, Princeton NJ 08540
[†] PAR Government Systems Corp, Beavercreek OH 45430
[‡] Raytheon BBN Tehcnologies, Cambridge, MA 02138
[§] University of California, Irvine, CA 92697
[**] Paul G. Allen School of Computer Science & Engineering , University of Washington, Seattle, WA 98195

out-of-distribution (OOD) data and in reinforcement learning and planning agents interacting with open world environments, inability to cope with unexpected and novel world states. It has been suggested that a key factor contributing to such failures is shortcut learning,[2] which exploits spurious correlations in datasets instead of learning patterns intended by the researchers. This is enabled by the common ML practice to both train and test on in-distribution data. More broadly, the causes of such fragility lie in the routinely used closed-world assumption.

In the ML papers that actually addresses open-world or open-set tasks, novelty assumes one of the following forms: 1) discrete anomalous observations or events, 2) change in the environment or context, 3) change in task. These forms of novelty have been tackled using the 1) framework for out-of-distribution (OOD) sample detection, 2) change detection methods and domain adaptation, and 3) transfer, continual and meta- learning for novel tasks.[3] These methods emerged as solutions to various empirically discovered limitations of ML models applied to specific engineering problems and thus are lacking in more general organizing principle.

We believe the research into cognitive and neuronal mechanisms of novelty detection and response can provide invaluable insights for developing such general organizing principles for novelty-aware AI systems capable of fast learning without the degradation of previously acquired memories and knowledge. Moreover, such insights might offer a way for designing radically autonomous AI agents.[4]

The rest of the paper is organized as follows. In Section 2 we present an exposition of the proposed metacognitive novelty detection cascade. Section 3 reviews some relevant cognitive and neuro-cognitive research, while section 4 offers a formalization of the notion of novelty and discussed novelty detection, response and adaptation mechanisms as an instantiation of the general framework of novelty theory[5] based on the Bayesian Brain hypothesis and Predictive Coding theory.[6,7]

2. Novelty metacognition in predictive coding models

Our perspective on novelty processing is based on an extension of the Bayesian Brain hypothesis.[8] The hypothesis posits that the brain not only learns a perception model $M_f(\theta)$ for inferring world state estimates $\hat{s}$ from observations o, but it is also equipped with an internal, generative, model $M_g(\phi)$ of the environment, which specifies a model for generating sensory observation predictions $\hat{o}$ from hidden state estimates $\hat{s}$ via the learned distribution $p(\hat{o}|\hat{s})$. The true hidden states in the environment are assumed to be drawn from a prior distribution $p(s)$, while the sensory observations are drawn from an observation distribution conditional on the hidden state, $p(o|s)$.[9]

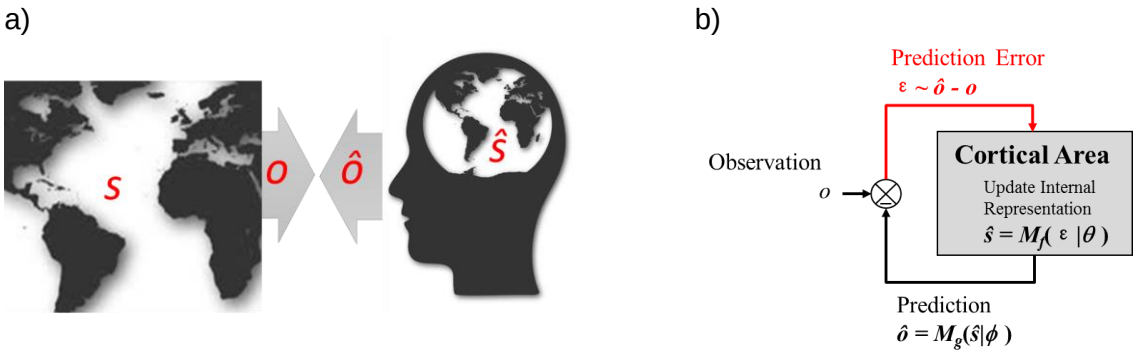


Figure 1. a) Generative World Model according to the Bayesian Brain hypothesis; b) predictive coding implementation.

Then the hidden world state is estimated using Bayes rule: $p(s|o) \sim p(\hat{o}\,|\hat{s})p(\hat{s})/p(\hat{o})$.

While the Bayesian Brain hypothesis can be implemented in a variety of ways, neurocognitive research has provided growing evidence that neuronal responses in many areas of mammalian neocortex are consistent with the predictive coding theory (PCT[6]). In PCT cortical neurons employ prediction error ε as the feed-forward signal used in inferring world state estimates $\hat{s}$. Effectively, such prediction errors are used for updating the state estimate and, following Kalman filter formalism, can be thought of as innovations. Note that Kalman filter-inspired predictive coding model can also track the variance of the prediction that can be used for setting a variable threshold for signaling novelty. To make the model $M_f(\theta)$ compatible with neurocognitive research, it can be instantiated to infer state estimates $\hat{s}$ as schemas, and depend not only on current observation, but also prior state estimates $\hat{s}_{t-1}$, $\hat{s}_{t-2}$,.. stored in episodic memory: $M_f(o_t,\hat{s}_{t-1},..|\,\theta)$. The predictive coding model is readily extendible to a hierarchical version where prediction error is calculated not only for predictions of sensory observations, but also for higher-level representations, such as schemas.

The prediction error depends on observations o_t and how well the generative model $M_g(\hat{s}_t,\hat{s}_{t-1},..|\,\phi)$ can predict future inputs $\hat{o}_{t+1}$, which in turn depends on both the state estimates $\hat{s}_t,\hat{s}_{t-1},..$ and model parameters ϕ. The state estimate $\hat{s}_t$ is updated from prediction errors and possibly previous state estimates $\hat{s}_{t-1},..$, using the perception model $M_f(\theta)$. In effect, prediction error is a function not only of observation o, but also of model parameters $\{\phi, \theta\}$ and state estimates $\hat{s}_t,\hat{s}_{t-1},..$. Thus, in order to discover the true source of the prediction error, one has to examine all of these factors. We propose that the novelty processing in the brain proceeds in a cascade that examines these potential sources of error in sequential manner so that each stage engages greater resources. The diagram in figure 2 depicts key components of such a cascade. Below we sketch the operation of the cascade.

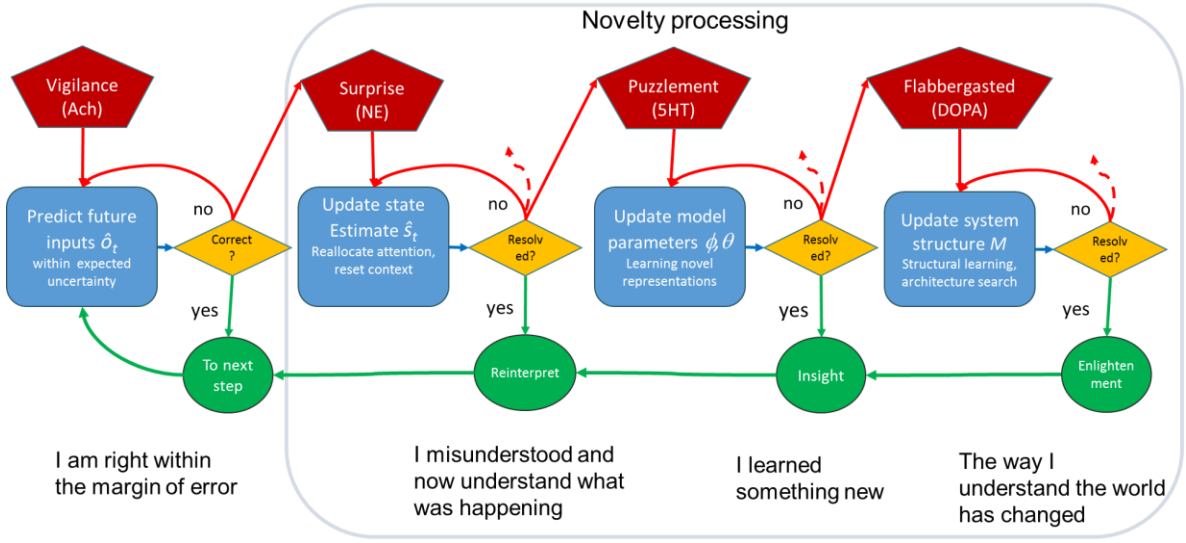


Figure 2. Proposed Novelty Processing Cascade. Red pentagons depict (meta-)cognitive control regulated by neuromodulatory systems of the brain and associated with epistemic emotions. Red arrows indicate either novelty response escalation (up-right), or iterative feedback (leftward-facing). The broken line arrows indicate quenching of novelty response (abort, postpone etc), which is associated with negative metacognitive feelings. Blue blocks are the core cognitive and meta-cognitive operations for novelty processing, yellow diamonds depict prediction error evaluation. The green ovals are novelty resolution outcomes and are associated with positive meta-cognitive feelings.

The ***novelty processing cascade*** (see figure 2) is triggered by the prediction error calculated in various neocortical areas. If the prediction error exceeds a certain threshold, which is a function of the uncertainty of prediction (e.g. for the Gaussian model the threshold can be specified as set at some distance from the mean in terms of standard deviation) that triggers the surprise response.[10] In addition, the prediction error may be modulated by task-specific attention (see the *Vigilance* pentagon of stage one) in such a manner that only task-relevant (i.e. high-reward) stimuli are contributing to the prediction error. This way, behaviorally irrelevant novel stimuli would not contribute to the surprise (albeit consider salient unexpected but irrelevant stimuli that can trigger orienting resposne; c.f.[11]).

Once surprise (a.k.a. orienting response) is triggered, the novelty processing cascade attempts to eliminate the prediction error by initially attempting to resolve lower level discrepancies using limited means, but on failure elevating processing level that may engage greater brain resources. Thus, the initial attempt to minimize prediction error is constrained to updating the inference process for re-evaluating the state $\hat{s}_t$ (updating the inference process may involve adjusting attention allocation, resetting context etc). If this fails to eliminate surprise, the novelty response is elevated to what we call "puzzlement." Puzzlement is aimed at improving model parameters (c.f. meta-update in meta-learning) and may involve covert and overt exploratory actions, such as scanning inputs and memory, seeking additional information and parameter updates using newly acquired information. If the puzzlement stage fails to identify and resolve causes of prediction error due to model parameter uncertainty, the novelty processing is elevated to the level of being "flabbergasted" (sometimes called "dazzled" in the literature[1]). At this stage the agent realizes fundamental lack of knowledge and understanding and in case of positive motivational factors proceeds with revising the world model *M* in depth that may include substantial reorganization (c.f. architecture search in ML).

Thus, epistemic emotions of novelty, mediated by the neuromodulatory mechanisms (figure 2, red pentagons), trigger meta-cognitive processing aimed at resolving the sources of prediction error (blue blocks); the prediction error is reevaluated repeatedly (yellow diamonds), until resolution is achieved resulting in positive meta-cognitive emotions (green ovals). Novelty processing escalation leading to the subsequent stage in the cascade, is resource-intensive, hence escalation is not always possible. This can lead to aborted novelty processing (broken line arrows), which subjectively is experienced as negative metacognitive feelings.[1]

Note that while surprise is a fast epistemic emotion consistent with "System 1" (fast and reflexive) operation, puzzlement and flabbergasted states are more consistent with slower operation of "System 2" (slow and deliberative) as they engage metacognitive structures.[12]

3. Neuro-cognitive mechanisms for processing novelty

In this section we review some evidence from cognitive and neuro- sciences supporting the proposed novelty processing cascade.

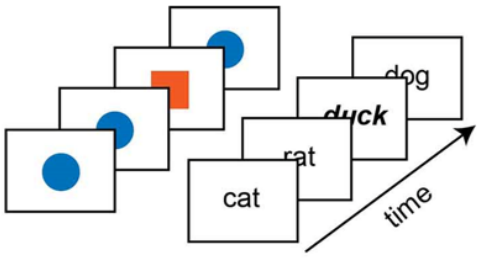
Representations of world model *M* in the brain encompass both semantic knowledge, or schemas, stored in neocortex, and episodic memories whose substrate is hippocampal formation. Indeed, people do not process and store every detail of events they have experienced, instead they use "schemata" – abstract knowledge structures. These schemata are based on expectations of the way things "should be" based on experience.[13] In machine learning (ML) context schemata can be defined as collections of objects bound together by a common context.[14]

Within schema theory, "novelty" can be defined as anything that breaks with these expectations.[15] This definition suggests that novelty is based on each person's past experience (as schemata are based on one's experiences)[1] and the world state estimate they are currently applying to their perceptions. Novelty can be a feature of objects themselves, but it can also be

a feature of new capabilities, or attributes of known objects. Novelty can also be a result of unexpected spatial elements (relationships, configurations, or environments), and unexpected temporal elements (actions, interaction, goals) of existing known objects. Therefore, within this paradigm, novelty describes an attribute we can apply to a stimulus that doesn't have a pre-existing representation.[15]

Cognitive psychology research refers to diverse novelty types, three of which have been studied extensively: **absolute, associative,** and **contextual** (a.k.a. relative) novelty.[16,17] *Absolute novelty* is the discovery of a brand-new class (of object, action, rule, etc.) and is something that has never been seen before, and hence requires learning a new schema. *Contextual novelty* is based on recent experience and arises from a mismatch between the components of a scene and activated schema and may require switching to a different schema. *Associative novelty* is detected when familiar objects are presented in novel configuration and thus may require rearranging familiar schemas (see figure 3).

a)

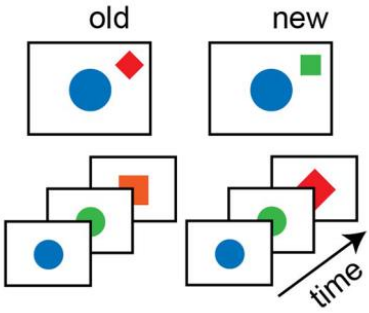


Temporal contextual novelty

b)

Spatial-relational contextual novelty

c)



Associative novelty

d)

Absolute (stimulus) novelty

Figure 3. Examples of visual novelties used in neuro-cognitive resaerch. While contextual novelties (a,b) trigger surprise, they can be resolved by updating satate/schema. Associative novelties may require learning new associations and modifying schemas (c). Absolute novelty requires instantiating entirely new schemas or paradigms (d).

In these settings, Kafkas[15] defined context in terms of "spatio-temporal or other information, that when repeatedly paired with a stimulus, or stimulus type, creates a representation." And *contextual novelty* as referring to "a familiar object in a new place (in space or temporal sequence – auth.)." Figure 3a shows a simple example of this type of novelty in the temporal domain (left) where a series of circles may have a random square interspersed within

or a series of words (in this case mammals) has one that doesn't fit the pattern. The square itself is not a novel object but its inclusion in this set is not expected and therefore novel within context. Panel 3b shows an example of contextual spatial novelty - a plane flying in front of a mountain is not novel, however the orientation of the plan is. Panel 3c shows an example of associative novelty, where familiar objects are combined in a novel configuration, while panel 3d shows an example of absolute (a.k.a stimulus) novelty.

Given these novelty types, novelty processing along the cascade of figure 2 then would proceed as follows. Within a goal directed task, the first step is vigilance – the monitoring of percepts within the current contextual paradigm defined by a task at hand. When there is a strong expectation violation by an oddball percept that falls outside predicted uncertainty tolerance, the result is surprise. This could be due to an absolute, associative or contextual novelty, all being schema discrepant. This unexpected stimulus causes an attention shift and the initial reaction would be a fast reflexive fast System 1 cognitive response that tries to quickly resolve the discrepancy through heuristic reasoning to quickly categorize the discrepancy and react appropriately. The unexpected stimuli can be resolved by registering the out-of-context deviant and updating/resetting *working memory* ($\hat{s}$) with a new modified contextual model/schema that incorporates the new information within the current paradigm. However, when this simple contextual updating fails, then the resultant response becomes puzzlement. Puzzlement is a higher-level metacognitive response when the observer realizes that what they think they know is incorrect and not easily resolved. These are likely due to absolute or associative novelties as a simple updating of context of the current model does not resolve the discrepancy. Here new information needs to be assimilated by modified schemas and incorporated into the world model, the model itself is updated (updates parameters) and insights are gained. This model expands the potential understanding of the world and results in fewer surprises in the future.

If, however, the novelty is truly absolute and does not elicit any stored schemas, updating the model can't resolve the prediction error, resulting in the person becoming flabbergasted. A completely new type of schema needs to be constructed, as a paradigm shift is required (what we might call "Paradigmatic Novelty"). This takes intense slow System 2 reasoning and potentially extensive exploration and learning to resolve the prediction error. The world model *M* must be re-learned so that it can encompass the predictive capabilities of the old model while adding a new level of understanding that can also predict the new information.

Cognitive and neurocognitive research have documented diverse effects of different novelty types on attention allocation, memory encoding, retrieval, and schema formation. For example, surprise-evoking novel stimuli compete with stimulus saliency[18] and have immediate effect on attention resulting in improved perception.[19] Cholinergic neuromodulatory system in basal forebrain (nucleus basalis of Meynert) is central to control of attention allocation,[20] but dopamine (D1 receptors) contributions to attention control have also been reported.[21]

Memory mechanisms are essential for novelty processing, as absolute novelty corresponds to experience that is not already contained in memory. Indeed, the novelty encoding hypothesis argues that novel information undergoes enhanced encoding in memory and thus leads to improved recognition performance.[10] Brain imaging studies show that novel stimuli (***absolute novelty***) elicit activations in hippocampal formation, medial dorsal nucleus of thalamus, and the anterior/inferior parts of cingulate cortex.[22] Animal studies further revealed the neuromodulatory mechanisms of memorization of novel stimuli, suggesting that novelty is detected by the hippocampus and through its connections to the ventral tegmental area, the detection of novelty can elicit ***dopamine*** release in the hippocampus, facilitating LTP at the activated synapses[23,24].

Furthermore, learning schemas is also affected by novelty manipulations. Tse and colleagues demonstrated that new information in mammals is learned extremely quickly if it matches a preexisting schema.[25] Such preexisting schemas could consolidate associative

memories as one-shot learning. The hippocampus (HPC) was necessary for learning schemas and any new information matching a schema. Plasticity in the medial prefrontal cortex (mPFC) increased when information was consistent with a familiar schema.[26] J. Krichmar et al.[27] recently demonstrated how a neural network model of mPFC develops representations of schemas and modulates indexing patterns in hippocampus to form schema-specific task representations. Neuromodulatory mechanisms were critical for rapid learning of information consistent with a familiar schema.

4. Towards agents with cascaded novelty processing

In this section we begin to formalize the proposed novelty processing cascade using the General Novelty Framework (GTF) of T. Boult et al.[5] instantiated using Bayesian Brain[8] and Predictive Coding Theory (PCT)[6,7] hypotheses. Table 1 below details mapping of the GTF concepts to the concepts compatible with PCT.

Concept	General Novelty Framework[5]	Agent-centric Probabilistic Grounding
World State and its distribution	$w_t \in W$	$p_w = p(w_t)$ $H(W) = -\sum p_w \log p_w$
Observations and observation distribution	$x_t \in O$	$p_x = -\sum_{w_t, a_t} p(x_{t+1}\|a_t, w_t)p(a_t\|w_t)p(w_t)$ $H(O) = -\sum p_x \log p_x$
Experience/ history tensor, parameterized as World Model $M(\varphi)$	$E_T = (E_{f,t}, E_{w,t})$	$M_T(\varphi) = p(\varphi\|E_T)$, $E_T = \{x_{\leq t}, z_{\leq t}, a_{\leq t}\}T$
Agent's state z (world state s + agent state) recognition function	$z_{t+1}, a_{t+1} = f_t(x_t, z_t)$	$p_t(z_{t+1}\|x_t, z_{\leq t})$ $\pi_t(a_{t+1}\|z_{\leq t})$, $a_t \in A$
Dissimilarity operator (world, perceptual)	$D_{y,T}(y', y; E_t) > \delta_y$ where $y \in \{w, x\}$	$p(x_t\|z_t; \varphi) < \delta_z$ contextual: $p(x_{t,s'}\|z_{\leq t, s\setminus s'}; \varphi) < \delta_{z_{t,s}}$
Regret (world, perceptual, agent)	$R_{y,T}: (O, A) \to \Re$ $y \in \{w, x, f\}$	$R_{\pi,T}: V(\pi^*) - V(\pi)$ where π^* is the optimal novelty policy

Table 1. Instantiation of a PCT novelty agent using the General Novelty Framework.[5]

The probabilistic model of a PCT-based novelty agent suggests criteria for triggering surprise, puzzlement and flabbergasted signals: Surprise is modelled as a prediction error exceeding threshold that depends on predicted uncertainty. High uncertainty warrants high surprise threshold and vice versa.

The puzzlement metric, *Pzz*, herein defined as a function of uncertainty of state z, can be formulated using Bayesian Surprise formula as follows:

$$Pzz = KL[p(z_t|x_s)||\mathrm{p}(z_\mathrm{t})]\ ,$$

where KL is Kulback-Leibler divergence between the pdfs of state z_t before and after Bayesian update of the state with an observation x_s.

Likewise, the metric for flabbergasted state, *Fbb*, can be defined as:

$$Fbb = KL[p(\varphi|x_s)||\mathrm{p}(\varphi)]\ ,$$

where KL divergence is calculated between pdfs of parameters φ before and after Bayesian parameter update with an observation x_s.

In future work we are planning to present instantiations of our novelty-aware agent for several task domains. We will attempt to demonstrate that AI systems designed using the cascaded novelty processing principle can be trained to identify sources of and respond to novelty in a robust manner.

5. Conclusions

Since novelty is a relationship between an intelligent agent that interacts with a dynamical environment, it depends not only on the history of the environmental states, but also agent observation, learning history and memory retention. Biological agents are endowed with complex neuronal mechanisms for detecting novelty and learning from it. Novel stimuli and experiences have great impact on attention, memory and schemas of these agents. Humans use a multi-stage cascade for processing novelties so that different levels of such cascade are evoked depending on the agent's ability to observe and represent world states in an effective, task-dependent manner, flexibility of their learning mechanisms and memory capacity and retention limits. We suggest tapping into such mechanisms for AI agent design to improve their "awareness" of and robustness to novelty.

Acknowledgements. This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA) under the SAIL-ON program. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

[1] Nerantzaki, K., Efklides, A. and Metallidou, P., 2021. Epistemic emotions: Cognitive underpinnings and relations with metacognitive feelings. New Ideas in Psychology, 63, p.100904.

[2] Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M. and Wichmann, F.A., 2020. Shortcut learning in deep neural networks. Nature Machine Intelligence, 2(11), pp.665-673.

[3] Geisa, A., Mehta, R., Helm, H.S., Dey, J., Eaton, E., Dick, J., Priebe, C.E. and Vogelstein, J.T., 2021. Towards a theory of out-of-distribution learning. arXiv preprint arXiv:2109.14501.

[4] Langley, P., 2020, April. Open-world learning for radically autonomous agents. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 09, pp. 13539-13543).

[5] Boult, T.E., Grabowicz, P.A., Prijatelj, D.S., Stern, R., Holder, L., Alspector, J., Jafarzadeh, M., Ahmad, T., Dhamija, A.R., Li, C. and Cruz, S., 2021, May. Towards a Unifying Framework for Formal Theories of Novelty. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 17, pp. 15047-15052).

[6] Rao, R.P. and Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), pp.79-87.

[7] Huang, Y. and Rao, R.P., 2011. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), pp.580-593.

[8] Doya, K., Ishii, S., Pouget, A. and Rao, R.P. eds., 2007. *Bayesian brain: Probabilistic approaches to neural coding*. MIT press.

[9] Gershman, S.J., 2019. What does the free energy principle tell us about the brain?. *arXiv preprint arXiv:1901.07945*.

[10] Reichardt, R., Polner, B. and Simor, P., 2020. Novelty manipulations, memory performance, and predictive coding: The role of unexpectedness. *Frontiers in human neuroscience*, *14*, p.152.

[11] Faraji, M., Preuschoff, K. and Gerstner, W., 2018. Balancing new against old information: the role of puzzlement surprise in learning. *Neural computation*, *30*(1), pp.34-83.

[12] Kahneman, D. (2011). Thinking Fast and Slow. Farrar, Straus and Giroux.

[13] Abbott, V., Black, J. B., & Smith, E. E. (1985). The representation of scripts in memory. Journal of Memory and Language 24, 179–199.

[14] Hwu, T. and Krichmar, J.L., 2020. A neural model of schemas and memory encoding. Biological cybernetics, 114(2), pp.169-186.

[15] Kafkas, A. & Montaldi, D. (2018). How do memory systems detect and respond to novelty? Neuroscience Letters 680, 6

[16] Barnard, W. A., Breeding, M., & Cross, H. A. (1984). Object recognition as a function of stimulus characteristics. Bulletin of the Psychonomic Society 22, 15–18

[17] Martindale, C., Moore, K., & West, A. (1988). Relationship of preference judgments to typicality, novelty, and mere exposure. Empirical Studies of the Arts 6, 79–96.

[18] Ernst, D., Becker, S. and Horstmann, G., 2020. Novelty competes with saliency for attention. Vision research, 168, pp.42-52.

[19] Schomaker, J. and Meeter, M., 2012. Novelty enhances visual perception. PloS one, 7(12), p.e50599.

[20] Thiele, A. and Bellgrove, M.A., 2018. Neuromodulation of attention. Neuron, 97(4), pp.769-785.

[21] Noudoost, B. and Moore, T., 2011. Control of visual cortical signals by prefrontal dopamine. Nature, 474(7351), pp.372-375.

[22] Tulving, E., Markowitsch, H.J., Craik, F.I.M., Habib, R., and Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. Cereb. Cortex. 6, 71–79, http://dx.doi.org/10.1093/cercor/6.1.71

[23] Lisman, J.E. and Grace, A.A., 2005. The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron, 46(5), pp.703-713.

[24] Shohamy, D. and Adcock, R.A., 2010. Dopamine and adaptive memory. Trends in cognitive sciences, 14(10), pp.464-472.

[25] Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P. and Morris, R.G., 2007. Schemas and memory consolidation. Science, 316(5821), pp.76-82.

[26] Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H. and Morris, R.G., 2011. Schema-dependent gene activation and memory encoding in neocortex. Science, 333(6044), pp.891-895.

[27] Hwu, T. and Krichmar, J.L., 2020. A neural model of schemas and memory encoding. Biological cybernetics, 114(2), pp.169-186.