Compressed information is all you need: unifying intrinsic motivations and representation learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

Humans can recognize categories, shapes, colors, grasp/manipulate objects, run or take a plane. To reach this level of cognition, developmental psychology identifies two key elements: 1- children have a spontaneous drive to explore and learn open-ended skills, called intrinsic motivation; 2- perceiving and acting are deeply intertwined: a chair is a chair because I can sit on it. This supports the hypothesis that the development of perception and skills may be continually underpinned by one guiding principle. Here, we investigate the consequence of maximizing the multi-information of a simple cognitive architecture, modelled as a causal model. We show that it provides a coherent unifying view on numerous results in unsupervised learning of representations and intrinsic motivations. This poses our framework as a serious candidate to be a guiding unifying principle.

1 Introduction

Children spontaneously explore their environment and learn skills in an open-ended way [Piaget and Cook, 1952, Ryan and Deci, 2000], driven by the so called intrinsic motivation (IM). During this process, perception and action selection are deeply intertwined. Humans gather observations as a result of their actions, select actions based on their representation of their state [Byrge et al., 2014] and perceive by looking at their action effect O'regan and Noë [2001], Gibson [1977]. It results a progressive construction of perception and skills.

Yet, in machine learning (ML), IMs objectives are often unrelated to unsupervised representation learning objectives like invariances learning Chen et al. [2020], covariances learning Dangovski et al. [2021] or disentanglement Kingma and Welling [2014]. In practice, even intrinsic motivations are highly heterogeneous; for example, one can maximize like bottleneck research [McGovern and Barto, 2001, Menache et al., 2002], expected cover time [Jinnai et al., 2019], saliency search [Bruce and Tsotsos, 2005] or novelty seeking behaviors Bellemare et al. [2016]. The information compression principle, which states that biological organisms aim to compress the data, is a natural candidate to unify IMs and representation learning objectives. This is because of its ability to explain diverse behaviors (*e.g* novelty) and feelings (*e.g* subjective beauty) [Schmidhuber, 2008]. However it lacks a quantitative instance of this principle that could guide unsupervised learning methods.

Here, we hypothesize that an agent is driven to maximize the information it compresses in its cognitive architecture. We model the cognitive architecture of an agent as a modular Bayesian network (BN) decomposed into modular goal/representation loops. We quantify the compressed information as the multi-information of the BN and propose that an agent maximize it through learning representations and skills. By introducing a lower-bound on the multi-information, we show that our framework unifies: 1- two information-theoretic formalisms of IMs [Aubret et al., 2021] accounting for exploration behaviors and the learning of a hierarchy of skills; 2- several approaches

Submitted to Information-Theoretic Principles in Cognitive Systems Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

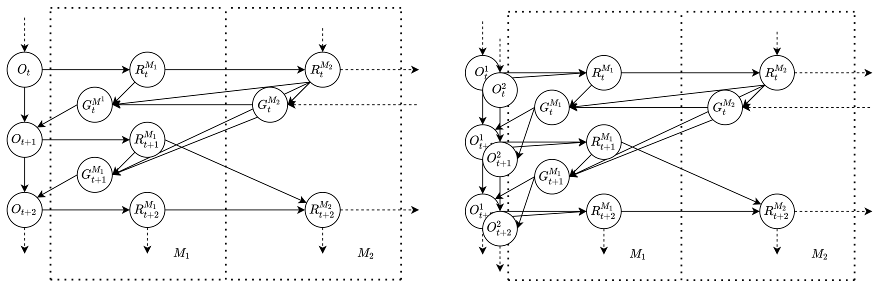

Figure 1: **Left:** BN B that sums up a simplified HRL-based cognitive model of an agent. We assume through the dotted arrows that the model is consistent through time and hierarchie. Please refer to the text for more information about the semantic of variables. **Right:** BN that considers initial independent modalities (multimodality). In our generic model, fusion of modalities may occur at an arbitrary level.

in unsupervised representation learning based on learning invariant, covariant and disentangled representations. 3- the framework of Active Efficient Coding (AEC) [Teulière et al., 2015], which simulates early infants behaviors. For an introduction to information theory, Bayesian networks and hierarchical reinforcement learning (HRL), please, refer to Appendix A.

2 Information compression in goal/representation loops

Causal model of the cognitive architecture. We consider a simple architecture endowed with high-level findings in the brain: we integrate neural-based representations [Kruger et al., 2012, Quiroga et al., 2005] and goals [Koechlin et al., 2003]. These groups of neurons perceive and act over increasingly larger time windows [Badre and D'Esposito, 2007] and increasingly larger number of modalities [Wallace and Stein, 1997].

Figure 1 shows the Bayesian network B of a goal-making step at the first and second level of a HRL framework, assuming for simplicity that the second-level goals last for two timesteps. Each module (e.g M_1) corresponds to a tuple (Representation, Goal) at one level of hierarchy with one set of modalities. First, an observation o_t is processed through a representation function to get a **representation** $R_t^{M_1} = f(o_t)$ and reprocessed with potentially other representations $R_{\Delta t}^{M_1} = (R_{t-1}^{M_1}, R_t^{M_1})$ into a higher-level representation $R_t^{M_2} = f^{M_2}(R_{\Delta t}^{M_1})$. Then it uses a high-level goal-conditioned policy, π^{M_2} to select a **goal** $G_t^{M_2} \sim \pi^{M_2}(\cdot|R_t^{M_2}, \dots)$. The goal-conditioned policy that achieves the goal $G_t^{M_2}$ in $R_t^{M_2}$, also named a **skill**, selects the ground action $a_t = G_t^{M_1} \sim \pi^{M_1}(\cdot|R_t^{M_2}, G_t^{M_2}, R_t^{M_1})$ and a_{t+1}. The agent assumes that the environment deals with the ground action and the previous observation to output a new observation. The sensori-motor loop continues this way

In Figure 1(right), each observation corresponds to a modality, which is initially assumed independent from other modalities by the agent. Actions can impact several observations (here all) and representations can aggregate different modalities (like visual, auditory, proprioceptive, ...).

Information compression. Our principle can be summed up as maximizing the information compressed by our causally represented modular cognitive model. We do so by deriving its multi-information. The multi-information between random variables of a BN can be rewritten as the sum of mutual information terms between a node and its parents [Slonim et al., 2001]. Then, we assume that our variables are stationary over time, *i.e* $p(X_t|Pa(X_t)) = p(X|Pa(X))$ where X represent any random variables and Pa represent the time-dependent parents (cf. Appendix B). It leaves us with:

$$MI(B) = \sum_{\substack{M \in \text{Mods} \\ t \in [T_0;T]}} \underbrace{I(G_t^M; G^{M_{next}}, R^{M_{next}}, R_t^M)}_{\text{Goal achievement}} + \underbrace{I(R_T^M; R_{\Delta t}^{M_{prev}})}_{\text{Compression/Novelty}} + \sum_{o \in \mathcal{O}} \underbrace{I(O_T^o; O_{T-1}^o, G_{T-1}^o)}_{\text{Controllability}}$$

where T_0 is the current step for a given module ($V_{T_0}) = V$ for all random variables V) and T the step corresponding to the computation of the next representation; O the set of modal-specific observations; M_{next} are the modules that directly follow M (immediate larger timescale or set of

Objective	Representation functions	Policies
$I(R_T^M; R_{\Delta t}^{M_{prev}})$	Compression	Novelty
$I(R_T^{M_{apnext}^{next}}; G^{M_{next}}, R^{M_{next}} \| R^{M_{apnext}})$	Covariance	Skill learning
$I(O_T^o; O_{T-1}^o, G_{T-1}^o)$	$\emptyset$	Controllability

Table 1: Summary of functions of the three parts of the lower bound.

modalities); M_{prev} the ones that precede M including observations and $R_{\Delta t}^{M_{prev}}$ the representations of M_{prev} between R^M and R_T^M. *Goal achievement* does not relate to current losses in the machine literature. To compensate this, we propose to relate information transmission of the *Goal achievement* to information transmission between two high-level representation. We can quantify the relation between the two and we exhibit our lower-bound in Equation 1 (cf. Appendix C for the derivation),

$$\sum_{M \in \text{Mods}} \sum_{t \in [T_0;T]} \underbrace{I(G_t^M; G^{M_{next}}, R^{M_{next}}, R_t^M)}_{\text{Goal achievement}} \geq \underbrace{I(R_T^{M_{apnext}^{next}}; G^{M_{next}}, R^{M_{next}} | R^{M_{apnext}})}_{\text{Covariance/Skill learning}} \quad (1)$$

with M_{apnext} being all the modules that precede those that follow M. In the next section, we interpret the different parts of our lower bound and multi-information relatively to IMs and representation learning methods.

3 Works unified through the compressed information

Let us look at the maximization of our lower bound with the representation functions and the policies. We sum up our set of objectives in Table 1.

Compression *Compression* typically refers to an **infomax** loss [Linsker, 1990, Hjelm et al., 2019]. Originally, *infomax* maximizes the mutual information between inputs X and ouputs Y of a neural network $I(X;Y)$. In our case, we set $X = R_{\Delta t}^{M_{prev}}$ and $Y = R_T^M$ such that an agent builds a representation that maximally keeps information about downstream trajectories and modalities. In some case, we may have $H_{max}(R_T^M) < H_{max}(R_{\Delta t}^{M_{prev}})$ because of some architectural constrains. It results in a lossy compression setup, *i.e* the agent necessarily lose information.

In the case of a multimodal agent, lossy compression induces an interesting property. Our architecture implies that incoming modalities are *a priori* independent, which may be wrong in practice: haptic feedbacks possibly give information about visual inputs. As a consequence of such wrong *a priori*, the agent may focus on redundant information across modalities [Wilmot and Triesch, 2021] (cf. Appendix D). In developmental psychology, this privileged perception refers to the intersensory redundancy hypothesis [Wilcox et al., 2007, Bahrick et al., 2004]. For example, an agent endowed with touch and vision and subject to this hypothesis may encode the shape of the object as a priority. We leave to Appendix E an analysis of how compression can induce disentanglement.

Covariance. The *Covariance* impacts what information is kept in a representation. We are only aware of works considering its non-hierarchical unimodal variant without considering ground observations ($M_{apnext} \approx \emptyset$) through contrastive learning approximations [Poole et al., 2018].

Let us first consider the time-contrastive learning lower-bound $I(R_T^M, R^M)$ of the *Covariance* Poole et al. [2019]. The loss discards temporally frequently changing features (*e.g* quick motion) and thus learn temporally invariant features [Wiskott and Sejnowski, 2002, Schneider et al., 2021]. We can list 3 application domains: 1) in reinforcement learning (RL), some implementation examples capture the action-based geometry of an environment, like (x,y) coordinates using ground observations [Li et al., 2021, Zhou et al., 2019]; 2) in video contrastive learning, agents manage to learn representations of time-extended behaviors displayed in a clip [Feichtenhofer et al., 2021, Qian et al., 2021, Liu et al., 2021]; 3) in visual contrastive learning, several works showed that, along with object interactions and gaze control, *Covariance* improves object recognition [Schneider et al., 2021, Wang et al., 2021] and categorization [Aubret et al., 2022].

Now let us focus on extensions that include the action A (or similarly goal G^M in our case) to maximize $I(R_T^M; G^M, R^M)$ [Nachum et al., 2019, Shu et al., 2020]. This makes R^M covariant to

G^M, meaning that the agent must keep both low-frequency information (as before) and the higher-frequency variation caused by the action (or augmentation). In RL, the resulting representation can be optimal to solve a RL task [Rakelly et al., 2021]. In visual/video contrastive learning, explicit sensitivity to some augmentations G^M can improve object categorization [Dangovski et al., 2021, Zhang et al., 2019, Xiao et al., 2020], even though they mostly apply biologically not plausible image transformation like color jittering. We note an exception [Jenni and Jin, 2021] which shows improved video representations. Our derivation essentially augments this loss to scale it to hierarchically ordered representations. Details (*e.g* ground observation $O \in R^{M_{apnext}}$) may be discarded by a low-level modules but re-encoded in high-level modules. For instance, "I saw shoes" (low-level representation) is informative about the fact that "I am going to the school" (high-level representation).

Novelty The *Novelty* has recently been formalized in Aubret et al. [2021] as maximizing $I(O; R) = H(O) - H(O|R)$. This tells us where an agent must go to improve its representation of the environment and is known to incite a wide exploration of an environment (maximizing $H(O)$) Aubret et al. [2021]. Our derivation emphasizes the need of looking for new trajectories and multi-modal combinations rather than simple observations.

In case of multimodal agents, we have seen with the *Compression* term that multimodal compression favors cross-modal redundant information. An agent may also actively look for such redundant information. It has been explored under the principle of Active Efficient coding (AEC) [Teulière et al., 2015], which is an intrinsic motivation that states humans act and perceive to better compress their sensory inputs. Several formalisms of AEC consider multimodal agents, an agent that separately perceives through its two eyes [Zhao et al., 2012, Eckmann et al., 2020, Vikram et al., 2014] or through visual and proprioceptive inputs [Wilmot and Triesch, 2021]. Our formalism of multimodal novelty directly matches [Eckmann et al., 2020], where they show that it can model the accommodation-convergence reflex in humans. This reflex describes the ability of humans to focus or defocus the vision on nearby or distant objects. Our loss suggests that similar results could be obtained for higher level behaviors.

Skill learning. *Skill learning* makes possible learrning of reusable skills, *i.e* goal-conditioned policies that have a predictable trajectory. In the DRL literature, an agent learns abstract skills by ensuring that low-level skills follow its high-level assigned goals [Aubret et al., 2021]. This is formalized as maximizing $I(G; R_{\delta t})$ where u extracts a subpart of the trajectory (*e.g* a state). For an agent that controls its torques (low-level actions), an example of time-extended skill can be to directly navigate in cardinal directions in a maze [Li et al., 2021]. Our derivation differs from the original formalism [Aubret et al., 2021] by including the previous representations $R^{M_{apnext}}$. This highlights two properties. First, skills are relative to their starting position, so that they apply a shift in the representation (like going to the right) rather than targeting a particular representation (reaching the wall). In practice, this may require to know which skill can be executed where, *i.e* skills affordances [Gibson, 1977, Khetarpal et al., 2020]. Second, it scales the loss to a hierarchy of time-extended skills as they also impact lower-level representations. At the lowest observations/actuators level, skill learning collapses to *Controllability* such that the agent tries to smartly select the actions to make according to how well the consequences are predictable [Touchette and Lloyd, 2004].

4 Conclusion

We hypothesized that an intrinsically motivated entity spontaneously tries to compress information. We propose to instantiate it as the maximization the multi-information of the constrained cognitive architecture of an agent. We showed that this amounts to maximize close-to previously known objectives within a modular hierarchy of goals/representations loops, thereby accounting for the autonomous learning of temporally/modality-extended representations and skills. By grounding our framework in ML research, our principle (thanks to the *Novelty*) avoids the dark-room problem faced by the Free-Energy Principle (FEP) [Friston, 2010] without adding ad hoc priors [Friston et al., 2012].

Our framework presents several limitations. First, we consider a very simple cognitive architecture. Second, our framework does not aim to explain how to maximize the local information theoretic terms. It implies to approximate information theoretic values which is known difficult [Belghazi et al., 2018, Poole et al., 2019]. Third, we hope future works will connect our contribution to studies about the impact of multi-information maximization at the level of neurons [Ay, 2002].

References

Arthur Aubret, Laetitia Matignon, and Salima Hassas. An information-theoretic perspective on intrinsic motivation in reinforcement learning: a survey. 2021.

Arthur Aubret, Markus Ernst, Céline Teulière, and Jochen Triesch. Time to augment contrastive learning. *arXiv preprint arXiv:2207.13492*, 2022.

Nihat Ay. Locality of global stochastic interaction in directed acyclic networks. *Neural Computation*, 14(12):2959–2980, 2002.

David Badre and Mark D'Esposito. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of cognitive neuroscience*, 19(12):2082–2099, 2007.

Lorraine E Bahrick, Robert Lickliter, and Ross Flom. Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, 13(3):99–102, 2004.

Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeswar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and R Devon Hjelm. Mine: mutual information neural estimation. *arXiv preprint arXiv:1801.04062*, 2018.

Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.

Marc Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Remi Munos. Unifying count-based exploration and intrinsic motivation. In *Advances in Neural Information Processing Systems*, pages 1471–1479, 2016.

Neil Bruce and John Tsotsos. Saliency based on information maximization. In *Advances in neural information processing systems*, pages 155–162, 2005.

Martin V Butz. Toward a unified sub-symbolic computational theory of cognition. *Frontiers in psychology*, 7:925, 2016.

Lisa Byrge, Olaf Sporns, and Linda B Smith. Developmental process emerges from extended brain–body–behavior networks. *Trends in cognitive sciences*, 18(8):395–403, 2014.

Kuo-Chu Chang and Robert M Fung. Node aggregation for distributed inference in bayesian networks. In *IJCAI*, pages 265–270. Citeseer, 1989.

Ricky TQ Chen, Xuechen Li, Roger Grosse, and David Duvenaud. Isolating sources of disentanglement in vaes. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 2615–2625, 2018.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

Cédric Colas, Tristan Karch, Olivier Sigaud, and Pierre-Yves Oudeyer. Intrinsically motivated goal-conditioned reinforcement learning: a short survey. *arXiv preprint arXiv:2012.09830*, 2020.

Paul Dagum, Adam Galper, and Eric Horvitz. Dynamic network models for forecasting. In *Uncertainty in artificial intelligence*, pages 41–48. Elsevier, 1992.

Rumen Dangovski, Li Jing, Charlotte Loh, Seungwook Han, Akash Srivastava, Brian Cheung, Pulkit Agrawal, and Marin Soljačić. Equivariant contrastive learning. *arXiv preprint arXiv:2111.00899*, 2021.

P Dayan and GE Hinton. Feudal reinforcement learning. nips'93 (pp. 271–278), 1993.

Samuel Eckmann, Lukas Klimmasch, Bertram E Shi, and Jochen Triesch. Active efficient coding explains the development of binocular vision and its failure in amblyopia. *Proceedings of the National Academy of Sciences*, 117(11):6156–6162, 2020.

Babak Esmaeili, Hao Wu, Sarthak Jain, Alican Bozkurt, Narayanaswamy Siddharth, Brooks Paige, Dana H Brooks, Jennifer Dy, and Jan-Willem Meent. Structured disentangled representations. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2525–2534. PMLR, 2019.

Christoph Feichtenhofer, Haoqi Fan, Bo Xiong, Ross Girshick, and Kaiming He. A large-scale study on unsupervised spatiotemporal representation learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3298–3308. IEEE Computer Society, 2021.

Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2): 127, 2010.

Karl Friston, Christopher Thornton, and Andy Clark. Free-energy minimization and the dark-room problem. *Frontiers in psychology*, 3:130, 2012.

Shuyang Gao, Rob Brekelmans, Greg Ver Steeg, and Aram Galstyan. Auto-encoding total correlation explanation. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1157–1166. PMLR, 2019.

James J Gibson. The theory of affordances. *Hilldale, USA*, 1(2):67–82, 1977.

Uri Hasson, Janice Chen, and Christopher J Honey. Hierarchical process memory: memory as an integral component of information processing. *Trends in cognitive sciences*, 19(6):304–313, 2015.

R. Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Philip Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019. URL `https://openreview.net/forum?id=Bklr3j0cKX`.

Simon Jenni and Hailin Jin. Time-equivariant contrastive video representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9970–9980, 2021.

Yuu Jinnai, Jee Won Park, Marlos C Machado, and George Konidaris. Exploration in reinforcement learning with deep covering options. In *International Conference on Learning Representations*, 2019.

Khimya Khetarpal, Zafarali Ahmed, Gheorghe Comanici, David Abel, and Doina Precup. What can i do here? a theory of affordances in reinforcement learning. In *International Conference on Machine Learning*, pages 5243–5253. PMLR, 2020.

Hyunjik Kim and Andriy Mnih. Disentangling by factorising. In *International Conference on Machine Learning*, pages 2649–2658. PMLR, 2018.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

Alexander S Klyubin, Daniel Polani, and Chrystopher L Nehaniv. Organization of the information flow in the perception-action loop of evolved agents. In *Proceedings. 2004 NASA/DoD Conference on Evolvable Hardware, 2004.*, pages 177–180. IEEE, 2004.

Etienne Koechlin, Chrystele Ody, and Frédérique Kouneiher. The architecture of cognitive control in the human prefrontal cortex. *Science*, 302(5648):1181–1185, 2003.

Norbert Kruger, Peter Janssen, Sinan Kalkan, Markus Lappe, Ales Leonardis, Justus Piater, Antonio J Rodriguez-Sanchez, and Laurenz Wiskott. Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE transactions on pattern analysis and machine intelligence*, 35 (8):1847–1871, 2012.

Te-Won Lee, Mark Girolami, Anthony J Bell, and Terrence J Sejnowski. A unifying information-theoretic framework for independent component analysis. *Computers & Mathematics with Applications*, 39(11):1–21, 2000.

Sergey Levine. Reinforcement learning and control as probabilistic inference: Tutorial and review. *arXiv preprint arXiv:1805.00909*, 2018.

Siyuan Li, Jin Zhang, Jianhao Wang, and Chongjie Zhang. Efficient hierarchical exploration with stable subgoal representation learning. *arXiv preprint arXiv:2105.14750*, 2021.

Ralph Linsker. Perceptual neural organization: Some approaches based on network models and information theory. *Annual review of Neuroscience*, 13(1):257–281, 1990.

Yang Liu, Keze Wang, Lingbo Liu, Haoyuan Lan, and Liang Lin. Tcgl: Temporal contrastive graph for self-supervised video representation learning. *arXiv preprint arXiv:2112.03587*, 2021.

Amy McGovern and Andrew G Barto. Automatic discovery of subgoals in reinforcement learning using diverse density. 2001.

Ishai Menache, Shie Mannor, and Nahum Shimkin. Q-cut—dynamic discovery of sub-goals in reinforcement learning. In *European Conference on Machine Learning*, pages 295–306. Springer, 2002.

Ferdinando A Mussa-Ivaldi and Emilio Bizzi. Motor learning through the combination of primitives. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355 (1404):1755–1769, 2000.

Ferdinando A Mussa-Ivaldi and Sara A Solla. Neural primitives for motion control. *IEEE Journal of Oceanic Engineering*, 29(3):640–650, 2004.

Ofir Nachum, Shixiang Gu, Honglak Lee, and Sergey Levine. Near-optimal representation learning for hierarchical reinforcement learning. In *International Conference on Learning Representations*, 2019.

J Kevin O'regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5):939–973, 2001.

Judea Pearl. Causal inference. *Causality: Objectives and Assessment*, pages 39–58, 2010.

Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

Jean Piaget and Margaret Cook. *The origins of intelligence in children*, volume 8. International Universities Press New York, 1952.

Ben Poole, Sherjil Ozair, Aäron van den Oord, Alexander A Alemi, and George Tucker. On variational lower bounds of mutual information. In *NeurIPS Workshop on Bayesian Deep Learning*, 2018.

Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International Conference on Machine Learning*, pages 5171–5180. PMLR, 2019.

Rui Qian, Yeqing Li, Liangzhe Yuan, Boqing Gong, Ting Liu, Matthew Brown, Serge Belongie, Ming-Hsuan Yang, Hartwig Adam, and Yin Cui. Exploring temporal granularity in self-supervised video representation learning. *arXiv preprint arXiv:2112.04480*, 2021.

R Quian Quiroga, Leila Reddy, Gabriel Kreiman, Christof Koch, and Itzhak Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107, 2005.

Kate Rakelly, Abhishek Gupta, Carlos Florensa, and Sergey Levine. Which mutual-information representation learning objectives are sufficient for control? *arXiv preprint arXiv:2106.07278*, 2021.

Richard M Ryan and Edward L Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67, 2000.

Jurgen Schmidhuber. Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In *Workshop on Anticipatory Behavior in Adaptive Learning Systems*, pages 48–76. Springer, 2008.

Felix Schneider, Xia Xu, Markus Roland Ernst, Zhengyang Yu, and Jochen Triesch. Contrastive learning through time. In *SVRHM 2021 Workshop@ NeurIPS*, 2021.

Rui Shu, Tung Nguyen, Yinlam Chow, Tuan Pham, Khoat Than, Mohammad Ghavamzadeh, Stefano Ermon, and Hung Bui. Predictive coding for locally-linear control. In *International Conference on Machine Learning*, pages 8862–8871. PMLR, 2020.

Noam Slonim, Nir Friedman, and Naftali Tishby. Agglomerative multivariate information bottleneck. In *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*, pages 929–936, 2001.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

Céline Teulière, Sébastien Forestier, Luca Lonini, Chong Zhang, Yu Zhao, Bertram Shi, and Jochen Triesch. Self-calibrating smooth pursuit through active efficient coding. *Robotics and Autonomous Systems*, 71:3–12, 2015.

Serge Thill, Daniele Caligiore, Anna M Borghi, Tom Ziemke, and Gianluca Baldassarre. Theories and computational models of affordance and mirror systems: an integrative review. *Neuroscience & Biobehavioral Reviews*, 37(3):491–521, 2013.

Hugo Touchette and Seth Lloyd. Information-theoretic approach to the study of control systems. *Physica A: Statistical Mechanics and its Applications*, 331(1-2):140–172, 2004.

TN Vikram, Céline Teulière, Chong Zhang, Bertram E Shi, and Jochen Triesch. Autonomous learning of smooth pursuit and vergence through active efficient coding. In *4th International Conference on Development and Learning and on Epigenetic Robotics*, pages 448–453. IEEE, 2014.

Mark T Wallace and Barry E Stein. Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience*, 17(7):2429–2444, 1997.

Binxu Wang, David Mayo, Arturo Deza, Andrei Barbu, and Colin Conwell. On the use of cortical magnification and saccades as biological proxies for data augmentation. In *SVRHM 2021 Workshop@ NeurIPS*, 2021.

Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of research and development*, 4(1):66–82, 1960.

Teresa Wilcox, Rebecca Woods, Catherine Chapa, and Sarah McCurry. Multisensory exploration and object individuation in infancy. *Developmental psychology*, 43(2):479, 2007.

Charles Wilmot and Jochen Triesch. Learning abstract representations through lossy compression of multi-modal signals. *arXiv preprint arXiv:2101.11376*, 2021.

Laurenz Wiskott and Terrence J Sejnowski. Slow feature analysis: Unsupervised learning of invariances. *Neural computation*, 14(4):715–770, 2002.

Tete Xiao, Xiaolong Wang, Alexei A Efros, and Trevor Darrell. What should not be contrastive in contrastive learning. In *International Conference on Learning Representations*, 2020.

Liheng Zhang, Guo-Jun Qi, Liqiang Wang, and Jiebo Luo. Aet vs. aed: Unsupervised representation learning by auto-encoding transformations rather than data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2547–2555, 2019.

Yu Zhao, Constantin A Rothkopf, Jochen Triesch, and Bertram E Shi. A unified model of the joint development of disparity selectivity and vergence control. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pages 1–6. IEEE, 2012.

Xiaomao Zhou, Tao Bai, Yanbin Gao, and Yuntao Han. Vision-based robot navigation through combining unsupervised learning and hierarchical reinforcement learning. *Sensors*, 19(7):1576, 2019.

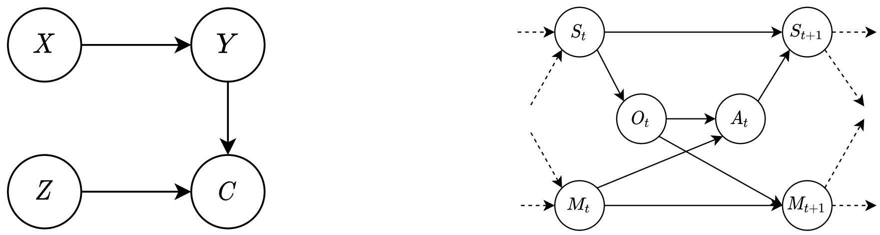

Figure 2: **Left:** Simple example of Bayesian network. **Right:** Simple example of Bayesian network.Bayesian network which sums up the perception-action loop of an agent, adapted from Klyubin et al. [2004]. t refers to the discrete time index, S are the hidden states of an environment, O the observations of an agent, A the ground actions of an agent, M the memory or internal state of an agent.

A Background

In this section, we introduce the basic components of our theory, namely Bayesian networks, information theory and hierarchical reinforcement learning.

A.1 Bayesian networks

Bayesian networks (or graphical models) are directed acyclic graphs for which 1-vertices correspond to random variables which represent part of the state of the system; 2-edges reflect statistical dependencies between these variables, *i.e* a causal relationship. Figure 2 (left) illustrates a simple example of Bayesian network. Given the dependencies, a joint probability $p(X, Y, C, Z)$ is conform to the graphical model if $p(X, Y, C, Z) = p(X)p(Y|X)p(Z)p(C|Z, Y)$. More generally, if $(V_0, \dots, V_N)$ are the $N + 1$ vertices of a graph, and $Pa(V)$ sums up the parents of V with respect to the edges, we can write [Pearl, 2014]:

$$p(V_0, \dots, V_N) = \prod_{i=0}^{N} p(V_i|Pa(V_i)). \tag{2}$$

These models are convenient to model action-perception loops [Touchette and Lloyd, 2004, Klyubin et al., 2004, Levine, 2018] and allow to compute information theoretic measures. In this setting, parameters, actions, states, decisions etc. . . are all random variables. Figure 2 (right) shows the Bayesian network induced by a typical perception-action loop which is unrolled through time. This kind of unrollment is typical through Dynamical Bayesian networks [Dagum et al., 1992]. In the following, while we could also use a Structured Graphical Model [Pearl, 2010], we assume that the standard Bayesian network is simpler and makes our results more reachable.

A.2 Measuring information

The Shannon entropy quantifies the mean necessary information to determine the value of a random variable. Let X be a random variable, its entropy is defined by:

$$H(X) = -\int_X p(x) \log p(x). \tag{3}$$

The entropy is maximal when $p(X)$ encodes a uniform distribution, and minimal when $p(X)$ follows a Dirac distribution. Similarly to the entropy, we can define the entropy conditioned on a random variable Y:

$$H(X|Y) = -\int_Y p(y) \int_X p(x|y) \log p(x|y). \tag{4}$$

Using these definition, we can introduce the mutual information, which quantifies the information that a random variable Y contains about another random variable X:

$$I(X; Y) = H(X) - H(X|Y) \tag{5}$$

The mutual information between two independent variables equals zero (since $H(X|Y) = H(X)$). Similarly to the conditional entropy, we can introduce the mutual information between variables X and Y knowing another random variable W:

$$I(X;Y|W) = H(X|S) - H(X|Y,W) \tag{6}$$
$$= D_{KL}\Big[p(X,Y|W)||p(X|W)p(Y|W)\Big]. \tag{7}$$

Following Equation 7, one can interpret the mutual information as being the difference between the joint distribution of two variables and the distributions of variables assuming they are independent.

It is straightforward to generalize mutual information to several random variables, it leads to the multi-information [Slonim et al., 2001], also named total correlation [Watanabe, 1960]:

$$MI(X_0,\dots,X_N) = D_{KL}\Big[p(X_0,\dots,X_N)||\prod_{i=0}^{N} p(X_i)\Big] \tag{8}$$

It quantifies the information that variables $X_0,\dots,X_N$ contain about each other. As above, it is described as the discrepancy between the joint distribution $p(X_0,\dots,X_N)$ and the same distribution if variables were independent.

A.3 Hierarchical and developmental reinforcement learning

A Markov Decision process (MDP) is defined through S the set of possible states; A the set of possible actions; T the transition function $T : S \times A \times S \to p(s'|s,a)$; R the reward function $R : S \times A \times S \to \mathbb{R}$; $\rho_0 : S \to \mathbb{R}$ the initial distribution of states.

An agent starts in a state s_0 given by ρ_0. At each time step t, the agent is in a state s_t and performs an action a_t; then it waits for the feedback from the environment composed of a state s_{t+1} sampled from the transition function T, and a reward r_t given by the reward function R. The agent repeats this interaction loop until the end of an episode. In reinforcement learning [Sutton and Barto, 1998], an agent aims to associate actions a to states s through a policy π in order to maximize the expected discounted reward defined by $\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t, s_{t+1}$.

In developmental machine learning [Colas et al., 2020], a **goal** is defined by the pair (g, R_G) where $G \subset \mathbb{R}^d$, R_G is a goal-conditioned reward function and $g \in G$ is the d-dimensional goal embedding. We can now define the **skill** associated to each goal as the goal-conditioned policy $\pi^g(a|s) = \pi(a|g,s)$; in other words, a skill refers to the sensori-motor mapping that achieve a goal [Thill et al., 2013]. This skill may by learnt or unlearnt according to the expected intrinsic rewards it gathers. It implies that, if the goal space is well-constructed (as often a ground state space for example, $R_G = S$), the agent can generalize its policy across the goal space, *i.e* the corresponding skills of two close goals are similar.

The approach can be extended to a hierarchy of RL agents, which is then denoted as hierarchical reinforcement learning. Typically, in feudal reinforcement learning [Dayan and Hinton, 1993], an upper-level *manager* learns a policy that assigns goals $g \in G$ to *workers*. To guide the learning process of the workers, the manager rewards them according to their actions. Once a worker achieves its goal, the manager sets another one and the execution loop continues. In practice, we can stack several levels of managers, so that there may be $k > 2$ levels in the hierarchy.

B Time independence

Let us derive of the time-independent mutual information between variables. Assuming we have $p(X_t) = p(X|t) = P(X)$ and $p(X_t|Pa(X_t)) = p(X|Pa(X))$ where X can represent any random variables and Pa represent the time-dependent parents, we have, with $t = 0,\dots,N$:

Symbol	Description
M_{next}	Modules that directly depend on M
M_{prev}	Modules that directly condition M
M_{aprev}	Modules that directly and indirectly condition M
M_{apnext}	Modules that directly/indirectly conditions a module that follows M
$G^M_{\Delta T}$	All goals generated in M between two higher level representations.

Table 2: The notations used to discuss modules.

$$\begin{aligned}
\sum_t I(X_t; X_{t-1}, Y_t) &= \sum_t I(X'; X, Y|t) \\
&= \sum_t H(X'|t) - H(X'|X, Y, t) \\
&= \sum_t -\sum_{X'} p(x', t)\log p(x'|t) + \sum_{X', X, Y} p(x', x, y, t)\log p(x'|x, y, t) \\
&= -\sum_{X'}\sum_t p(x'|t)p(t)\log p(x'|t) + \sum_{X'}\sum_{X,Y}\sum_t p(x', x, y|t)p(t)\log p(x'|x, y, t) \\
&\overset{(1)}{=} -\sum_{X'}\sum_t p(x')p(t)\log p(x') + \sum_{X'}\sum_{X,Y}\sum_t p(x', x, y)p(t)\log p(x'|x, y) \\
&\overset{(2)}{=} -\sum_{X'}\sum_t p(x')\frac{1}{N}\log p(x') + \sum_{X'}\sum_{X,Y}\sum_t p(x', x, y)\frac{1}{N}\log p(x'|x, y) \\
&= \sum_t \frac{1}{N}\big[-\sum_{X'} p(x')\log p(x') + \sum_{X'}\sum_{X,Y} p(x', x, y)\log p(x'|x, y)\big] \\
&= I(X'; X, Y)
\end{aligned} \tag{9}$$

where we applied $p(X|t) = p(X)$ in (1) and noticed that $p(t) = \frac{1}{N}$ in (2).

C Proof of Equation 1

Before startign the proof, let us rigorously define a module and its corresponding notations. A module represents a bio-inspired *building block* [Mussa-Ivaldi and Solla, 2004, Mussa-Ivaldi and Bizzi, 2000, Butz, 2016] associated to 1- a temporal/inter-modal receptive field over incoming observations [Hasson et al., 2015]; 2- an independent part of a lateral decomposition. The temporal receptive field refers to the number of temporally different indirectly processed observations, while the inter-modal receptive field refers to the set of indirectly processed modalities. This generic module is composed of a representation and goal variables (and their respective functions f and π) with the temporally-dependent causal relationships displayed in Figure 1 that makes them interact with each other. Let us define a module formally, assuming causal relationships are consistent over time:

Definition 1. A module M is a set of random variables (G^M_t, R^M_t) with a causal relation $R^M_t \to G^M_t$ and a conditioning module.

Definition 2. A module M depends on an other module M_p according to a temporal scale ΔT if we have for a given T 1- one causal relation $R^{M_p}_{\Delta T} \to R^M_T$ where $R^{M_p}_{\Delta T} = \{R^{M_p}_{T-\Delta T}, ..., R^{M_p}_T\}$; 2- several $(R^M_{T-\Delta T}, G^M_{T-\Delta T}) \to G^{M_p}_t$ for each $t \in [T - \Delta T, ..., T[$. Inversely, M_p conditions M.

Table 2 sums up the notations that derive from these two definitions.

Now, let us recall Equation **??** and start our derivation.

$$MI(B_G) = \sum_{M \in \text{Modules}} \Big[\sum_{G_t^M \in [R^{M_{next}}; R_T^{M_{next}}]} [I(G_t^M; G^{M_{next}}, R^{M_{next}}, R^M)] + \underbrace{I(R_T^M; R_{\Delta t}^{M_{prev}})}_{\text{Compression/Novelty}} \Big]$$
$$+ \sum_{o \in \mathcal{O}} \underbrace{I(O_T^o; O^o, G^o)}_{Controllability}$$

To derive the lower-bound of the *Skill learning*, we mostly take advantage of the data processing inequality (DPI):

We want to lower-bound:

$$\sum_{t \in [T_0; T]} I(G_t^M; G^{M_{next}}, R^{M_{next}}, R_t^M)$$

Thanks to conditional independence, add representations conditioning $R^{M_{next}}$:

$$= \sum_{t \in [T_0; T]} I(G_t^M; G^{M_{next}}, R^{M_{next}}, R^{M_{apnext}}, R_t^M)$$

Apply DPI to condition all goals on the same representations:

$$\geq \sum_{t \in [T_0; T]} I(G_t^M; G^{M_{next}}, R^{M_{next}}, R^{M_{apnext}})$$

Bring together sequential goals into one group of variables:

$$\geq I(G_{\Delta T}^M; G^{M_{next}}, R^{M_{next}}, R^{M_{apnext}}) \tag{10}$$

We separate the term in two parts with the chain rule:

$$= I(G_{\Delta T}^M; G^{M_{next}}, R^{M_{next}} | R^{M_{apnext}}) + I(G_{\Delta T}^M; R^{M_{apnext}})$$

Thanks to conditional independence, we can extend the left part:

$$= I(G_{\Delta T}^M, G^{M_{apnext}}, O^{M_{apnext}}; G^{M_{next}}, R^{M_{next}} | R^{M_{apnext}}) + I(G_{\Delta T}^M; R^{M_{apnext}}) \tag{11}$$

We can aggregate all goals and observations at the current timestep like in Figure 3 [Chang and Fung, 1989]:

$$= I(E^M; G^{M_{next}}, R^{M_{next}} | R^{M_{apnext}}) + I(G_{\Delta T}^M; R^{M_{apnext}})$$

We apply the DPI on E^M to exhibit $R_T^{M_{apnext}}$:

$$\geq I(R_T^{M_{apnext,next}}; G^{M_{next}}, R^{M_{next}} | R^{M_{apnext}}) + I(G_{\Delta T}^M; R^{M_{apnext}}) \tag{12}$$

The first and last lower-bounds becomes narrow when lower-level modules maximize their own objectives. For the second lower-bound we use the fact that $I(X, Y; V|W) \leq I(X; V|W) + I(Y; V|W)$.

Finally, we leave a rigorous proof to future work and hypothesize that the lower-bound in Equation 14 is narrow. It may be because the right-hand term is unlikely to also maximize our *Novelty/Covariance* term.

$$I(R_T^{M_{apnext,next}}; G^{M_{next}}, R^{M_{next}} | R^{M_{apnext}}) + I(G_{\Delta T}^M; R^{M_{apnext}}) \tag{13}$$
$$\geq I(R_T^{M_{apnext,next}}; G^{M_{next}}, R^{M_{next}} | R^{M_{apnext}}) \tag{14}$$

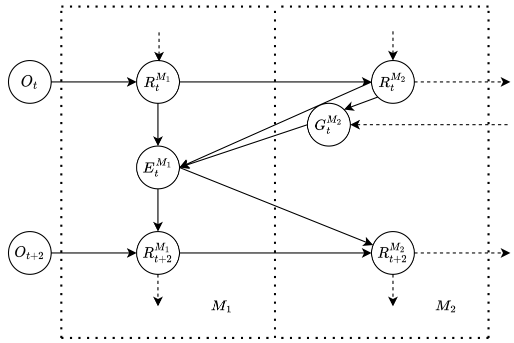

Figure 3: Bayesian network that replicates B, but illustrates how we can aggregate different variables in E.

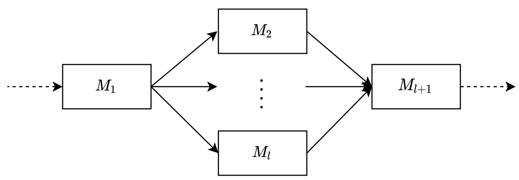

Figure 4: Module-based Bayesian network inducing the disentanglement of a representation.

D Multi-modal Compression/Novelty

According to our objective, a multimodal agent subject to lossy compression favors cross-modal redundant information with its novelty/compression term. Let us formalize this observation. In Figure 1 (right), our agent processes each modality independently and maximizing our *Novelty* amounts to compression maximization. Since our modal-specific modules M_1 and M_2 have a timescale equal to one, our *Novelty* becomes $I(R_T^{M_1}, R_T^{M_2}; R_T^{M_3})$ where M_3 depends on M_1 and M_2. We can decompose it into

$$\begin{aligned} I(R_T^{M_1}, R_T^{M_2}; R_T^{M_3}) &= I(R_T^{M_1}; R_T^{M_3}) + I(R_T^{M_2}; R_T^{M_3}) \\ &\quad - I(R_T^{M_1}, R_T^{M_2}) + I(R_T^{M_1}, R_T^{M_2} | R_T^{M_3}) \\ &= I(R_T^{M_1}; R_T^{M_3}) + I(R_T^{M_2}; R_T^{M_3}) \end{aligned} \tag{15}$$

where we apply the independence assumption in Equation 15. The best way to maximize Equation 15 under lossy compression is to first maximize $I(R_T^{M_1}, R_T^{M_2}) - I(R_T^{M_1}, R_T^{M_2} | R_T^{M_3})$, *i.e* the information shared by $R_T^{M_1}$ and $R_T^{M_2}$ about $R_T^{M_3}$ as it will simultaneously maximize both terms. We will now review this objective through concrete works applied to multimodal agents.

E Disentanglement

Let us consider the architecture in Figure 4, displayed with notations introduced in Appendix C. It displays a typical separate/rebranch pattern without extension of the temporal receptive field. Unlike when compressing multimodal data, we can not assume $M_2, ..., M_l$ to be independent because of their common predecessor. Let us compute the compression term of M_{l+1}:

$$\begin{aligned} \mathcal{L}_{\text{dis}} &= I(R_T^{M_{l+1}}; R^{M_2}, ..., R_{M_{l-1}}) \\ &= \sum_{i=2}^{l-1} [\underbrace{H(R^{M_i})}_{\textit{Individual information}}] + \underbrace{H(R^{M_2}, ..., R^{M_{l-1}} | R^{M_{l+1}})}_{\textit{Global information preservation}} \underbrace{- MI(R^{M_2}, ..., R^{M_{l-1}})}_{\textit{Independency}} . \end{aligned} \tag{16}$$

Interestingly, *Independency* objective has been explored in the context of (Beta-) Variational Auto-Encoder (VAE) [Esmaeili et al., 2019, Chen et al., 2018, Gao et al., 2019, Kim and Mnih, 2018] and independent component analysis [Bell and Sejnowski, 1995, Lee et al., 2000]. These work suggest that it rules the disentanglement of a representation, *i.e* its decomposition into semantically different parts of the data (color, shape, motion, size, orientation . . .).