Physics of Life Reviews 51 (2024) 343–381

Contents lists available at ScienceDirect

Physics of Life Reviews

journal homepage: www.elsevier.com/locate/plrev

Review

Disentangled representations for causal cognition

Filippo Torresan [a], Manuel Baltieri [b,a,*]

[a] University of Sussex, Falmer, Brighton, BN1 9RH, United Kingdom
[b] Araya Inc., Chiyoda City, Tokyo, 101 0025, Japan

ARTICLE INFO

Communicated by J. Fontanari

Keywords:
Causal cognition
Animal cognition
Causal reinforcement learning
Disentangled representations
Disentanglement

ABSTRACT

Complex adaptive agents consistently achieve their goals by solving problems that seem to require an understanding of causal information, information pertaining to the causal relationships that exist among elements of combined agent-environment systems. Causal cognition studies and describes the main characteristics of causal learning and reasoning in human and non-human animals, offering a conceptual framework to discuss cognitive performances based on the level of apparent causal understanding of a task. Despite the use of formal intervention-based models of causality, including causal Bayesian networks, psychological and behavioural research on causal cognition does not yet offer a computational account that operationalises how agents acquire a causal understanding of the world seemingly from scratch, i.e. without a-priori knowledge of relevant features of the environment. Research on causality in machine and reinforcement learning, especially involving disentanglement as a candidate process to build causal representations, represents on the other hand a concrete attempt at designing artificial agents that can learn about causality, shedding light on the inner workings of natural causal cognition. In this work, we connect these two areas of research to build a unifying framework for causal cognition that will offer a computational perspective on studies of animal cognition, and provide insights in the development of new algorithms for causal reinforcement learning in AI.

1. Introduction

Causal cognition, the ability to acquire and exploit causal information about oneself and the world, is a core aspect of adaptive and intelligent behaviour, in both human and non-human animals [1–5]. Artificial systems displaying various forms of seemingly intelligent behaviour have also emerged in the last few years [6,7]. However, these systems still fall short of performing at the level of most non-human animals, which often showcase various kinds of *causal* cognitive abilities [8–11]. It has thus been suggested that an understanding of the mechanisms of causal cognition will play a crucial role in cognitive science and artificial intelligence for the next decades [12,8,9,11,13–16].

The study of causal cognition both in human and non-human animals has a long history, with roots in behavioural studies trying to establish the extent to which an organism's behaviour reflects proper causal understanding of the world instead of simpler forms of associative learning [17,1,18–22,4]. Some of the most influential studies in this area have combined theoretical and modelling work based on the formalism of causal Bayesian networks to account for the cognitive performance of various subjects in tasks designed to test causal cognition abilities [23,24,18,19,25,20].

* Corresponding author at: Araya Inc., Tokyo, Japan.
E-mail addresses: f.torresan@sussex.ac.uk (F. Torresan), manuel_baltieri@araya.org (M. Baltieri).

https://doi.org/10.1016/j.plrev.2024.10.003
Received 9 October 2024; Accepted 14 October 2024

Available online 21 October 2024

Unlike research on learning from reinforcement in behavioural psychology, which served as foundation to the mature and rich theory of modern reinforcement learning [26], the wealth of experimental findings in causal learning has not yet been translated into a set of computational principles for a coherent, pragmatic framework showing how agents *acquire* a causal perspective of the world by acting in it while equipped with little or no knowledge of the relevant features of the environment. In other words, most approaches set aside one of the core questions about causal cognition: how are causal models acquired, or *learnt*, by agents when they are not endowed with a priori knowledge of the structure of a model? Mainstream Bayesian frameworks often overlook this question because they tend to describe processes of causal inference *with a model* [27], i.e. assuming that the cognitive capacities of agents in a certain context can be assessed using a model encoding causal variables and relationships postulated by a scientist with a priori knowledge of the causal structure deemed necessary in that context. This descriptive approach however fails to provide a clear (and testable) account of how a causal viewpoint can emerge from an agent's first-person experience, i.e. *within a model* [27], or as the machine learning analogy goes, *from pixels* [28–30,6,31].

On the other hand, the agent perspective focusing on computational models that implement processes of causal learning and reasoning from first-person experience is emerging as a dominant area of research in the field of artificial intelligence (AI), following a surge of interest in causality by the machine learning and deep learning communities [32–34]. In these areas, several works have proposed new unsupervised methods to learn causal structure from data (causal structure learning or causal discovery) [35–40], while others have designed new reinforcement learning algorithms based on some of the principles and ideas developed in causality research [41–49]. Others have also recognised the importance of causality for robotics [50,51] and new benchmarks and datasets have been introduced to study more rigorously causal inference in AI [52–56,10].

In this work, we introduce a unifying treatment of research in causal cognition in non-human animals and causal machine learning. Our focus will be almost entirely on non-human animals, exploring "lower-level" causal capabilities that we believe provide at the moment a more fertile ground for a comparative study of current AI systems' capabilities and the skills of causal reasoners in the animal world. Towards the end we will briefly touch on possible extensions of the current proposal to human cognition, with "higher-level" causal skills involving language, abstract reasoning and various other abilities that at the moment still seemingly escape most non-human animals and AI systems. Our goal here is to lay the foundations for a research program where 1) computational work in causal machine learning can help us to gain clarity on the principles driving the acquisition of causal knowledge in adaptive agents, and where 2) causal cognition can in turn inspire the development of new algorithmic implementations in causal machine learning. In particular, we will provide an explicit blueprint for a theoretical and computational framework centred around the notions of disentanglement and causal representation learning [33,57,58], with relevant connections to the process of functional specialisation in neural cells [59–61], that can form a more rigorous basis for conceptual characterisations of causal cognition in terms of explicitness, sources and integration of causal information [62–66,5].

In what follows, we will adopt a pragmatic notion of causal information, understood as information about the nature of causal relationships describing the causal structure of the environment and the effects of an agent on different variables, respectively physical and difference-making information [67,65,64].

Our proposal will bridge the gap between studies of causal cognition, on one side, and mathematical and computational models of adaptive behaviour, on the other, allowing hypotheses about the nature and emergence of different constitutive blocks of causal learning and reasoning to be tested using the power of modern causal machine learning models. By connecting classic works on causality and causal cognition with recent developments in machine and reinforcement learning, the proposal on offer moves beyond a descriptive approach on how non-human and human animals display and use causal knowledge [5], shifting the focus instead on what could be the computational principles that promote *learning* of essential causal models/representations in a simple agent that interacts with an environment.

In section 2 we will start with an overview of the literature on causal cognition in non-human animals, briefly going through some of its most relevant landmark studies leading to modern proposals addressing causal understanding. Section 3 will go through a conceptual framework proposed to unify different accounts of causal understanding and developed in order to characterise causal understanding in terms of causal information on a more fine-grained scale that includes three dimensions: explicitness, sources and integration of causal information. Using some of the formal work presented in section 4, collecting the useful notions of disentanglement, structural causal models and (partially observable) Markov decision processes, in section 5 we will see how it is possible to operationalise computationally causal understanding in modern work on deep (reinforcement) learning. Section 6 will provide a comparative analysis of work on natural (animal cognition) and artificial agents (machine/reinforcement learning) showing common areas of interest between causal cognition and causal machine learning and highlighting the main differences between the two. Finally, in section 7 we discuss how these two lines of research could benefit from each other's insights moving forward, speculating on proposals that combine them in new ways.

2. Causal cognition

2.1. Causal cues and the debate on associative vs. cognitive explanations

Early work on causal cognition in human and non-human animals focused on how subjects learn about the strength of cue-reward relationships, where some of the given cues could be attributed the "causal power" of eliciting rewarding outcomes [17,68,69]. A significant part of this early work can be contextualised within an old debate on whether causal learning is just a form of associative or contingency learning, the dominant theoretical framework to study animal learning [70], or a form of learning that requires more cognition-laden processes.

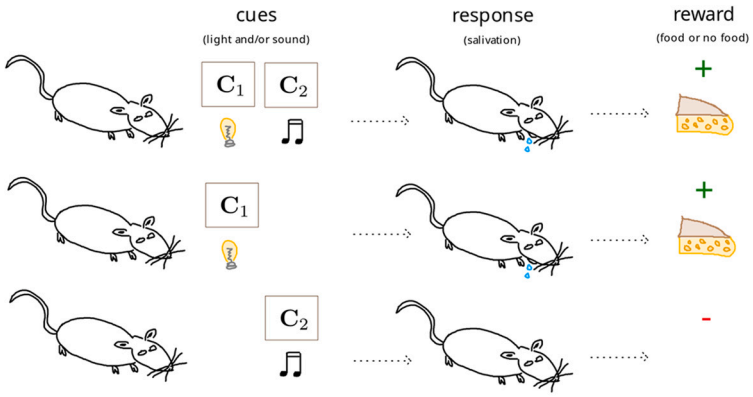


Fig. 1. An illustration of the phenomenon of backward blocking in rats. Subjects are conditioned to elicit a response (salivation) to a stimulus (presence of food) by means of a compound cue (light + acoustic tone) as well as a single cue (light). When tested with the other cue (acoustic tone), the rats do not react as strongly as if they understood that in the compound-cue trials the only cause of the reward was the first cue (light).

The crux of this debate was not whether there are causal relationships or structures in the world, which is a more philosophical type of issue. Granted that there are, associative accounts have usually tried to show that the successful performance of some agents in seemingly causal learning tasks can be modelled, and ultimately explained, by means of purely associative learning mechanisms. Roughly, these would track the relevant causal associations between certain variables without the need of invoking more sophisticated cognitive processes or structures involving a notion of causality [71–74]. In contrast, other works have highlighted how certain behavioural responses, especially from humans, are indicative of **causal models** that the agent in question would exploit to reason about causal relationships of various sorts [75–78].

A paradigmatic example of how this debate has typically unfolded can be found within the analysis of *backward blocking* in conditioning experiments with rats. In these experiments, test subjects are exposed to cues, say $\mathbf{C}_1$, $\mathbf{C}_2$, and the compound cue given by their combination, i.e. $\mathbf{C}_1\mathbf{C}_2$, that may or may not lead to a rewarding outcome, indicated by + (see Fig. 1). After some trials, the subjects, often rats, learn about the relationships between those cues and the reward, and react accordingly when similar cues are shown. In the backward blocking scenario, after rats have been trained with cue-outcome sequences like $\mathbf{C}_1\mathbf{C}_2+$, $\mathbf{C}_1+$ (in that specific order), one finds that they don't react to the presentation of $\mathbf{C}_2$ alone in subsequent trials. The response to $\mathbf{C}_2$ is "blocked" after witnessing $\mathbf{C}_1+$ because the causal role of $\mathbf{C}_2$ is reconsidered in light of evidence that suggests it was not involved in producing the reward.

This sort of retrospective evaluation (of what happened in a earlier trial) is a problem for associative accounts because they usually assume that a cue-outcome association can increase or decrease in strength only when the cue is present (together or without the reward). However, in backward blocking scenarios the change in behavioural response to $\mathbf{C}_2$ occurs following the presentation of $\mathbf{C}_1+$ and despite the fact that $\mathbf{C}_2$ has always (or most of the time) appeared at the same time as the reward +. An advocate for causal models would see retrospective evaluation as an example of their influence on cue-outcome learning. Given the evidence represented by $\mathbf{C}_1+$, the decreased response towards $\mathbf{C}_2$ could be explained in terms of a re-evaluation of the role played by that cue when it appeared as part of the sequence $\mathbf{C}_1\mathbf{C}_2+$. Such evidence would suggest that $\mathbf{C}_2$ did not have a part in the causal relationship producing the rewarding outcome.

At the same time, over the years several revisions of traditional associative accounts to model retrospective evaluation have been proposed to account for backward blocking and more complex scenarios involving higher-order relations between cues, i.e. relations between two cues that never occur together but that appear in combination with another cue, see for instance [73]. However, these revisions usually depart from traditional associative principles in significant ways, e.g. requiring more sophisticated information-processing capabilities, see for instance the discussion in [22] and references therein. This thus leaves us with architectures based on, or inspired by associative principles [79], begging however the question of whether these models are still associative, or not.

More recently, the whole dichotomy between associative and cognitive explanations has been put into discussion, as different works argue that empirical and theoretical grounds for such a distinction are too weak [22,80–82]. Indeed, one could point out that the whole debate is a consequence of a narrow view of what counts as cognitive, e.g. exclusively representations and processes that have an "internal semantic or propositional structure" and that shape thought and behaviour via "structured inference" [72], hence a legacy of cognitivism. It is this narrow and demanding view of what defines a cognitive representation or process that generates an unhelpful contrast with simpler and more basic ones in the explanation of mental faculties.

Focusing on the friction between these two views makes us blind of the possibility that cognition ought to be considered on a spectrum. To witness, more recent perspectives have often extended the realm of what counts as cognitive [83–85] to more basic, low-level biological processes, or have described cognition, from perception to action and higher-order functions, as primarily a matter of Bayesian inference on different spatiotemporal scales [86–88]. Pursuing this line of reasoning suggests then that the contrast between associative and cognitive accounts could be simply reframed as that between lower-level and higher-level cognitive processes placed on a common spectrum. In this view, the fundamental questions become where processes regarded as part of causal cognition can be found on the cognitive spectrum, whether they can be characterised in terms of simpler building blocks, and how they might interact with each other to realise some form of causal understanding.

2.2. Causal understanding as a building block for causal cognition

Moving past the associative vs. cognitive debate using a more comprehensive definition of cognition at different levels has however brought forward a perhaps more fundamental dispute about the presence, or not, of forms of **causal understanding** in agents, and what such an understanding amounts to. This is especially evident in the behavioural research on causal cognition in non-human animals, where the goal is to design behavioural tasks specifically intended to try and measure some manifestation of causal understanding [17]. In other words, tasks that would ascertain whether a subject is capable of feats of causal cognition, where the assumption is that a solution of the task requires certain causal, cognitive strategies.

An example of this research is represented by studies on capuchin monkeys using the trap-tube task [89–92], where causal cognition is characterised as the comprehension of key cause-effect relationships within the task. The trap-tube task consists in pushing a food reward out from a transparent tube (anchored to the floor) using some kind of tool (e.g. a stick), by inserting it into one of the tube's two openings, see Fig. 2a. In general, capuchin monkeys struggle to learn how to solve the task, either because they would pick the wrong kind of tool (a stick that could not be inserted into the tube because of its shape) or because they would pick the wrong side to put the stick, making the reward fall into a trap positioned underneath the tube.

The persistent error patterns of the (few) subjects that could solve the task after extensive trial and error are thought to be evidence of a distinction between 1) successful performance based on a "stroke of luck" after extensive active experimentation, and 2) successful performance based on an understanding of relevant causal variables inherent within the task requirements [89,91]. It is in fact well known that capuchin monkeys have a propensity to produce a wide variety of actions and complex combinations thereof, even involving tools, to the point that they could be described as expert tool-users. Because of this, it is unclear whether they have an appreciation of the causal relationships between their behaviour and the resulting outcome. In other words, they might learn that using certain tools is an effective way to achieve certain results, but they may not appreciate the reasons for why their actions are successful [91].

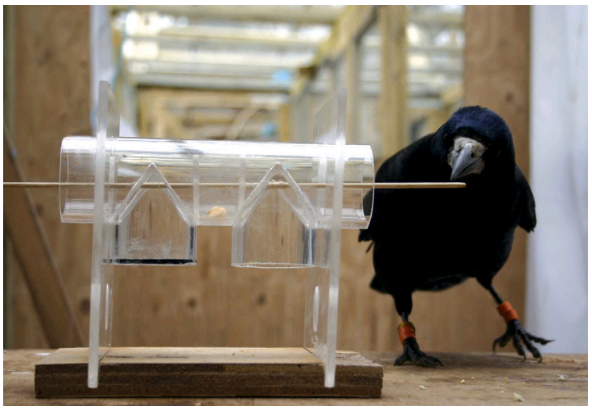
In contrast, experimental evidence in chimpanzees suggests that they may have an understanding of the causal relationships between certain actions and their associated outcomes [93–95]. The key finding here is that some subjects, tested with different configurations of the trap-tube task, were able to select the right side of insertion (almost) immediately, allegedly displaying an ability to plan their actions according to the different causal relationships present in the task configuration. Consequently, this evidence suggests that the successful subjects were not using heuristics such as a distance-based rule, which would for example determine the correct action based on the distance of the reward from the tube's openings without an understanding of the causal structure of the problem. Instead, subjects appeared to take into account the causal features of the task configuration and choose beforehand what action to perform. This would thus amount to a representational strategy that delineates the key requirements of the task in advance and results in the correct behaviour without the need of extensive trial-and-error learning. More specifically, one could argue that those successful chimpanzees exhibited some kind of causal understanding of the consequences of pushing the stick inside the tube (but see [96,95] for different views). Unlike for instance the capuchin monkeys of other experiments [89,91], where a constant monitoring of the effects of one's ongoing action (to check one is on the right track) and attempting a variety of actions' combinations (in the hope to find the right one) were instead necessary.

In a subsequent review paper, Visalberghi and Tomasello [97] argue that an organism can be said to understand causal relationships, or to posses a causal understanding of (parts of) the world, if and only if they are able to see or posit some mediating forces (or variables) between two associated events. This kind of explanatory attitude is informally described as the key component of causal understanding, one that helps an organism to envisage and navigate the **web of causal possibilities** of the "how" and "why" events at different points in time may be connected. This kind of understanding has an impact on the behavioural strategies an agent might pick to reach certain goals, as novel ways of manipulating the environment are disclosed, targeting those specific mediating forces [97]. In this view, an organism equipped with causal knowledge is therefore capable of dealing with unexpected signals from the environment in a farseeing way, i.e. by taking into account different possibilities in advance [97]. Following this conceptual definition, it is thus worth focusing on the relation between causal understanding and behavioural strategies, and more specifically on whether agents can plan ahead how to obtain action-driven outcomes arising from a mastery of the causal structure of the world as opposed to a more basic perception-based understanding of only the causal structure governing one's observations (cf. the distinction between model-based and model-free reinforcement in section 6.1.1).

2.3. Causal interventions and tool use

A strong candidate for the presence of plans based on action-driven outcomes is the ability to produce a **causal intervention**, an action that involves causal control on a particular effect [97]. On the one hand, this seems to provide strong evidence for causal cognition since producing an intervention requires some form of causal understanding. In particular, it requires an agent to understand that its actions, in the form of movements of its own body, could be used as external probes for the causal texture of the world (*cf.* second rung of the *causality ladder* in [9]).

On the other hand, the attribution of causal interventions to cognitive agents appears still controversial because there are only limited reports that hint at intervention-like abilities in, for instance, rats [78,98] and primates [99,5]. At the same time it is unclear whether **tool-use**, the ability to skilfully manipulate objects, common in species like corvids [106,100,101], should count as a form of intervention or not. In general, it is not entirely obvious what the markers of causal interventions are and how to design experiments that could determine their presence or absence.



(a) **A variation of the trap-tube task**. An animal subject learns to retrieve a reward from a transparent tube by pushing it out by means of a stick. In this variation of the task, the crow has to slide the reward over the open hole (the one on the right). Image credit: [106] under Elsevier user license.

(b) **Floating-reward task**. An animal subject learns to drop stones into a water-filled cylinder to raise the water level and reach a piece of food. Image credit: [105], adapted under the terms of the Creative Commons Attribution License.

Fig. 2. Causal cognition tasks.

Work on rats, for example, suggests that these animals can learn a common-cause model, where a light being turned on is perceived as the cause of two effects, a noisy sound and the release of some food. After exposure to patterns of causal relationships for a certain number of (training) trials, the rats enter a test condition characterised by a lever that produces a noise when pressed. Interestingly, it has been reported that after (accidental) lever presses, rats exhibit a less resolute search for food (measured by the number of nose-poking in the cage's hopper) than when the noise is presented alone. A possible explanation for this behavioural response would regard these rats as capable of recognising their action (the level press) as an intervention, an independent self-generated perturbation on one variable of the learnt causal model. In fact, an effect (noisy sound) cannot be an indication that a cause is present (light) when that effect is produced by an intervention (lever press). Therefore, by conceiving of their action as an intervention on one variable of the learnt causal model, the rats do not expect that the other effect (food release) will occur, which then induces a less vigorous search for food [78]. While these findings are consistent with the claim that rats can differentiate between predictions based on observations and predictions based on interventions, they do not exhaustively prove that rats can produce interventions to activate a certain causal path, in this case the one leading from the light to the food dispensation (as discussed in [78]).

Work on corvids on the other hand, see for instance [102], testing New Caledonian crows with a few variations of the trap-tube task (see Fig. 2a), suggests that they possess critical causal understanding abilities, e.g. an appreciation of causal relationships involving object-hole interaction, on which their exceptional tool-using skills might be built. Similarly, [103,104] report positive results on a task (inspired by Aesop's fables) in which crows have to learn to drop some objects (e.g. stones) into the right water-filled tube so that the water displacement brings a floating reward (e.g. a piece of meat) closer to the tube opening (see Fig. 2b). The results here point at the fact that the birds managed to solve the task, seemingly by attending to the relevant causal information, e.g. the fact that larger and not hollow objects will produce a bigger water displacement. For a variation of the task however, where the setup instead consisted of three water-filled tubes arranged in a row on a wooden board, with some space between each other, results were less clear. In this task, the baited tube is the one in the centre and, crucially, it is connected with one of the others by means of a U-shape tube hidden from view (located underneath the board). Dropping objects into one of the lateral tubes will have as an effect a water-level rise in the baited tube. Since the central tube is too narrow to drop anything in it, to bring the reward on the surface it is crucial to recognise this counter-intuitive effect and exploit it, i.e. to infer the presence of and reason about **hidden causal mechanisms**. Here, all tested birds performed at chance level, meaning that they dropped objects randomly on either of the two lateral tubes [103], see also [104] for similar conclusions on a slightly modified setup.

All this evidence points at the fact that causal understanding could be a key notion for a more systematic study of causal cognition. At this stage, however, debates on its presence, role and features make it unsuitable for more practical investigations (see, e.g. [105] for evidence that crows' performance on floating-reward tasks is affected by pre-existing preferences on which object to pick, and [107,108] for a critical meta-analysis of the works on corvids described above). After reviewing some of the main themes driving current research on causal cognition, moving away from the debate on associative vs. cognitive and embracing the challenge of determining what constitutes causal understanding and how to infer it from behavioural experiments (e.g. using tasks involving causal interventions or tool-use), we thus next look at how a large body of research in these areas has been recently organised in a new conceptual framework attempting to capture the more fundamental dimensions of causal understanding.

3. A conceptual framework for causal cognition

After decades of research in the field, by now it is evident that different works on causal cognition have often appealed to different conceptualisations of the subject matter, to the point that a consensus has yet to be formed on what even constitutes causal understanding, see [22,4] for some reviews, and the contributions to [1,3] for other perspectives. As briefly illustrated by the previous

section, different lines of work place causal understanding at different levels of a hypothetical cognitive spectrum, and have portrayed very diverse views on how to characterise it in terms of key cognitive functions and behavioural outputs. Most researchers might agree on the idea that a causal agent has the kind of behavioural flexibility that is unattainable by agents lacking causal understanding. Yet, the variety of positions trying to describe what underpins it appears to only contribute to the confusion.

Some works point to a representational strategy for agents to picture in advance what the key causal features for solving a task are. This would then characterise a distinction between performing and understanding, i.e. whether the solution of a task is achieved via some sort of shortcut, or in a robust and reliable way [89–91,93]. Others are more demanding, and see causal understanding as the result of some form of **causal reasoning**, yet another ambiguous expression that has been described in different ways. For instance, causal reasoning could involve structural or symbolic (causal) knowledge abstracted from perceptual cues [109,110,3]. In other words, it would underpin an ability to search for cause-effect relations that reveal how and why two events are connected, or why some actions lead to certain outcomes (i.e. diagnostic causal reasoning), requiring thus the presence of some causal beliefs [97,111]. Others would further maintain that without an ability to perform causal interventions, perhaps even involving unobservable entities (see hidden causal mechanisms in section 2.3), it is unlikely that an agent is able to grasp causality as opposed to just behave *as if* it does. Going back to tool-use then, the question of whether adaptive tool use may reveal the presence of some of the abilities just described or whether it may be a confounder instead [95,100,101] remains unanswered.

In an attempt to put causal cognition research on a more precise and coherent footing, we find different proposals discussing experimental findings framed with respect to a few recurring themes, drawing attention to key aspects of causal cognition [63–66,62] (see also [5] for a recent review). Starzak and Gray [62] in particular dissect the main disagreements over causal understanding in non-human animals, proposing a more precise way to study causal cognition using a three-dimensional conceptual framework inspired by and overall consistent with other conceptual treatments [63–66]. The starting point is to regard causal cognition as the processing of causal information, which is to be conceived of as any piece of evidence, or data (acquired via experience), that discloses aspects of the causal structure of the world. This includes physical aspects, indicating how objects/events are connected and interact mechanically, and difference-making aspects, for example how much impact a system, or agent, acting on a certain variable can have on another [67,65,64]. The latter play a predominant role in an intervention-based account of causality working with structural causal models [112], which are indeed a core part of the stance we take throughout the paper, and thus help make the idea of causal information more concrete. However, we note that the debate on the exact relation between interventions in a causal model and actions (or policies) of an agent remains open (*cf*. distinction between acting and intervening in [5], and discussion in section 2.3). [1]

One of the advantages of this move is to put on one side normative issues, e.g. what really counts as causal understanding, and instead focus on empirically tractable parts of the matter [62]. To a first approximation, the main idea is to score the performance of subjects on causal tasks (see section 2) along three dimensions that have the potential to cover the full spectrum of causal cognition. With these as a way to ground the discussion of different empirical results, Starzak and Gray [62] then suggest ways to draw a more fine-grained comparison of the extent to which non-human animals and humans process causal information. More specifically, following Starzak and Gray [62], causal information processing can be characterised along three dimensions:

1. the level of **explicitness** of causal information,
2. the **sources** of causal information, and
3. the level of **integration** of causal information.

3.1. The explicitness of causal information

The explicitness dimension, refining some intuitions presented in [65], aims to capture a spectrum of causal information where on the one end, **implicit models** are essentially blind to causal relationships. These models represent cases where actions and outcomes/rewards are entangled or "fused" [65], i.e. models based on associative correlations where the causal structure (see the web of causal possibilities in section 2.2) is essentially hidden and inaccessible to the agent. In this class of models, agents cannot necessarily come up with a complex plan on how to adjust their actions that is sensitive to the web of causal possibilities in order to achieve a certain goal, since they lack or have a limited understanding of their own actions and other variables in the environment as causally relevant for bringing about certain outcomes or rewards [65]. They can however take actions in a less structured way, for example using knowledge acquired from repeated trial-and-error in an associative manner, leading to a continuum of explicitness that is apparent in several experimental studies as seen in section 2. For instance, associatively pairing the action of pressing a button (cause) with the presence (very often, but not always) of some food (end goal, an effect) can be considered as an example of implicit model.

On the other side of the spectrum we find **explicit models**, models where the causal structure is completely unpacked and relations between actions and outcomes/rewards are disentangled and available for an agent to take advantage of. Looking at the previous example, we can imagine a different scenario where an agent realises that a button press will activate a food dispenser mechanism and that the food will become available if and only if there is no obstructing object in the mechanism. In this case, the action of clearing the dispenser from the obstructing piece is an action that can be said to require a more explicit understanding of the causal

[1] A more detailed analysis would first require an appreciation of the fact that information itself can be defined in different ways in cognitive science, e.g. by means of a distinction between Shannon information and semantic information [113]. With respect to the framework of structural causal models, there have been recent attempts at formulating a causal version of Shannon's notions of entropy and information gain, quantifying how much control a certain variable could provide over another [114,115]. In general, the notion of causal information we consider in the manuscript can be seen more closely related to these definitions.

structure of the world, at least compared to the first situation, an operation that is directed at altering one of the intermediate causal variables, i.e. the object obstructing the food dispenser, so as to obtain the food.

A qualitative description of explicitness thus amounts to establishing what an agent can do with the acquired causal information, for example by investigating an agent's degree of flexibility in using what it has learned in a causal task (e.g. clearing the dispenser mechanism of the obstructing object). To a first approximation, the key idea is that the more explicit a model is, the more causal information is available to an agent, because the means to reach a certain goal have been recognised as distinct from each other and from the goal itself (the mediating variables of a certain causal influence have been identified, *cf.* [112]), thereby leading to a higher degree of flexibility in behavioural responses.

To see how different degrees of explicitness appear in the animal cognition literature, we can take a closer look at the research on the trap-tube task described in section 2.2. Facing the trap-tube, an agent that can only form implicit models where actions and outcomes are entangled, i.e. where actions $\mathbf{a}$ leading to states $\mathbf{s}$ are a-causally associated to outcomes $\mathbf{x}$, has a very limited ability to discern the possibly relevant intermediate variables that could be exploited to reach the goal. These include for instance the position of the trap, necessary for an understanding of whether its opening (the hole) is on the tube's lower surface or not, i.e. whether it affects the desired outcome (recall that if the opening is on the top surface, e.g. due to a rotation of the tube, then the trap is ineffective).

Since the literature on these experiments suggests that most capuchin monkeys do not seem to appreciate the relevance of the trap, and consequently perform poorly when the tube is inverted, we could say that these agents rely only on implicit models of the form "*insert the stick, out comes the reward*" [91]. While one could say that this associative rule encodes some causal information, it is evident that this only happens in a very implicit and vague manner, leading to maladaptive behaviour in most other contexts, especially without retraining.

In contrast, causal representations capture the relevant difference-making relationships present in the task at hand, e.g. the role that the trap plays in the trap-tube. Such relationships express fine-grained information that shows the different causal links between instrumental/intermediate variables and outcome/reward variables, enabling more flexibility in planning or action selection.

More in general, explicitness can be evaluated by assessing an agent performance in contexts where some kind of knowledge transfer is required. One scenario might involve adjusting a learned strategy to solve a similar task. For example, an agent might have learned that a rake can be used to fetch a coconut that fell into a pond without getting wet. But the rake could also be used to detach a coconut from a palm tree without having to climb to its top, and make it fall on the ground for easy retrieval and tasting. One could also imagine a more drastic context change requiring completely different means for performing successfully in the same task, e.g. if the rake is not available, the subject could look for a similar or different object that may play the same functional role.

Another scenario might instead require using previous knowledge or learned strategies to solve new tasks, e.g. exploiting the same means for a different end. For example, if an agent has learned that the rake can be used to bring desirable items closer, the same agent should be able to use the rake in other contexts, e.g. to retrieve other types of food. Finally, one might think of scenarios where knowledge transfer also demands paying attention to different functional properties of the same means (or others) because such properties are now relevant to the solution of a different problem. In this case, according to Starzak and Gray [62], the cognitive capacities in question would amount to some form of insight learning. For example, after having retrieved the coconut, our agent could use the rake to discourage greedy conspecifics or other animals from stealing the just-earned meal.

3.2. Sources for learning causal information

The sources dimension of Starzak and Gray [62] is also inspired by Woodward [65], where causal learning for an agent is proposed to involve information about both the agent itself and other systems, leading to a distinction between egocentric and non-egocentric sources of causal information with several consequences for animal cognition studies, see for instance [116,117].

Egocentric causal information captures the idea of an agent that can acquire an understanding of the causal structure of the world from its own behaviour, focusing on the effects (desirable or not) of the actions they can control and perform. For example, a subject could learn that performing action $\mathbf{a}_1$ makes a difference for obtaining outcome $\mathbf{x}_1$ but not for $\mathbf{x}_2$. This is the realm of instrumental conditioning (or learning) investigated extensively in animal research [65].

Non-egocentric causal information can, on the other hand, be obtained from two main external sources: the behaviour of other agents and the unfolding of natural events. **Social causal information** implies that an agent can learn about important action-outcome contingencies by paying attention to other agents' behaviour, possibly with limited form of interaction to influence the behaviour of those agents. For instance, observations of a conspecific performing action $\mathbf{a}_1$ and reliably obtaining outcome $\mathbf{x}_1$, but not $\mathbf{x}_2$, provide important causal information for a subject that is aiming at outcome $\mathbf{x}_1$ (or $\mathbf{x}_2$).

Natural causal information similarly suggests that events in the natural world, in principle uncontrollable from the perspective of an agent as in the case of the weather or, say, gravitational force, can disclose ecologically important causal relationships. For instance, observing a piece of fruit falling from a tree-branch shaken by the wind could reveal to an attentive observer important causal information on how to get some food, as long as it is capable of performing a causal analysis of the situation [118]. Compared to the acquisition of the previous types of causal information, natural information imposes, arguably, a higher cognitive load on the subject because the event in question does not tell the subject what action may be (causally) relevant and for what reason(s). In other words, there is an additional cognitive effort that the subject needs to make to appreciate that, given certain observations, action $\mathbf{a}_1$ can produce outcome $\mathbf{x}_1$.

Empirical evidence so far suggests that egocentric causal learning is the most widespread in the animal kingdom [62,5], whereas social causal learning and natural causal learning (the latter called "observational" in [62]) are fully present in adult human beings only.

Importantly, an agent that appreciates this broad demarcation is able to recognise that portions of the sensory stream are to be attributed to entities other than itself, in the sense that they are *about* those entities, despite the fact that all sensory information is acquired via the same sensory channels. In section 5, we will point out the extent to which both the source dimension and integration dimension (described next) are intimately linked to explicitness, which we will be understood as disentanglement. This will also suggest a way an agent has to distinguish between the three sources, i.e. by a learning a disentangled representation, despite all information being acquired through the body (in a "egocentric" manner).

3.3. The integration of causal information

Integration appears more ambiguously in [62] but can be understood, in general, as consisting of operations of update, combination, extension, etc. of one source of information with another one to form a coherent structure. More concretely, in our framework integration will be later framed in terms of meaningful combinations of different sources of information (egocentric, causal and natural) that can describe if and when agents are capable of translating observations from different perspectives into their own (egocentric) perspective, forming **egocentric + social causal information**, **egocentric + natural causal information** or **complete causal information** (egocentric + social + natural), or whether social and natural information can be integrated without a direct effect on the agent own egocentric perspective to form **social + natural causal information**.

While laying the foundations for a formal characterisation of causal cognition across the animal kingdom, the inherent ambiguity of not only integration, but of explicitness and sources too, mixed with the more general focus on high-level discussion over a clear operationalisation, pushes the idea that the conceptual framework proposed by Starzak and Gray [62] is in several ways still heavily relying on interpretative work to be done by the reader. In the next section we fix a few core ideas in a mathematical language that will form the basis of our proposal to formalise Starzak and Gray [62]'s conceptual framework in a pragmatic way in section 5.

4. A mathematical framework for causal cognition

4.1. Disentanglement in machine learning

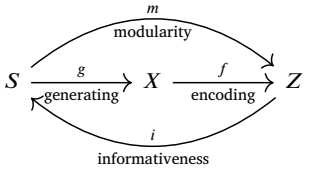
A key proposal in modern approaches to deep (reinforcement) learning is that of **disentanglement** [119], roughly stating that in order to acquire an understanding of the causes behind some given observations, it is necessary to interpret those causes as distinct (high-level) factors, and recognise the different causal power they exert when giving rise to observations [33], see [120] for a recent review. For example, if we see a red ball made of rubber bouncing on the ground, what makes it bounce? While different factors including colour, shape, and material, are intertwined and together produce observations captured by our eyes, some of these factors have no causal influence on the bouncing behaviour, i.e. colour. According to the disentanglement hypothesis, the ability to discern different factors is thus a crucial step in a theory of causal understanding.

Similarly to other influential proposals, disentanglement has been used to characterise a general intuition based, however, on different implementations and interpretations. Following Zhang and Sugiyama [121], we thus look at some of the common structure behind different definitions of disentanglement. To do so, we focus in particular only on sets and functions between them. This is technically equivalent to stating we are working in the category of sets and functions, **Set** [122], and while in various ways limiting, this allows us to focus on the central parts of our proposal to connect disentanglement to explicitness down the line using a relatively simple mathematical toolkit (sets and functions, without introducing more advanced tools from category theory).

We start by defining S, X, Z as sets of (generative) **factors**, **observations** and **codes** respectively. The sets of factors and codes are further assumed to be (Cartesian) products of $n \in \mathbb{N}$ factors and $l \in \mathbb{N}$ codes:

$$\begin{aligned} S &= S_1 \times S_2 \times \cdots \times S_n \\ Z &= Z_1 \times Z_2 \times \cdots \times Z_l \end{aligned} \tag{1}$$

At a high level, the main idea driving this framework is that representing distinct factors, i.e. having disentangled codes that faithfully map to disentangled factors, is the starting point for acquiring a causal understanding of the world: an agent with knowledge of what factors generated its observations is an agent that understands what data-generating mechanisms brought observed data to its sensory peripheries, see Fig. 3 for a way to frame the example of the red ball in this initial setup, to be unpacked next. The setup for disentangled representations can then be captured, in a compact form but with more specific constraints to be imposed below, by the following commutative diagram:



S : factors
X : observations
Z : codes (2)

We introduce next the formal definition of disentanglement that we will be referring to throughout this work, following Zhang and Sugiyama [121].

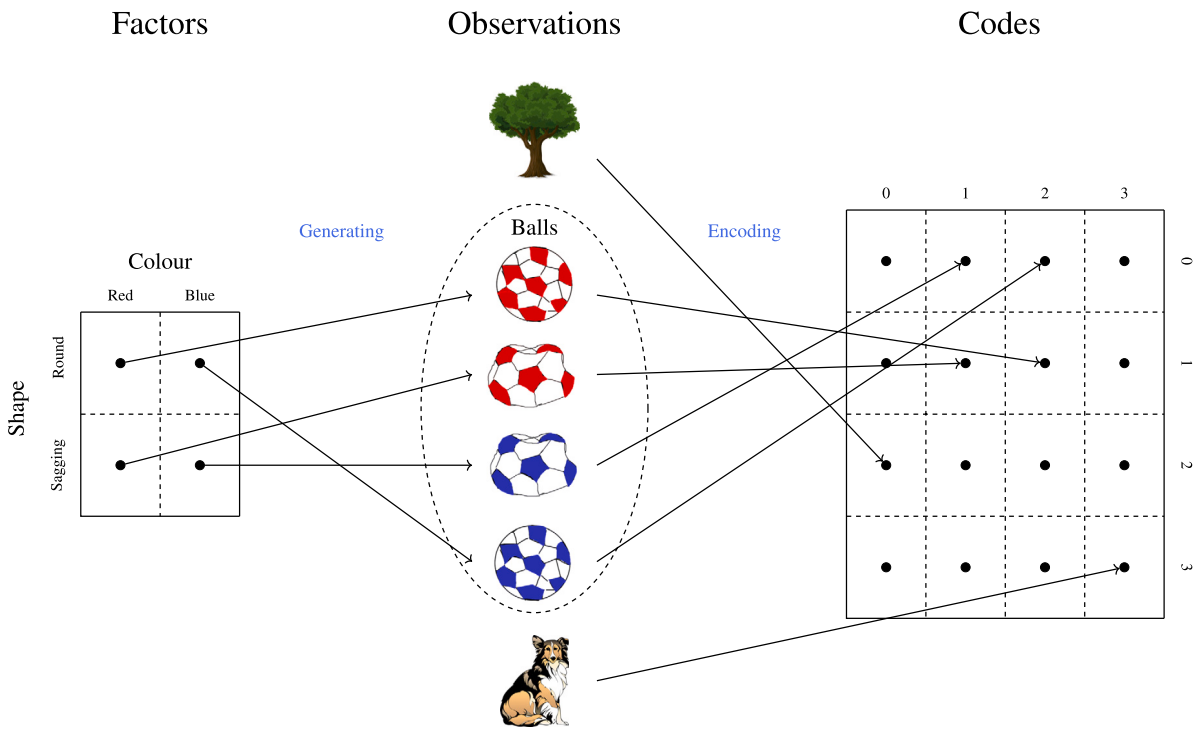


Fig. 3. Disentanglement for a bouncing red ball. Following the example of a system whose goal is to understand what factors cause a ball to bounce (colour, shape, material, etc.) described in the text, we sketch here (some possible) factors, observations and codes for a system disentangling factors from observations into codes. For such a system to understand what makes a red ball bounce, we consider shape and colour as factors generating observations about different balls (round and punctured balls, and of different colours, red and blue), inside the dashed-line oval shape at the centre of the figure. Other observations can be part of the standard repertoire of observables for our system (say, dogs or trees), but their factors are not explicitly drawn as we only focus on (some aspects of) a system capable of disentangling factors that generated observations of different kinds of balls. A disentangled representation is one that *faithfully* maps factors to codes via the given observations, according to the assumptions provided in the main text, see Definition 4.1. In this scenario we have a list of codes that can be mapped to appropriately (i.e. injectively) from a list of factors through observations. As explained in the text, however, this is an extremely idealised situation meant to show the theoretical, optimal requirements for a definition of disentanglement. In practice, machine learning models instantiated in a causal setting, see for instance the variational autoencoder in section 5.1, can only partially meet these requirements, for example due to limited capacity of an encoder or due to noisy observations mixing factors.

Definition 4.1 *(Disentangled representations).* A disentangled representation is a product of codes Z_i for $i \in \mathbb{N}$ defined through the following:

- a **generative process** or data generating process, as a function

$$g : S \to X \tag{3}$$

that gives rise to observations X from the product of n factors $S = S_1 \times S_2 \times \cdots \times S_n$; this function is assumed to be injective since we ask that if two observations are identical, they must be mapped to by two identical factors, but also allow for the possibility that some observations may not be mapped to from any factor (i.e. g is not surjective), which can be seen as stating that the set of observations can be of much higher cardinality compared to the set of factors; intuitively, the set of observations could include all the possible scenes a subject has ever been exposed to, but the factors S producing data, e.g. about a ball bouncing around, only map to a small part of all possible observations (all possible variations of the red bouncing ball, while for example saying nothing about a tree whose leaves are moved by the wind, see Fig. 3),
- an **encoding**, as a function

$$f : X \to Z \tag{4}$$

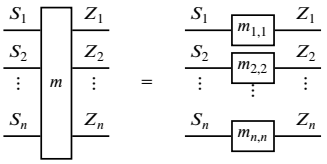
that maps observations X to the product of l codes $Z_1 \times Z_2 \times \cdots \times Z_l$ while also requiring f to be injective on g's image, $g(S)$; for a more intuitive interpretation we consider the following scenarios:
  - $l \geq n$, the case considered here (in [123], see section 5.1, and in [121] who focused on the special case $l = n$), meaning that we assume there are potentially more codes than factors and thus all factors can be encoded by different codes (independently, after we impose more structure below),
  - $l < n$, relevant for practical implementations (see for instance variational autoencoders in section 5.1), where we may not have (or want, for reasons including for instance lossy compression) enough codes to model all factors due to constraints or wrong

modelling choices, in this case we need to consider approximations to disentanglement, either mathematically [124] or via some (deep) reinforcement learning implementation, which will be discussed in section 5.1,

- a **modularity** map,

$$m : S \to Z \tag{5}$$

defined as the composition $m = f \circ g$,[2] i.e. using n factors S to generate observations X and then encoding these observations, $g(S)$, in $n(\leq l)$ codes Z, the constraint can also be visually expressed in the following (string diagrammatic) form to be read left to right



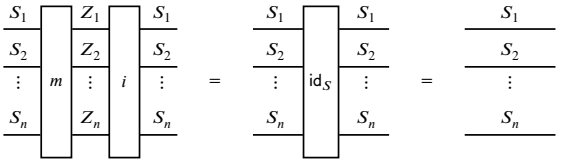
(6)

stating that in the case of $l \geq n$,[3] the map m factors into $m = m_{1,1} \times m_{2,2} \times \cdots \times m_{n,n}$, meaning that the n-th code only encodes the n-th factor, note that m is injective as the composition of an injective function, g, and a function injective under its image $g(S)$, f,

- an **informativeness** requirement, modelled as the existence of as a function

$$i : Z \to S \tag{7}$$

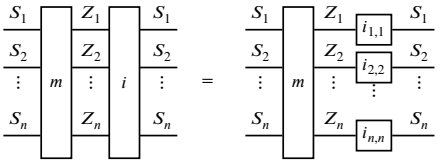
that maps codes to factors in such a way that $i \circ m = \mathsf{id}_S$[4] for the identity map on factors id_S, see the following[5]



(8)

i.e. i is a left inverse for m, or in other words (post)composing m with its left inverse i means that factors can be recovered (we have identities on the right hand side of eq. (8)), it doesn't however tell us that they will be disentangled (i is not itself modular/factorised),

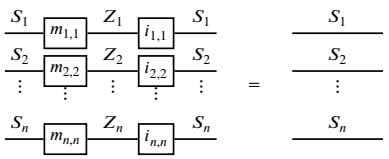
- a **disentanglement** requisite on informativeness, this ensures that the codes are *independent* faithful representations of different factors and can be stated using the following diagram



(9)

where $i = i_{1,1} \times i_{2,2} \times \cdots \times i_{n,n}$; this condition is particularly relevant because there may be cases where, for instance, without a proper causal understanding of the factors generating a bouncing ball, one might mistakenly assume material and colour, together, are a relevant factor: if all the balls an agent ever saw bouncing were red and made of rubber, colour and material could be assumed to be *jointly* necessary factors to understand how the ball bounces.[6]

Combining these conditions, we obtain finally the following diagram for **disentangled representations** $Z_1 \times \cdots \times Z_n$ satisfying the following equation



(10)

[2] Think of $f \circ g$ as $f(g(\mathbf{s}))$ for some element $\mathbf{s} \in S$.

[3] If $l < n$, we can simply replace n with l.

[4] Technically, we say that m is a split monomorphism [122].

[5] Notice that since we treat modularity and informativeness as separate criteria, we don't include the modular/factorised version of m in eq. (8).

[6] Further constraints, of general importance for a more precise definition of disentangled representations that can deal with undesirable special cases (e.g. redundancy of information in the codes) can be found in [121] but will not be further discussed here.

In this view, a disentangled representation is thus one that captures or reflects in a meaningful way the expressivity of data-generating factors underlying some given observations. While useful as a general guideline for a more precise notion of disentanglement, it is however still unclear at this stage how to relate this notion of disentanglement to different dimensions of causal understanding [62] in a more formal way. So far, we haven't in fact discussed how this idea can be connected to formal accounts of causality. To form the basis of this connection we thus introduce next some key concepts from causal representation learning that allow us to go beyond the somewhat idealised illustration of disentanglement just outlined, involving an exact notion of disentanglement fulfilling the requirements for full identifiability of factors via the existence of an injective mapping to codes.

Due to the probabilistic setup underlying the following definition(s), one could argue that the notion of disentanglement we reviewed here (using sets and functions) can't be applied directly to the models in the following sections. However, one of the strengths of the categorical approach proposed by [121] is the presence of relatively straightforward generalisations of the notion of disentanglement to categories other than **Set**, including (Markov) categories handling probabilistic reasoning in a "function-like" manner. This means that a definition such as the one given above can be generalised to probabilistic models of interest in machine learning and causal reasoning.

It is however evident at this point that a definition like the one for disentanglement provided here can hardly be considered practical, since most of its conditions appear hard to meet by any realistic model. For this reason, we will from here onwards use "disentanglement" in a slightly looser way, taking it to capture different flavours of disentangled representations with, crucially, the same structural desiderata, allowing at the same time for "disentanglement" to include different levels of approximations for practical implementations. For instance, we will talk about disentanglement for probabilistic setups, where exact disentanglement still appears quite difficult, but different levels of success have been reached with approximations given by the variational autoencoders reviewed in section 5.1. For extensions to the above definition that treat approximations to the identifiability of factors more formally, e.g. without the injective mapping assumption, we refer the reader to [124], where approximations to the modularity assumption are studied in the formal settings of enriched category theory, expressing measures of "failure" for a map to be injective and distances from the idealised (injective) case, and [125], for extensions of this approach to Markov categories involving probability theory.

4.2. Causal representation learning

As highlighted by Zhang and Sugiyama [121], disentanglement has been described in various different ways across the literature. In one of the most influential accounts, disentanglement can be viewed as a component of causal models recovering the causal factorisation of a process generating a collection of observations of interest [33,126]. In this view, disentanglement is thus a crucial part of the answer to the question of how causal models are acquired, providing a way to operationalise a process deemed necessary for an agent to *learn* a causal characterisation of the world it acts in [33,126].[7] As we will see, taking this perspective allows us to relate explicitness, one of the dimensions of causal cognition (see section 3.1), to disentanglement, in light of the nascent field of causal machine learning.

Causal machine learning is a collection of methods and applications based on the notion that exploiting causal information in data can lead to a more robust, accurate, and efficient kind of (data or system) modelling, thereby viewing causality as a fundamental notion to move past some of the limitations of machine learning methods based on statistical learning (from now on we will refer to these methods as "traditional machine learning") [32,128,129].

Within this line of research, the subfield of causal representation learning can be regarded as a way to recover disentangled representations from data. Traditionally, representation learning has been conceived as the task of learning a **generative model** in the form of a low-dimensional feature space (codes) of high-dimensional data (observations) produced by a **generative process** whose features (factors) remain hidden. The idea driving this approach is that if those codes capture key, informative, aspects of a dataset, they would aid in solving downstream tasks (i.e. predicting a label) [119]. However, these models have often sidestepped questions regarding the origins of particular datasets, overlooking structural knowledge of the data-generating process that could have produced them. This in turn affects what a generative model can account for, often limiting its scope to only statistical correlations with little to no causal power.

Causal representation learning extends these ideas by bringing into representation learning (and, more generally, deep learning) some of the principles, methodologies, and objectives of classic causal inference research [112,34,130], with the goal to learn a low-dimensional vector of *causal* codes from high-dimensional observations generated by *causal* factors[8] [33,128,131]. In this paradigm, the data-generating process can be formalised as a **structural causal model** capturing the causal relationships between factors underlying the data distribution. Importantly, recalling the distinction between generative process and generative model, learning a generative model means to represent something about the generative process described as a structural causal model. In the best case scenario, a generative model recovering the full gamut of causal information assumed to exist in the generative process can itself be described as a structural causal model of the same form as the generative process. This particular scenario assumes however *causal sufficiency*, i.e. that there are no hidden common causes (also referred to as hidden **confounders**) on factors in the generative process, meaning that every common cause of any two or more variables is already accounted for and included in the model (see [132, Ch.

[7] An insightful, independent view of some of the possible reasons for disentanglement to be a necessary condition to learn causal models can be found in [127], where sets expressible as Cartesian products of other sets ("factored sets") are used as the starting point of a theory for causal and temporal inference consistent with the semigraphoid axioms (graphoid minus intersection) of conditional independence expressed in [112].

[8] Note that in the causal representation learning literature, both causal codes and causal factors are usually addressed as simply "causal variables". Here we will instead maintain an explicit distinction.

10] and [34, Ch. 9]). More often, we will instead only look at cases where hidden confounders are present, thereby violating causal sufficiency, which will then allow us to distinguish between weak and strong disentanglement in section 5.1.[9] To better understand the role confounders can play, we define next a structural causal model.

Definition 4.2 *(Structural causal model of the generative process).* Given the following:

- a collection $S = (S_1, \dots, S_n)$ of $n \in \mathbb{N}$ causal factors (or causal variables),
- a collection $X = (X_1, \dots, X_d)$ of $d\,(\geq n) \in \mathbb{N}$ observables,
- a collection $C = (C_1, \dots, C_d)$ of $m \in \mathbb{N}$ confounders,
- a collection $N^S = (N_1^S, \dots, N_n^S)$ of n noise variables on causal factors,
- a collection $N^X = (N_1^S, \dots, N_d^S)$ of d noise variables on observables,

and assuming that all the noise variables are jointly independent, a structural causal model $\mathcal{C}$ [112] of the data-generating process is:

- a collection $(h_1, \dots, h_n)$ of n structural assignments, each assigning a value to a corresponding causal factor S_j for $j \in \{1, \dots, n\}$ based on
  - its parents (direct causes) in the set

$$\mathbf{PA}_j \subset \{S_{\setminus j}, C\} \quad (\text{where } S_{\setminus j} := S \setminus S_j) \tag{11}$$

  and,
  - the noise variable N_j^S,
- and an emission map (or mixing function) g generating observables X, cf. the generating process in Definition 4.1,

that is:

$$\begin{aligned} S_j &= h_j(\mathbf{PA}_j, N_j^S), \qquad j \in \{1, \dots, n\} \\ X &= g(S, N^X). \end{aligned} \tag{12}$$

Importantly, one can show that a structural causal model $\mathcal{C}$ entails a corresponding directed acyclic graph, $\mathcal{G}$, and a unique probability distribution[10] $P_{S,X}$ defined over factors S and observables X [34]. Concisely, a structural causal model induces a causal Bayesian network, where the conditional distributions relating causes to effects, e.g. $P(X|S), P(S_i|S_j)$, etc., are often referred to as **causal mechanisms**, capturing ways in which causes produce their effects. The underlying structural assignments of the structural causal model describe those mechanisms in greater detail, providing a functional specification that allows one to consider both linear and non-linear forms of causation with additive or non-additive noise [133–136]. Conversely, we can also think that any empirical probability distribution has an associated structural causal model that induces it. In this case, however, it can be shown that such a structural causal model is not unique. It is nonetheless possible to define an equivalence class of structural causal models consistent with that same distribution [34]. Starting from a joint probability distribution $P_{S,X}$ rather than a specific model $\mathcal{C}$, there is a corresponding equivalence relation, or partition, of directed acyclic graphs with respect to which the $P_{S,X}$ can be factorised. A particular graph can be chosen, then, to reflect the particular causal mechanisms of the true structural causal models.

While crucial for unpacking a notion of disentangled representations with connections to the literature on causal representation learning, the tools we introduced so far remain fundamentally rooted in a body of work in machine learning that does not take into account how agents make use of these tools for decision making over time. To introduce a notion of agent as a system with goals capable of solving a particular class of problems we thus look at the framework of reinforcement learning [26] and review some of its basic components that will be later involved in our account of disentanglement for learning agents. Following Zhang and Sugiyama [121], one can reasonably expect that the definition of disentanglement given in section 4.1 can also be generalised to classes of models with dynamics (i.e. time-dependent models),[11] however at this stage the possibly non-trivial interplay between the categorical (in the sense of being based on category theory) disentanglement of Zhang and Sugiyama [121] and a comparable categorical reinforcement learning setup [137] remains admittedly unclear.[12]

[9] Note that a causal factor or code does not have to be hidden to be a confounder. Any common cause of two or more variables has a confounding effect on certain causal relationships (see the definition of confounding in [34, 113]).

[10] For the use of probabilistic language from now onwards, the following conventions are adopted:

1. P stands for a probability distribution or function, and P_X is the probability distribution for a random vector X (if one-dimensional, X should be understood as a random variable), the subscript for P_X will be omitted if the random variable is clear from context,
2. $P(X = x)$ is shortened by simply writing $P(x)$, i.e. using only the value the random variable takes,
3. $p(x)$ is either the probability mass function or probability density function evaluated at x for the probability distribution P_X.

[11] For example, by using monoids instead of groups in [121].

[12] For instance, we can consider bisimulations of POMDPs, described in section 6.1.2 as a possible implementation to achieve a high degree of disentanglement, and whether the definition of equivalence classes of states it embodies includes rewards or not. If it doesn't, one can obtain equivalence classes of states for all actions

4.3. (Causal) reinforcement learning

Reinforcement learning (RL) provides a natural avenue for ways to combine work from machine learning and decision making in adaptive, learning agents [26]. Similarities between classical RL and causality have been put forward in previous works [138–140], however a clear-cut notion of causality appears to be missing [128]. Here we provide some background for standard RL implementations, which will be then placed in context and used once we overview recent work in causal RL in section 6.

A typical reinforcement learning setup involves the definition of a problem in terms of (a model of) an environment represented by a (discrete-time) **Markov decision process (MDP)**.

Definition 4.3 *(Markov decision process (MDP)).* A Markov decision process is a tuple (S, A, T, γ, r), where:

- S is the state space,[13]
- A is the action space,
- $T : S \times A \to P(S)$ is the transitions dynamics, where $P(S)$ is the set of distributions over S with finite support such that for a given state $\mathbf{s}_t$ and $\mathbf{a}_t$, $T(\mathbf{s}_t, \mathbf{a}_t)$ gives a probability distribution of states an agent can transition to from state $\mathbf{s}_t$ while taking action $\mathbf{a}_t$, often written as $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$,
- $\gamma \in [0, 1)$ is called the discount factor,
- $r : S \times A \to \mathbb{R}$ is the reward function, giving a reward every time a transition is taken.

Alternatively, it is also common to define a problem as a **partially observable Markov decision process (POMDP)**, where information of the environment is only indirectly available through some observations.

Definition 4.4 *(Partially observable Markov decision process (POMDP)).* A partially observable Markov decision process is a tuple $(S, A, X, T, M, \gamma, r)$, where S, A, T, γ, r follow the definition of an MDP and

- X is the observation space,
- $M : S \to P(X)$ is the observation or measurement map.

The goal of agents in an RL setup is to select sequences of actions that maximise expected cumulative discounted reward, also known as expected return, based on past and current experiences acquired through meaningful interactions with the environment. Action policies representing sequences of actions are defined by the following

Definition 4.5 *(Action policy).* Given an MDP (S, A, T, γ, r), a policy π is defined as either

- a deterministic function $\pi : S \to A$, or
- a stochastic map $\pi : S \to P(A)$ assigning a distribution of actions to each state in S.

For a POMDP $(S, A, X, T, M, \gamma, r)$, a policy π is usually defined instead as either

- a deterministic function $\pi : B \to A$, or
- a stochastic map $\pi : B \to P(A)$,

where *beliefs* $B : H \to P(S)$ play the role of sufficient statistics of histories, i.e. observation-action sequences $H := (X, A)^*$ where * is the Kleene star, used to represent zero or more combinations of its argument, in this case observation-action pairs. In the case of MDPs, these beliefs are trivialised by full observability and the Markov property, so that states S observed at some time t are sufficient statistics of histories of state-action sequences in $H := (S, A)^*$ up to t. As we will see, in practice policies are often parameterised by the weights $\boldsymbol{\omega}$ of a neural network that outputs the action probabilities after processing $\mathbf{s}$, and in this case, they will be defined using the following notation: $\pi_{\boldsymbol{\omega}}$.

For probabilistic state transitions and action policies, the expected cumulative discounted reward is defined as follows.

Definition 4.6 *(Expected cumulative discounted reward).* Let $S_t \subseteq S, A_t \subseteq A$ be state and action subspaces of their respective spaces, indexed by time $t \in T$. We define time-indexed rewards $R_t \subseteq \mathbb{R}$ for $t \in T$ as

with transition dynamics leading to the same equivalence classes of states, i.e. a task-irrelevant partition of states consistent with their dynamics that does not depend on the reward signal perceived by a particular agent. If it includes rewards on the other hand, one in general obtains a finer partition that considers states *having the same reward* from which transitions for all actions lead to the same equivalence classes of states, i.e. a partition of states sensitive to instantaneous rewards and thus to task-relevant aspects of a problem for a particular agent.

[13] Our choice of using S to represent states of the environment conforms with the distinction between factors S of a generative process and codes Z representing states of an agent's generative model.

$$R_t := r(S_t, A_t), \tag{13}$$

which, in turn, can be used to define the cumulative discounted reward, or return G_t, indicated by

$$G_t := \sum_{i=t}^{T} \gamma^{i-t} R_t \tag{14}$$

with $i = 1$ for rewards over a full trajectory, and $i > 1$ for rewards over a partial one, respectively (the former appears in the next equation, while the latter will be used in eq. (21)).

The *expected* cumulative discounted reward is then defined as

$$J(\boldsymbol{\omega}) = \mathbb{E}_{\tau \sim p_{\pi_{\boldsymbol{\omega}}}(\tau)} \left[G_1 \right] \tag{15}$$

with respect to a probability distribution over trajectories, $p_{\pi_{\boldsymbol{\omega}}}(\tau)$.

After introducing all the necessary, mathematical background for disentanglement and causality as they are treated in this work, in the next section we will look at how these notions have been described, implemented and studied in deep (reinforcement) learning. This step will in particular help to bridge the gap between the formal, but somewhat simplified definition presented in this section (with agents modelled with, or without, disentangled representations and in a causal or simply a-causal way) and measures on degrees of disentanglement that can tell us *how much* disentanglement and causal understanding can be found in different systems, i.e. how far subjects are from the ideal disentangled and causal representations introduced in this section.

5. A computational framework for causal cognition

The computational framework for the study of causal cognition outlined in the following sections, using tools from modern deep (reinforcement) learning, heavily relies on our proposal to identify the explicitness dimension of causal information in natural agents with disentanglement, as studied in the field of (causal) representation learning for artificial systems. Based on this, we conjecture that a certain level of explicitness/disentanglement ought to be necessary for a system to recognise (the presence of) different sources and for the integration of causal information of different kinds. Thus, in our view, the dimensions representing sources and integration should be seen as inherently dependent on the availability of disentangled representations in the first place.

In light of this, our proposal will seek to mainly unpack the explicitness dimension, building on a qualitative and quantitative distinction of degrees of disentangled representations, defined below as “weak” and “strong” disentanglement respectively. Notably, this doesn’t imply that disentanglement is sufficient for a complete account of causal cognition, as additional computational processes need to be considered on top of causal representation learning for a treatment of sources (section 5.2) and integration (section 5.3). However, in our view, the fact that an agent can distinguish between sources of causal information, and decide what components of its experience to combine to succeed in certain tasks, is a natural consequence of establishing a way to discern distinct causal factors. We will thus be providing some concrete proposals for computational processes involved in discovering sources and ways to use and integrate different types of information, but not cover the full spectrum of possible ways to implement these dimensions, especially since for both sources and integration we will see in section 7 that there are various areas where current deep reinforcement learning frameworks could take advantage of ideas from animal cognition.

5.1. Explicitness as degrees of disentanglement

One of the main practical instantiations of disentanglement originates with models built on the architecture of **variational autoencoders (VAEs)** [141,142], where a disentangled representation is defined as one where single latent units (codes) of a VAE are independently responsive to single factors generating observations [143].

Following the notation in section 4.1, the goal of a VAE is to learn, given a dataset $\mathcal{D}$ of observations X, a probabilistic generative model that can approximate factors S using hidden variables (codes) Z. To do so, in a standard VAE architecture we find two neural networks, aptly named encoder and decoder, see Fig. 4: an **encoder** with weights $\boldsymbol{\phi}$ parameterising a distribution $Q^{\boldsymbol{\phi}}_{Z|X}$, and a **decoder** network with weights $\boldsymbol{\theta}$ parameterising a distribution $P^{\boldsymbol{\theta}}_{X|Z}$. The encoder plays the role of the function f in Definition 4.1, mapping observations to codes, while the decoder is a clever construction that corresponds to a map

$$k : Z \to X' \tag{16}$$

where X' can be seen as **reconstructions** of X, i.e. observations that should be “as close as possible” to the original ones according to some measure, in this case given by the VAE optimisation objective provided below. Notice that, since we only required f to be injective for $g(S)$, k need not be a function and thus is not well-defined for the simple setup of sets and functions we adopted in section 4.1. It is however a map that can easily be defined for more general setups [121].[14] Combining these, a VAE is trained to learn weights $(\boldsymbol{\phi}, \boldsymbol{\theta})$ so as to maximise the *evidence-lower bound* (ELBO):

[14] One could also just impose further restrictions on f, for example making it surjective so that k becomes its right inverse. Alternatively, one could see k simply as a *partial multi-valued* function and work in the category of sets and relations, or partial multi-valued functions, however this may not be the best choice for an

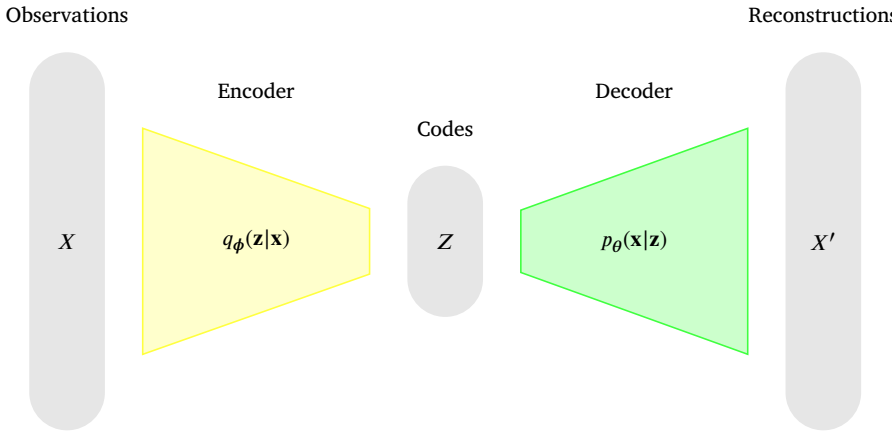


Fig. 4. Variational autoencoder. An intuitive representation of a variational autoencoder, combining an encoder $q_{\phi}(\mathbf{z}|\mathbf{x})$ taking observations $\mathbf{x} \in X$ as inputs and producing $\mathbf{z} \in Z$ as outputs, these outputs are then used by a decoder $p_{\theta}(\mathbf{x}|\mathbf{z})$ providing reconstructions of observations $\mathbf{x} \in X'$ with the goal of making them "as close as possible" to the original observations.

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}) = \mathbb{E}_{\mathbf{z} \sim q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})} \Big[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}) \Big] - D_{KL}[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p(\mathbf{z})] \tag{17}$$

via stochastic gradient descent using (batches of) observations sampled from a dataset [144,142]:

$$(\boldsymbol{\phi}, \boldsymbol{\theta}) \leftarrow (\boldsymbol{\phi}, \boldsymbol{\theta}) + \frac{1}{|D|} \sum_{\mathbf{x} \in D} \nabla \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}). \tag{18}$$

Looking at the ELBO more closely, one can notice how the VAE is tasked with competing objectives. Trying to maximise the first term, usually called the reconstruction loss, amounts to tweaking the weights of both the encoder and the decoder, $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ respectively, in such a way that the latent variables (codes Z) inferred by the former are more likely given a certain input (observations X) and that the decoder can then use those variables to reconstruct it (reconstructed observations X'). In other words, weight values that lead to bad inferences and/or that do not afford a good reconstruction of the original inputs will be penalised. The second term is negative (because the KL divergence, measuring closeness between distributions, is always greater than or equal to zero) so its maximisation tries to bring the divergence to zero, i.e. bringing the posterior distribution close to the prior. This is asking the VAE to map the inputs to latent variables that are as close as possible to the given prior distribution. Overall, the VAE should learn hidden variables, or codes, $\mathbf{z}$ that lead to good reconstructions *and* whose probability can be brought close to the one indicated by the prior.

To improve the disentanglement performance of VAEs, one usually adds a hyperparameter $\beta > 1$ to the KL divergence term of the original objective, leading to what is usually addressed as β-VAE [143]. Despite the empirical confirmation of disentanglement, however, the rationale for the β tweak is not entirely clear. The intuition is that by putting more emphasis on learning statistically independent variables in the latent representation, one could also get a disentangled representation [145,143]. Other proposals to improve on disentanglement, based on the VAE, have seen the introduction of different architectures or regularisation terms to the optimisation objective [146–150]. However, fundamental limitations remain because (1) without some supervision or crucial inductive bias disentangled representation learning cannot be achieved in practice [151–155] and (2) there is no general agreement on how to quantitatively measure disentanglement in the units of the latent representation [156,157,126].

More formal insights can nonetheless be obtained by looking at the optimisation objective of the β-VAE using the **information bottleneck principle**, a method to find a generalised version of sufficient statistics (in this case, codes Z) for a statistical model (in this case, the generative process $g : S \to M$) by defining a trade-off between compression and descriptiveness of a model given some data [158,159], with connections to fundamental laws of thermodynamics [160,161]. More formally, Alemi et al. [162] show that one can derive a version of the β-VAE objective from an unsupervised variation of the information bottleneck principle. In this version, instead of maximising a mutual information term involving a latent representation of codes Z and labels in a dataset D, we minimise a mutual information term capturing how much information the latent representation encodes about a single given data point from the dataset, because the generative model is supposed to learn a class of codes compatible with all the data points, not just one. Using this principle, one can also understand the KL-divergence term $D_{KL}[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p(\mathbf{z})]$ modulated by β [143] as imposing an information bottleneck limiting the amount of information that can be encoded from the data into the latent representation or, in information theoretic terms, as limiting the channel capacity of the encoder [145]. Disentanglement is thus a form of information compression that strives for a semantic alignment of the latent codes with the causal factors underlying the data-generating process (i.e. making latent codes a good approximation of a sufficient statistic for causal factors), providing a minimal (causal) description of a dataset of observations that is maximally informative (see [163] for a connection between data compression and causal modelling, making explicit the sense in which a causal model offers a minimal description of a probability distribution).

introductory treatment of formal notions of disentanglement as the presence of monoidal (and not Cartesian) products complicates the definition of disentanglement [121].

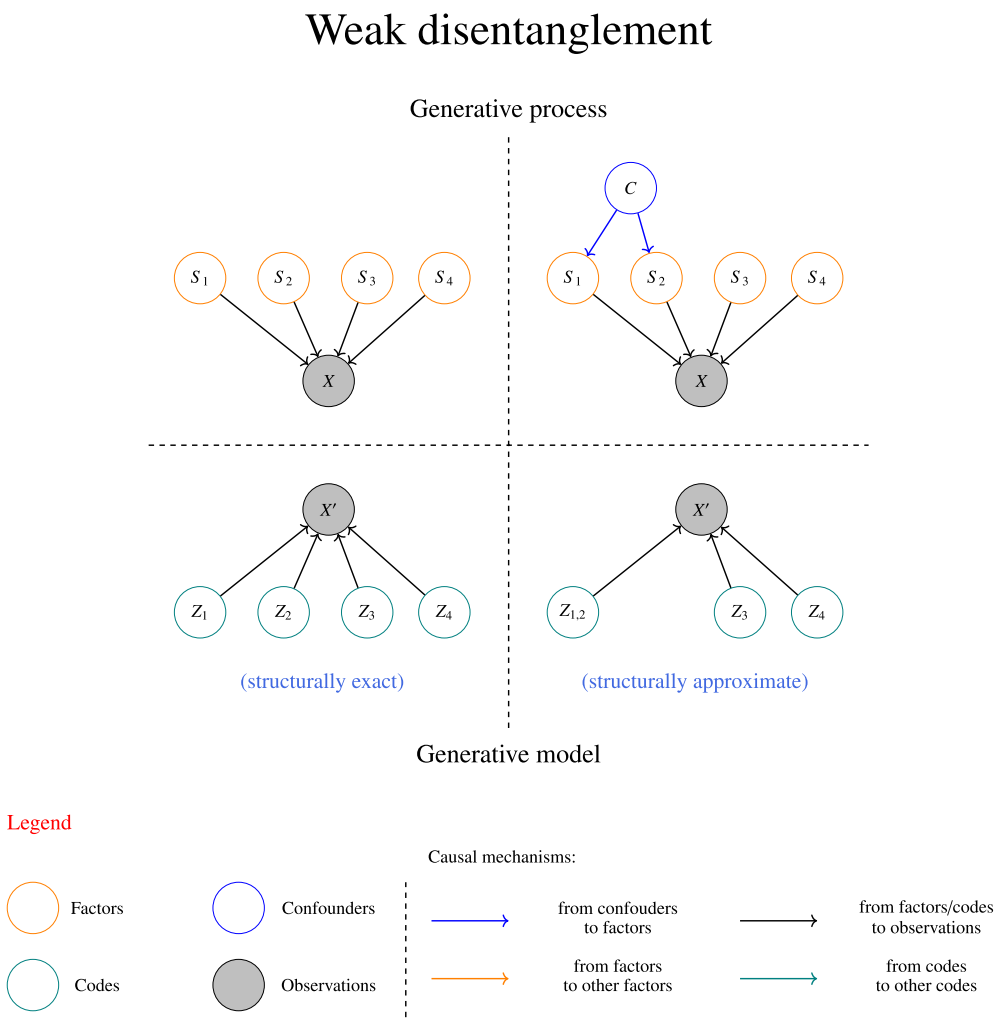


Fig. 5. Weak disentanglement. Weak disentanglement as a variant of causal representation learning concerned only with learning causal codes Z from high-dimensional inputs (observations X) generated by causal factors S, and the identification of causal mechanisms that map causal factors to observations (no causal mechanisms between causal codes). A generative model can be considered structurally approximate (with respect to the assumed generative process) if it fails to recover completely disentangled codes (some codes remain entangled, e.g. the code $Z_{1,2}$ suggestively standing in for two factors S_1, S_2) and/or if it leaves out causal mechanisms relating codes to observations. Note that in an actual implementation, like the VAE, entanglement might manifest as correlations among two or more codes in the latent representation, all capturing the same factors at the same time. In other words, the node $Z_{1,2}$ may in practice represent a group of nodes with possibly bidirectional influences amongst each other.

As explicitly argued in section 4.2, treatments of causal representation learning can be regarded as attempts at learning a *causal generative model* that identifies causal factorisations of a *causal generative process*. Depending on the goals and features of a specific implementation however, we will see that different architectures recover different kinds of latent representations. To highlight what we believe to be the biggest difference, we will define two macro categories, of weak and strong disentanglement, based on whether a model can recover only factors and their relations to observables or factors and causal mechanisms involving observables and other factors of a generative process, respectively. Note that in line with the definition of structural causal model provided above (Definition 4.2) both kinds of relations can, and in practice nearly always do, involve non-linearities and non-additive noise.

Definition 5.1 *(Weak disentanglement).* Weak disentanglement captures the idea of learning a mapping between observations X and causal codes Z, without requiring that the causal relationships among the factors (potentially involving confounders C) are also recovered (see Definition 4.2). This approach appears for example in [123], where confounders in the generative process cannot be encoded by the generative model, and there are no causal mechanisms between causal codes or between confounders and causal codes, thus rendering all assignments h_j trivial (i.e. identities) (cf. eq. (11)), see Fig. 5:

$$\mathbf{PA}_j \subset \{\}, \qquad j \in \{1, \dots, n\}. \tag{19}$$

In weak disentanglement approaches, causal factors are modelled by causal codes as elementary ingredients that independently influence the observations through the mapping g.[15] The goal of an agent is then to learn a causal generative model that reflects this scenario, e.g. by relying on a measure of robustness with respect to interventions [123] or by enforcing an independence (orthogonality) constraint on the Jacobian of the appropriate version of the g map, whose elements quantify the influence of each causal factor on observations from the dataset $\mathcal{D}$ [136]. Further, if observations provide information only about a subset of the causal factors (partial observability), thereby introducing potential confounders, a sparsity constraint imposed on the latent representation can help to recover the ground-truth causal factors [167]. Despite the absence of any reference to causality, practical examples of disentangled representation learning, illustrated by means of the β-VAE earlier, achieve something similar insofar as one prepares a dataset with known causal factors and usually shows that their method can recover all those factors in the latent representation (in this case the correspondence is expected to be one-to-one, in contrast with [123] where it is assumed to be one-to-many).

Definition 5.2 *(Strong disentanglement).* Strong disentanglement aims to recover not only the causal factors in the latent representation, but also confounders and the causal relationships present among them all (i.e. the causal mechanisms). In other words, in strong disentanglement we consider the more general scenario described in section 4.2 where the causal relations derived from the sets of parents $\mathbf{PA}_j$ of each variable are not assumed to be only confounders

$$\mathbf{PA}_j \subset \{S_{\setminus j}, C\} \quad (= eq.\ (11)). \tag{20}$$

This is implemented for instance by the CausalVAE architecture [168], where a modified VAE is augmented with a mask layer, in the form of an adjacency matrix, trained and applied to the latent vector of codes Z to implement the structural assignments defining the causal mechanisms among the codes (effectively, this step amounts to sampling from a structural causal model), see Fig. 6. The vanilla CausalVAE architecture is however limited to static data, e.g. images, and does not thus lend itself to treatments of causality for dynamical processes where causal factors can be causally related in time, typical for example of reinforcement learning treatments of decision making in agents. To tackle this, building on the CausalVAE and other similar works, extensions have been proposed to handle temporal data, see Fig. 6. In one of these [169], the VAE-based architecture is more elaborate, and designed 1) to encode and decode features of objects (the causal factors) from observations using convolutional neural networks, CNNs, 2) to integrate temporal information in the latent representation with gated recurrent units, GRUs, and 3) to infer the causal codes while enforcing constraints on their latent causal dynamics (e.g. that the noise terms are mutually independent or follow a particular distribution), to ensure that the causal generative process is identifiable from data (see section 4.3).

At a high level, differences between the two approaches can be described in terms of causal understanding of properties vs objects:

- models with weak disentanglement can essentially account for *properties*, in the red bouncing ball example these would be colour, shape and material, and these properties are by definition independent as we wouldn't want or expect them to posses causal relationships among them (if they did, they would not be different properties), while
- models of strong disentanglement aim to describe *objects*, in the red bouncing ball example these could be the wind, the ground and a player, and unlike for properties, for objects we would also want a model to capture causal relations among them, in a way that makes them noticeable and relevant for an interacting agent.

Overall, in the spectrum of models implementing disentanglement, from weak to strong, a causal representation that only identifies causal factors and their relations to observations can be said to be less explicit than one also recovering the causal mechanisms that exist among them. The latter is in turn less explicit than a causal representation that identifies the functional form of the causal mechanisms, etc. In other words, a more fine-grained and detailed causal representation encodes more causal information, thereby describing more explicitly the causal structure of a certain domain. This is in line with the informal depiction of explicitness offered by section 3.1, where the notion was related to the idea of having more causal information to operate in a certain context, e.g. knowing what means can be manipulated to reach a certain goal, and will be used hereafter as the foundation of our proposal to relate work on causality in animal cognition and deep (reinforcement) learning (see section 6.1).

5.2. Trajectories as sources of causal information in RL

In **online learning** an agent uses the current policy to perform an action in the (training) environment, which responds with a reward signal, at each time step. Trajectories of state-action or observation-action pairs, defined as $\tau := [\mathbf{s}_0, \mathbf{a}_0, \dots, \mathbf{s}_T, \mathbf{a}_T]$ and $\tau := [\mathbf{x}_0, \mathbf{a}_0, \dots, \mathbf{x}_T, \mathbf{a}_T]$ respectively, often called histories (see section 4.3), can be stored in a **replay buffer**[16] as sequences of tuples together with their respective rewards at each time step, e.g., $(\mathbf{s}_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1})$ or $(\mathbf{x}_t, \mathbf{a}_t, r_t, \mathbf{x}_{t+1})$, where one sequence corresponds to

[15] If confounders are not present (or not considered), see left panel in Fig. 5, then the problem of weak disentanglement resembles that of independent component analysis (ICA) [164], for connections between ICA and causal representation learning, see [165,136,166].

[16] This procedure is also known as *experience replay* and was part and parcel of one of the first breakthroughs of deep reinforcement learning, see [6].

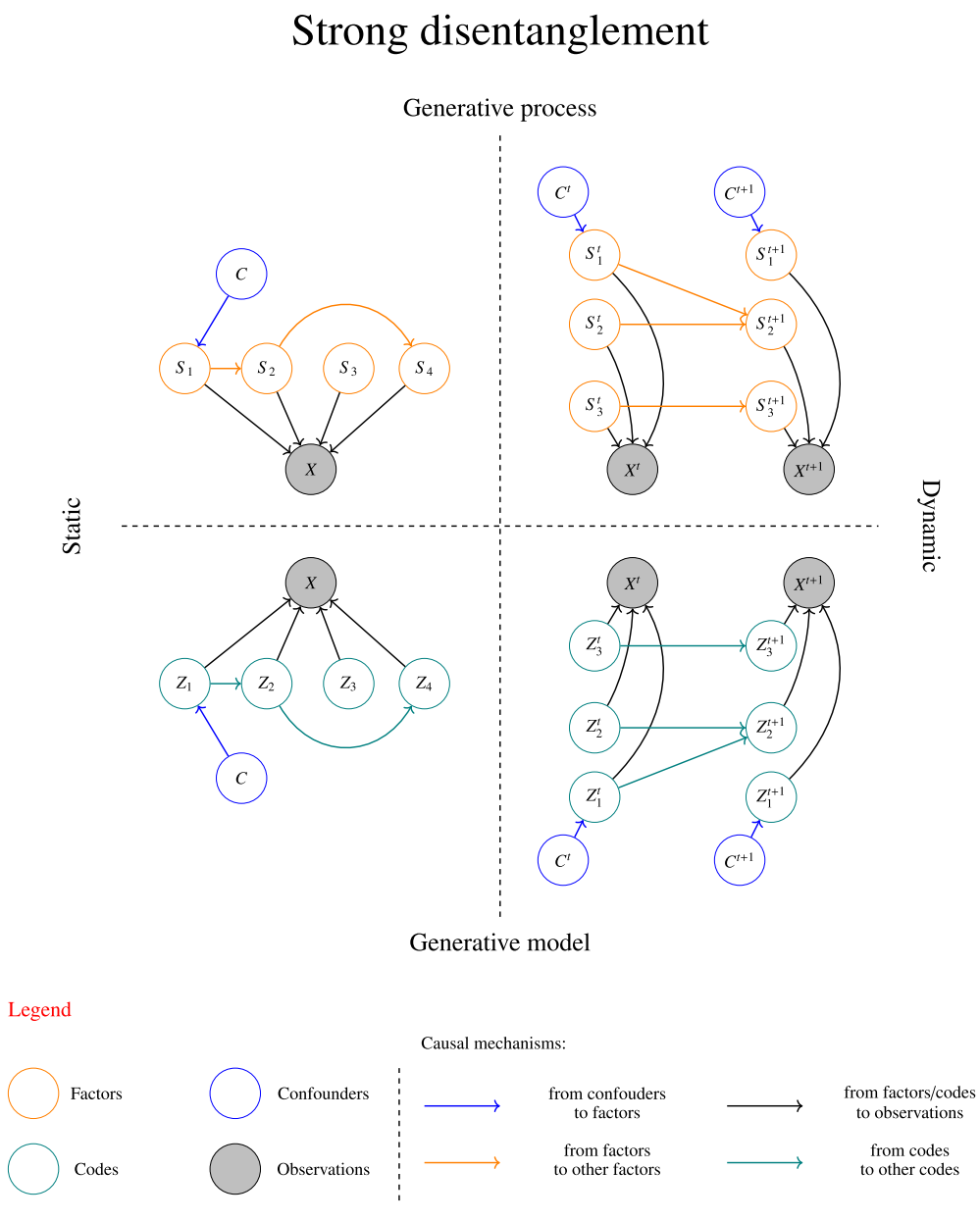


Fig. 6. Strong disentanglement. Strong disentanglement as a variant of causal representation learning concerned with learning causal codes and causal mechanisms of different kinds, to observations and to other causal codes. In a static setting (left panel), strong disentanglement involves discovering the causal mechanism among causal factors and, potentially, the structural assignments of the underlying structural causal model. In a dynamic setting (right panel), strong disentanglement denotes finding causal structure in the transition dynamics (e.g. of an MDP), determining causal mechanisms between causal factors at different time steps. Importantly, we don't include causal mechanisms within the same time slice as we assume that there can be no instantaneous interactions among factors. Note that even when we include dynamics there is still an underlying SCM that can be partitioned into sets of causal factors "over time", i.e. causal factors are processes extended over time, or sets of variables (in this case conveniently given the same name but a different time index) belonging to different time steps. Intuitively, in the dynamics setting more causal structure is uncovered because one is considering causal mechanisms over more causal factors (assigned to different time slices).

a trajectory. This information forms an agent's experience: the state the agent was in, $\mathbf{s}_t$, the action it performed, $\mathbf{a}_t$, the reward it collected, r_t, and the next state $\mathbf{s}_{t+1}$ reached from $\mathbf{s}_t$ by performing $\mathbf{a}_t$ [170,6,31].

Concretely, this experience can be used to obtain a Monte Carlo estimate of the expected cumulative discounted reward in eq. (15) based on trajectories sampled from the replay buffer, which is regularly updated and acts as a rudimentary storage of memories for the agent. To see this idea in action, we look at a popular class of approaches represented by **actor-critic** methods for which the approximate gradient of the RL objective (see eq. (15)), used to update policy parameters $\boldsymbol{\omega}$, is computed as follows [26]:

$$\nabla_{\boldsymbol{\omega}}\hat{J}(\boldsymbol{\omega}) = \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\nabla_{\boldsymbol{\omega}}\log\pi_{\boldsymbol{\omega}}(\mathbf{a}_{n,t}|\mathbf{s}_{n,t})\left(\sum_{i=t}^{T}\gamma^{i-t}r(\mathbf{s}_{n,i},\mathbf{a}_{n,i}) - V^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i})\right) \tag{21}$$

where N is the number of trajectories sampled from the replay buffer; $\pi_{\boldsymbol{\omega}}$ is the current policy whose parameters $\boldsymbol{\omega}$ will be updated with the computed gradient; $\sum_{i=t}^{T}\gamma^{i-t}r(\mathbf{s}_{n,i},\mathbf{a}_{n,i})$ is the discounted sum of rewards (G_t, cf. eq. (14)) evaluated for the (partial) sampled trajectory n and acquired from time step t until the terminal state T; and $V^{\pi_{\boldsymbol{\omega}}}$ is a (state) value function that acts as a baseline with respect to the discounted sum of rewards, defined as the expected return G_t from a chosen state sampled from trajectory n at time $t = i$, $\mathbf{s}_{n,i}$, if a policy $\pi_{\boldsymbol{\omega}}$ is followed from that point onwards, i.e.

$$V^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i}) := \mathbb{E}_{\pi_{\boldsymbol{\omega}}}[G_t|S_t = \mathbf{s}_{n,i}]. \tag{22}$$

In eq. (21), it is common [26,171–173] to approximate the discounted return $\sum_{i=t}^{T}\gamma^{i-t}r(\mathbf{s}_{n,i},\mathbf{a}_{n,i})$, the sum of rewards obtained from a particular, realised trajectory n, with

$$Q^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i},\mathbf{a}_{n,i}) := \mathbb{E}_{\pi_{\boldsymbol{\omega}}}[G_t|S_t = \mathbf{s}_{n,i}, A_t = \mathbf{a}_{n,i}]. \tag{23}$$

This is the Q-function (or action value function) for the policy under consideration, quantifying the value of performing an action in a certain state, after which the policy is followed until the end of the episode. With this substitution, one defines the advantage $\mathrm{Adv}^{\pi_{\boldsymbol{\omega}}}$, specifying how much better it is to take an action $\mathbf{a}_{n,i}$ as opposed to an average action[17]

$$\mathrm{Adv}^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i},\mathbf{a}_{n,i}) := Q^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i},\mathbf{a}_{n,i}) - V^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i}) \tag{25}$$

that can be approximated using a *critic* neural network trained to estimate only the state value function from the reward (because the Q-function can be rewritten as the sum of the reward at the current state and the expected state value function at the next, i.e. $Q^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i},\mathbf{a}_{n,i}) = r(\mathbf{s}_{n,i},\mathbf{a}_{n,i}) + \mathbb{E}_{\mathbf{s}_{n,i+1}\sim P(\mathbf{s}_{n,i+1}|\mathbf{s}_{n,i},\mathbf{a}_{n,i})}\left[V^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i+1})\right]$) [26,171–173]. The *actor* part is represented instead by the policy $\pi_{\boldsymbol{\omega}}$, parameterised by a policy neural network that outputs the most suitable action given a certain state. A learning step then involves sampling a batch of trajectories from the buffer, using them to evaluate $\nabla_{\boldsymbol{\omega}}\hat{J}(\boldsymbol{\omega})$, i.e. the gradient of the estimated objective with respect to the policy parameters, and updating these parameters, $\boldsymbol{\omega}$, using the gradient to derive a better policy, conducive to expected cumulative reward maximisation.

Importantly, the learning problem becomes **off-policy** if the sampled trajectories used to compute the gradients are not collected by the current policy (as implemented by the policy network at the current time step) but by a different one (implemented, for instance, by an old parameters configuration of the policy network). In practical situations, this is often the case because the replay buffer does not store only the most recent history, acquired by the current policy, but also past experience. Therefore, sampling a batch of trajectories from the replay buffer turns the learning problem into an off-policy one. Similarly, in an imitation learning scenario, policy optimisation is also by default off-policy because the gradients are computed using sampled trajectories that come from an expert demonstrator and not from the policy currently followed by the agent (see section 6.2.2 for a more detailed overview on imitation learning).

Effectively, an off-policy learning problem highlights the extent to which a structure analogous to the replay buffer could act as a "storage" of experiences derived from different sources, e.g. from other agents (including the agent's past self). In turn, the challenge arises as to how to weigh and/or combine this experiential data, i.e. the challenge of integration (addressed more extensively in the next section).

In fact, in the off-policy setting, estimating the gradient of the objective considered above, $\hat{J}(\boldsymbol{\omega})$, is problematic because the parameters of the current policy could be updated based on actions and/or reward information (e.g. value functions) that in reality characterise a different/earlier actor (technically, this would be like trying to obtain a Monte Carlo estimate of an expectation with respect to a certain probability distribution, the current policy, with data collected when *a different* distribution/policy was in place). In other words, the gradient information in $\hat{J}(\boldsymbol{\omega})$ might be inaccurate for updating the current policy. Corrections can be applied to the gradient depending on the exact RL method considered, going from a variety of importance sampling techniques (for policy gradient methods, see section 5.3) to the use of appropriate value functions, i.e. using the actions of the current actor (in actor critic methods).

5.3. Combining different sources of causal information in RL

Following the idea of a buffer containing stored trajectories representing experience to update a value function or policy, integration can be interpreted as the ability of an agent to combine different kinds of experience into its own decision-making process, appropriately weighted based on context, task demands, origin, resources, etc. In principle, these kinds of experience can include single-source causal information, e.g. when an agent uses experience acquired at different points in time or in different environments/tasks, such as in multi-task or meta-RL (see [174]). However, in this context we focus on integration of different sources (see

[17] This can be seen by writing down explicitly the relationship between V and Q, in our case

$$V^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i}) = \mathbb{E}_{\mathbf{a}\sim\pi_{\boldsymbol{\omega}}}\left[Q^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i},\mathbf{a}_{n,i})\right] \tag{24}$$

section 3.2), with the goal of highlighting synergistic forms of causal understanding that truly take advantage of the amalgamation of different causal perspectives. This allows us thus to delineate an operational, computational account of the notional idea of integration presented in [62].

In computational terms, the question of how best to integrate and use information coming from different sources is a foundational aspect of **offline RL**, where the replay buffer can store trajectory data collected by any policy in a variety of virtual environments, more or less realistic, or from real-world tasks. For example, recent datasets for offline RL tend to include trajectories from experts (e.g. hand-designed controllers or human demonstrators), from other RL agents trained online in a certain domain, from the same agent operating in the same environment but performing slightly different tasks (multi-task, past experience), and from unsupervised (i.e. reward-free) exploratory policies [175–178].

The challenge of designing an offline RL algorithm is precisely that of exploiting the collected data in such a way that the learned policy can be safely applied to a given environment. This means that the learning algorithm has to acquire and integrate causal information from various (PO)MDPs (those in the training set) in such a way that the most appropriate actions for new downstream tasks/environments can be extrapolated successfully from past experience and generalised into unfamiliar contexts. A vanilla approach consists of using **importance sampling**, originally tailored for dealing with off-policy data (see previous section) [179–183,26].

In this context, importance sampling corresponds to the introduction of importance weights, ratios (computed over a trajectory) between the current policy to be optimised, $\pi_{\boldsymbol{\omega}}$, and a behavioural policy, $\pi_{\boldsymbol{\beta}}$, used to collect the transitions sampled from the replay buffer:

$$\boldsymbol{w}_T^i = \prod_{t=1}^{T} \frac{\pi_{\boldsymbol{\omega}}(\mathbf{a}_{i,t}|\mathbf{s}_{i,t})}{\pi_{\boldsymbol{\beta}}(\mathbf{a}_{i,t}|\mathbf{s}_{i,t})}. \tag{26}$$

The application of these weights in eq. (21), to compute a Monte Carlo estimate of the RL objective's expectation and the corresponding policy gradient, could be seen as a mechanism to facilitate the integration of different sources of information, obtaining:

$$\nabla_{\boldsymbol{\omega}} \hat{J}(\boldsymbol{\omega}) = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{w}_T^i \sum_{t=1}^{T} \nabla_{\boldsymbol{\omega}} \log \pi_{\boldsymbol{\omega}}(\mathbf{a}_{n,t}|\mathbf{s}_{n,t}) \left(\sum_{i=t}^{T} \gamma^{i-t} r(\mathbf{s}_{n,i}, \mathbf{a}_{n,i}) - V^{\pi_{\boldsymbol{\omega}}}(\mathbf{s}_{n,i}) \right) \tag{27}$$

Weights $\boldsymbol{w}_T^i$ adjust gradient information coming from trajectory collected *off the current policy* (i.e. using the behavioural policy) according to how much the two policies are in agreement. This approach works under the assumptions that the marginal state probabilities with respect to the current and old policy networks' parameters, say $p_{\boldsymbol{\omega}\,\mathrm{cur}}(\mathbf{s})$ and $p_{\boldsymbol{\omega}\,\mathrm{old}}(\mathbf{s})$, are sufficiently close to each other, and that the dynamics of the respective (PO)MDPs are the same. In the off-policy case, these assumptions hold true because during learning the replay buffer is regularly updated with newly experienced trajectories and cleared from the older ones, and because the (PO)MDP dynamics are usually assumed to be fixed. In the offline setting, this is however not guaranteed, see [184] for a recent review of some of these and other open challenges, and the next section for a more complete overview of relevant works in causal RL on integration, explicitness and sources across causal reinforcement learning and animal cognition.

Importantly, this is only one possible strategy that could facilitate the integration of different experiences for the purpose of adaptive behaviour. More generally, how to integrate efficiently causal information coming from different domains/sources into flexible causal models is an active area of research in causal machine learning. For instance, there are amortised approaches to causal discovery that approximate a posterior over graph structures in various settings, relying on differentiable frameworks (e.g. variational inference) for a quantification of uncertainty, and some can leverage and integrate both observational and interventional data during training [185,38,39,186,36,47]. There are also some initial studies focusing on the modularity of these causal models, on how they can be continuously refined and/or composed, to enable successful generalisation [187,57,58]. The extent to which these techniques can be adopted by the problem setting of reinforcement learning, i.e. whether they have a bearing on what it means for an embodied agent to integrate causal information, remains to be seen.

6. Bringing together causality in natural and artificial agents

Recent work in (deep) reinforcement learning, in the area now called causal reinforcement learning [188], can help us shed light on ways to translate algorithms from machine learning into a more systematic study of causal learning agents. Using this line of work, we thus review a series of algorithmic implementations and models from causal RL and place them on a spectrum of increasingly high disentanglement, providing a comparative analysis with empirical and conceptual works in the animal cognition literature, see Fig. 7. This will in turn suggest a more concrete connection with the explicitness dimension of Starzak and Gray [62]'s framework, paving the way for an understanding of causal information from different sources and possible strategies to integrate them sensibly.

6.1. Explicitness of causal representations

6.1.1. Weak disentanglement

At the lower end of the explicitness spectrum (see Fig. 7), we find agents of traditional (non-causal) deep RL setups that are successful at solving a variety of narrow tasks by engaging in forms of **dense learning**, meaning that they often appear to learn at least some of the dependencies between their actions and desired outcomes/rewards [26,6,221,217], and in some cases they are augmented with more sophisticated forms of planning, curiosity-based exploration and the ability to achieve a variety of goals in high-dimensional environments [222,28,29,218,219,30,220]. Nonetheless, the web of dependencies learned by these agents is usually

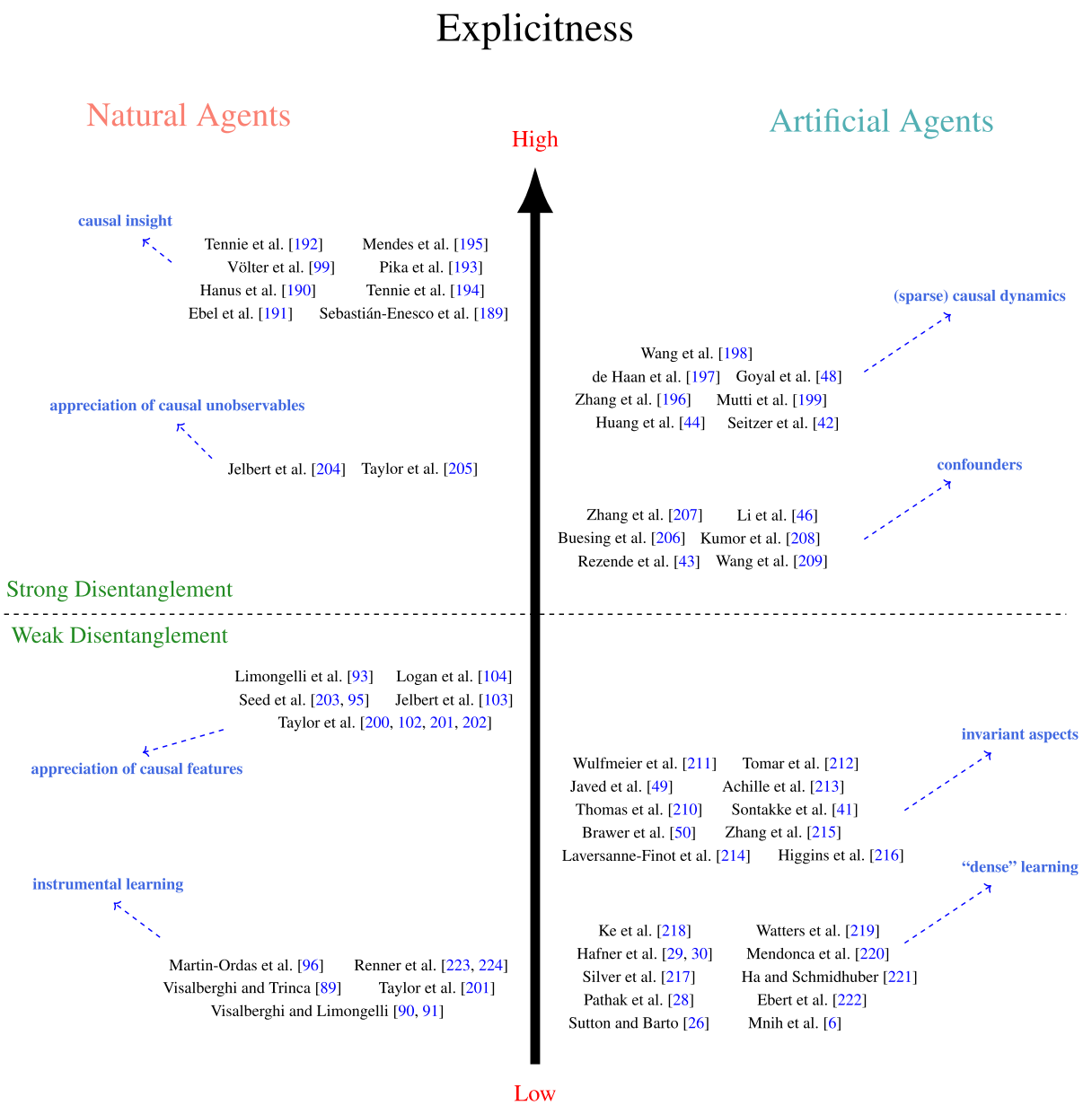


Fig. 7. The explicitness spectrum.

dense because, as dependencies that are associative in nature, they include spurious features and/or relationships. In other words, dense learning in these agents goes hand in hand with a lack of causal information processing. These agents are in several ways akin to animal subjects engaging in **instrumental learning** [89–91,201,224,223,96] (see section 2), except for the amount of data samples used during training.

To have a better understanding of algorithms and empirical results higher in our explicitness scale, and their relation to weak and strong disentanglement, it is then useful to look at more specific features of causal representations. In classical RL, particularly when the problem is presented as a POMDP, a representation can be understood in two different ways:

- in *model-free RL*, these are representations of factors (i.e. codes) given as inputs to a policy, $\pi(\mathbf{a}|\mathbf{s})$, i.e. the state representations (implemented as vectors of state variables) fed to the policy network to produce an action, while
- in *model-based RL*, the term representation points at a model of the transition dynamics (involving, in turn, the state representations) $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$.

Based on this distinction, we suggest that there are two different ways to understand what a causal representation involves in causal reinforcement learning: 1) in causal model-free RL, a causal representation describes a particular encoding, often a compression, of the observations into latent states (with a causal flavour) while 2) in causal model-based RL, a causal representation models both the latent states and the causal dynamics of the environment. Here we suggest that the first kind of causal representation has a lower degree of explicitness than the second one, since it fails to capture the causal dynamics of the environment. We thus view it as a possible example of *weak disentanglement*. The latter instead is defined precisely in terms of its ability to map causal dynamics and is thus an example of *strong disentanglement* (section 5.1).

In causal model-free RL, in particular for a partially observable setting, an agent is said to learn an explicit causal representation if it can map high-dimensional observations to state representations that are disentangled, uncovering codes that represent some parts of the causal structure (the causal factors) of the data-generating process, but not including a complete disentangled representation (see section 5.1). Thus, an agent can be said to exploit this (partial) causal information if such information facilitates policy learning or has a positive impact on policy execution when the causal representation is fed to the policy network.

More specifically, the benefits of this kind of causal representation would follow from capturing **invariant aspects** of an environment, represented by the causal codes, making learning and using a policy more robust to changes in secondary, irrelevant variables (i.e. improving on generalisation).[18] There are many works on invariant representation learning for RL [207,229–234], but it is not entirely clear whether all these invariant representations qualify as disentangled representations in the weak sense discussed here.

Conversely, as long as causal codes can be considered invariant aspects of the environment, disentangled representations in a RL agent are necessarily some kind of invariant representations. Empirically, the extent to which this is key to credit assignment, i.e. the ability of an agent to determine which actions (and/or states) contributed the most to successful performance in a certain domain, remains however to be proved. At present, there is some evidence indicating that learning a causal representation, one achieving weak disentanglement, provides several benefits in RL, e.g. a better exploration of the state space and more robust learning [213,214,216,211,210,41,215,50,212,49].

At a comparable level of explicitness, in the animal cognition literature we find subjects that can successfully solve trap-tube tasks, water-displacement tasks, and/or tasks with similar setups where an attention to objects' physical and functional properties is essential. Recognising these objects' invariants and their causal role for the purpose of solving a given task is suggestive of an **appreciation of causal features**, constituting factors of the generative process, that disambiguates them, at least in part, from non-causal ones [93,200,102–104,95,203,202,201]. For instance, there is empirical evidence that crows can drop stones into baited, water-filled tubes according to stones' width and water levels. Lower water levels and wide tubes hinder in fact water displacement with the available stones, which is necessary to reach the reward [104], see section 2.3.

6.1.2. Strong disentanglement

In causal model-based reinforcement learning the agent is specifically trained to learn a *world model*, a model of the dynamics of the environment, then used for planning and decision-making (i.e. selecting the next action). In this context, causality-inspired approaches involve revealing and exploiting more causal aspects of the (modelled) environmental dynamics, regarded as crucial for having more capable learning agents that, for instance, do not fall prey to spurious correlations like agents with less explicit models might.

To achieve this, we have attempts to handle **confounders**, hidden common causes that can have an impact on factors and their state transitions (see section 4.2), by deconfounding the dynamics of a POMDP. Practically, this entails modelling state-transitions as affected by confounders, whose presence is either assumed from the start [46,206–208], or can emerge from initially unaccounted parts of the dynamics/predictive model through a process of decomposition of observations into confounding and relevant state information [43,209]. This leads to more explicit models because the effects of confounding factors are isolated to obtain a more robust understanding of how events in the environment unfold.

More in detail, we can look at [46] as an example of the first kind (known confounders), based on object-centric learning (using graph-neural networks) combined with a model of the transition dynamics that is assumed to be confounded by time-invariant hidden variables, e.g. the object's masses, friction coefficients, etc. The goal of the agent here is to solve a POMDP, but this requires learning a generative model that is deconfounded, estimating the confounding variables for each object (using tools from do-calculus [112,9]), which in turn can be used to generate accurate observation trajectories had the initial conditions been different (e.g. the objects' position). Effectively, this causal world model enables a kind of counterfactual planning that starts with the question of what would have happened under alternative initial conditions, i.e. given an intervention that changes the starting states. On the other hand, for the second group (unknown confounders) we can consider [43] where agents with partial models, i.e. models learnt using past actions and the initial agent's state as opposed to the full trajectory of past observations, are shown to be less robust to policy changes. These partial models are in fact confounded by past observations, which are not used to train the model but do anyway influence the policy picked by the agent, but can be adjusted for such confounders by using once again techniques from do-calculus.

In the animal cognition literature, understanding the influence of potential confounders can be linked to an **appreciation of causal unobservables**, such as in crows adjusting their actions depending on changes in experimental variables that are not visible to them [205,204]. In one study, crows were tested on a task that consisted in extracting some food from a box, placed on a table

[18] The connection between invariant aspects of an environment and causality has been established, for instance in [112,34,132]. However, there is ongoing disagreement on how traditional machine learning methods should be adjusted to capture invariance in data (with some theoretical guarantees) [225–228]. Similarly, it remains unclear how the RL framework could be impacted by such theoretical advancements in the long term. See also works cited in the main text.

and in front of a curtain. From behind the curtain, a human could operate a wooden stick that through a hole in the curtain could come close to the food box, therefore causing trouble for the crows trying to reach the food. The presence/absence of the human thus confounds whether it is "safe" to go and retrieve the food from the box (because in principle a stick's movement does not create danger, unless it is intentionally used to poke through a hole, for example by a human experiment), therefore it would be useful to be able to reason about what is behind the curtain. The evidence reported in [205] suggests that crows can attribute the movement of the stick to a hidden agent behind the curtain and act accordingly, e.g. being more cautious when they do not observe anyone leaving the experiment's room (because the stick could move again). Similarly, in a context where a food dispenser is activated by means of placing objects on it and where an object's weight confounds the food release (only heavy enough objects activate the dispenser), crows can learn to infer the weights of the objects from their movements in a breeze and pick the appropriate ones to get the food from the dispenser [204]. In both studies, the animal subjects were able to adjust their behaviour by paying attention to the reward dynamics, i.e. to whether narrow or wide tube were to be preferred (according to the respective water level), or to whether light or heavy objects were activating the dispenser.

Beyond confounding factors, model-based reinforcement learning can be improved by observing that key causal relationships in the environment, relevant for solving a particular problem, do not involve all state variables and transitions among them. That is, the causal dependencies among variables in the environment that an agent can have an effect on, for the purpose of reaching a certain goal state, form a causal structure that is *sparse*, in the sense that it only captures some deeper facts about a whole class of problems (or environments) for a particular agent. For example, an agent might be capable of accessing a certain area of a building by means of a detailed world model that accurately predicts what happens when a red button located next to a glass door is pressed, while standing on a floor with hexagonal tiles. This detailed model keeps track of all possible dependencies among the colours of buttons, the material of adjacent doors, and the geometric shape of the floor tiles on which the agent stands (and it might predict with some confidence that only when a combination of those features is encountered, then access to a certain area will be granted). The problem is that most of those dependencies are likely just spurious correlations, hiding the most fundamental (causal) fact that a button next to a door in general tends to open that door when pressed.

Thus, instead of learning indiscriminately every conditional dependence relation (causal or not) among state variables at adjacent time steps, an agent should strive to learn a causal transition model that identifies the causal relationships that matters for the class of tasks at hand. Precisely, focusing on the **(sparse) causal dynamics** present in a given environment means to identify the subset of latent state variables, or causal factors S, that are likely to form generalisable causal relationships, exploitable not only for the task at hand but for similar tasks as well [44,198,196,42,48,197,199]. Therefore, in contrast to modelling dense dynamics, leveraging causal sparsity is more computationally efficient and can help an agent to avoid learning spurious correlations over time.

One influential approach hinges upon notions of state abstraction [212] (see also [235] for more background). A state abstraction can be regarded as a compact (latent) representation that is invariant to task-irrelevant information (i.e. only information relevant to a specific problem is encoded), and is technically defined as a (probabilistic) bisimulation. A bisimulation (equivalence) is a structure-preserving equivalence relation of states of a (PO)MDP, $B \subseteq S \times S$, describing equivalence classes of states S/B, for all actions in the action space A, with transition dynamics leading from states with the same reward R to the same equivalence classes of states, i.e. for $\mathbf{s}_1, \mathbf{s}_2 \in B$ and $\forall i, a \in S/B, A$ the following conditions apply [215,196,232,212]

$$P(i \mid \mathbf{s}_1, a) = P(i \mid \mathbf{s}_2, a)$$

$$r(\mathbf{s}_1, a) = r(\mathbf{s}_2, a) \tag{28}$$

see also the notion of "causal states" for ϵ (epsilon) machines and transducers in computational mechanics [236], roughly a (unifilar, minimal) version of probabilistic bisimulations for stationary stochastic processes, without rewards.

This line of work can lead to a higher level of explicitness, via strong(er) disentanglement, as exemplified for instance by [198], showing empirically that their architecture can learn codes identifying causal factors as well as the causal mechanisms between them, across times steps. This means that the RL agent can determine whether a state at time step t has a causal influence on another state at $t+1$, and whether the causal relation is relevant for solving a particular problem.[19]

In [62], a high-level of explicitness is connected with adaptive behaviour supported by a flexible use of causal information, which enables an animal to re-use acquired knowledge or past behaviours (with the appropriate changes, if necessary) to reach new goals, the same goals but in a slightly different context, and/or solve tasks never encountered before. Therefore, at the top level of the explicitness spectrum, we find **causal insight**, a term we use to refer broadly to this type of generalisation abilities, chiefly involving a deeper realisation of what a problem/scenario entails, based on causal knowledge. These can be often seen in transfer learning and innovative/insightful problem solving, e.g. in the floating-peanut task [190,194,191,189,195,193,99,192], in which the solution to a task is allegedly reached via an adaptive restructuring of one's experience [237,238]. The main idea here is that non-human animals appear to be capable of cognitive feats thanks to highly explicit causal models of both state variables and transition dynamics, though there is some disagreement in the experimental literature on the extent to which this effectively happens [202,239,240].

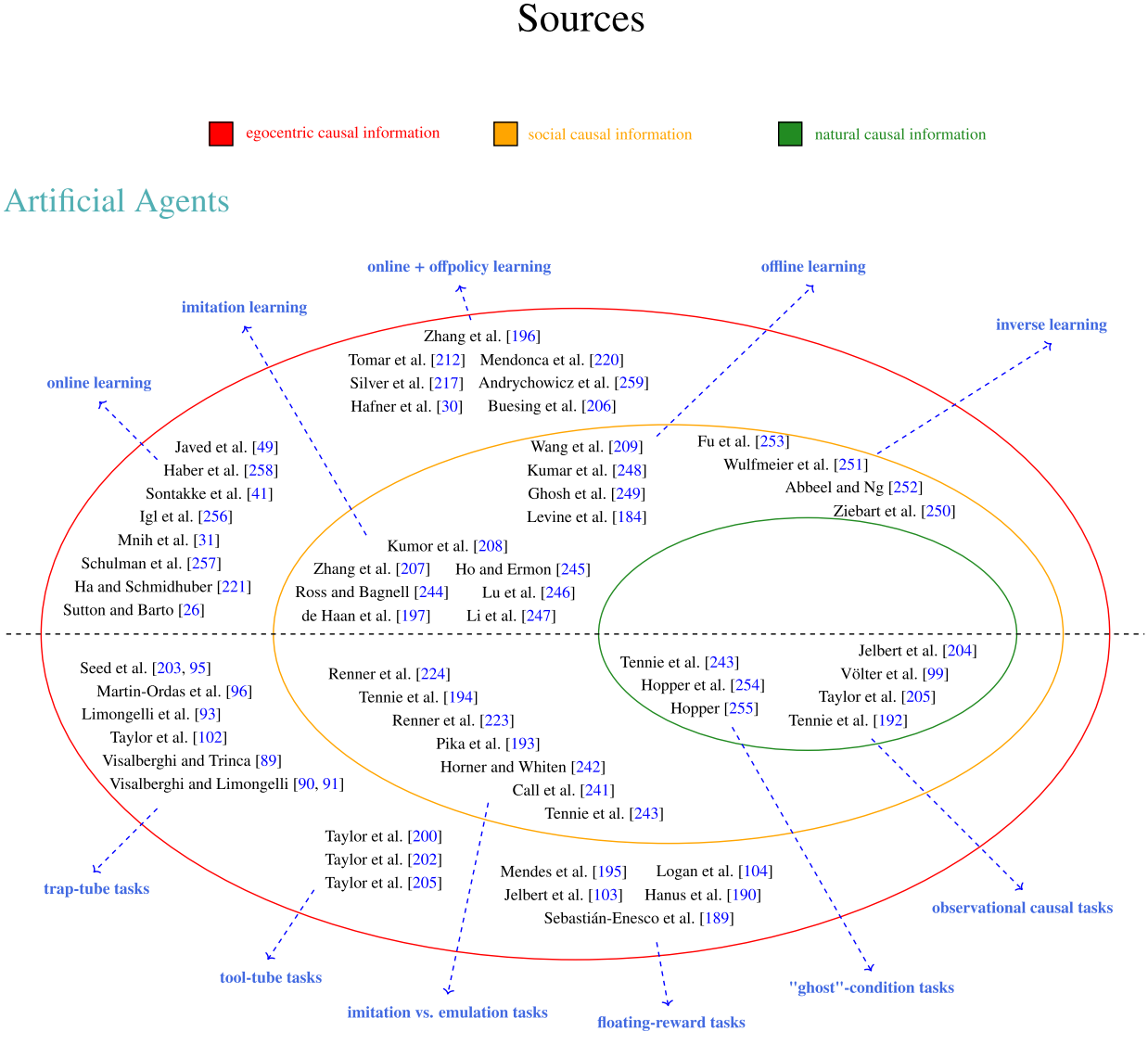


Fig. 8. Sources of causal information.

6.2. *Sources of causal information*

6.2.1. *Egocentric causal information*

Successfully learning from online interactions in an environment implies appreciating, to some extent, the relevance of certain action-outcome contingencies for reward maximisation or reaching a certain goal. Learning from online interaction amounts to instrumental learning, which chiefly involves egocentric causal information (see Fig. 8) and has been extensively studied in animals (see section 3.2), particularly with a variety of tasks including **trap-tube tasks** [89–91,93,102,95,203,96] (see section 2.2), various **tool-use tasks** [200,202,205] and **floating-reward tasks** [103,104,195,190,189] (among several others).

Similarly, most RL agents are designed to be egocentric causal learners, with varying abilities to latch onto causal information provided by their own experience, which is ultimately shaped by the provided reward signal. These abilities come for instance from strategies to boost **online learning** via particular training/optimisation techniques (e.g. uncertainty-based or curiosity-driven

[19] To be precise, in [198] the state abstraction is learned on the states of a MDP, so the challenge of deriving a disentangled latent representation (from high-dimensional observations) is bypassed. In other words, weak disentanglement is taken for granted.

exploration) [26,221,257,31,256,258,49,41], or from methods to maximise the benefits of **online + off-policy learning** [206,217,212,196,220,259,30].

With the exception of some of the works cited in section 3.1, most of these approaches have not adopted a causal terminology. Furthermore, the agents in question do not process feedback from the environment as causal information, e.g. by paying attention to key causal relationships with techniques from causal machine learning. Nonetheless, we refer to them as egocentric causal learners because they have the ability to process (at least partially) the consequences of their actions.

6.2.2. Social causal information

As already mentioned in section 5.2, artificial agents can be designed to learn to solve a task through **imitation learning**, e.g. by relying on demonstrations of the expected behaviour for the given task. The imitation learning pipeline can be implemented in various ways, tailored to the specific domain of application (for a recent review of the main techniques, see [260]). In RL, the general idea is to allow a learning agent to have access to the experience of an expert, i.e. trajectories of optimal interactions for solving the task at hand, which are conveniently pre-processed in the same representational format of information in the replay buffer, so that they can guide the learning process towards a policy that achieves similar rewards [244,245,247]. While broadly successful, imitation learning approaches do not necessarily entail the processing of social causal information in a comparable way to natural agents. Imitation learning in and by itself does not in fact prevent a learning agent from simply exploiting correlations between state variables and actions present in the dataset of expert's demonstrations to learn an optimal policy for a certain task. In the presence of distributional shift, which arises every time trajectory information used for training comes from a policy different from the one currently used by the learning agent, agents that learn by imitation, but without causal knowledge, are usually prone to causal confusion or misidentification (e.g. of what prompted the expert to act in certain ways) [197]. If a correlation ceases to exist, performing the same action in response to a certain state could in fact turn out to be inappropriate in most cases. This knowledge deficit has been highlighted and studied in depth by a few recent causal RL works, making a first important step towards artificial agents trained via imitation that are better equipped to deal with confounders and spurious correlations [197,207,208,246], making them more "aware" of the causal structure of the problem under consideration.

The emergence of **offline RL** has marked another milestone in approaches to learning from imitation insofar as the emphasis is placed on the ability to learn from a dataset of previously recorded trajectories, potentially coming from other agents performing similar or different tasks [184,209]. This represents a more challenging problem because during training the agent can no longer receive feedback from the environment, using its current policy to collect more trajectories through trial-and-error, as is typically done in imitation learning. Optimal behaviour must be learned from a dataset that is not updated during training, and that inevitably will not provide a complete picture of the environment/task in which the agent will be deployed. Techniques to ensure that a policy will perform well enough when deployed include conservative methods to bound the learned value functions (to avoid the risk of assigning high values to wrong states) [248], algorithms that take into consideration the agent's uncertainty about the identity of the test environment (enabling a kind of policy adaptation at test time) [249], and causal approaches to off-policy policy evaluation (see [184] and [261] for comprehensive reviews).

To gain a better understanding of the extent to which current imitation learning approaches in RL are linked to causal cognition, it is instructive to consider a line of research in the animal cognition literature directed at investigating what kind of learning strategy is adopted in a social context by non-human primates, using **imitation vs. emulation tasks**. The distinction between imitation and emulation revolves around the particular "copying" strategy used by the learning agent when observing the behaviour of a conspecific, i.e. either adhering to the demonstrator's actions (imitation) or focusing more on the action's results or outcomes (emulation) [262,263,118,264,265]. The imitating agent will reproduce virtually the very same actions of the demonstrator whereas the emulating agent will try to reproduce the results of those actions, e.g. a rewarding outcome, using the same or different behavioural strategies, depending on context [241,243,194,223,224,193,242]. For instance, to collect a floating peanut from a water-filled tube (an example of a floating-reward task), one has to increase the water level in the cylindrical container; a higher water level is the key instrumental result (or precondition) required to solve the task. In a social setting with expert demonstrators, a subject that overlooks that piece of information and learns to solve the task by copying all the particular actions of the expert conspecific (e.g. the ambulatory behaviour to collect the water) will fail at the task if those actions are no longer appropriate, or available, to produce the desired outcome (e.g. the water can be accessed only by climbing) [194]. Importantly, there is empirical evidence suggesting that adopting an emulative vs. imitative learning strategy can depend on the availability of causal information about the effects of certain actions, and their connection with the final, desired outcome. For instance, in [242] chimpanzees witnessed a human demonstrator securing a reward from a puzzle-box using a tool. When the box was opaque, hiding the relevant tool movements necessary to unlock the reward, at test time the subject reproduced all the actions seen in the demonstrations (learning by imitation). Conversely, with a clear box the subjects learned to ignore the irrelevant actions, thereby solving the task more efficiently. Thus, learning by emulation implies attending to goal/instrumental information (e.g. higher water levels in the tube) and being able to act upon it whether relevant behaviour has been demonstrated or not, i.e. attending to causal information pertaining to the causal states and/or variables that form a sort of precondition to reach a final outcome. As such, this learning strategy affords efficiency and flexibility because the learning subject is free to explore and select the best course of action to reach an end goal.

An alternative RL framework for imitation learning can be found in **inverse RL**, where the aim is to design an agent capable of inferring the objective of the expert demonstrator, i.e. what reward function is shaping its behaviour, and of looking for an optimal policy based on that [252,250,251,253].

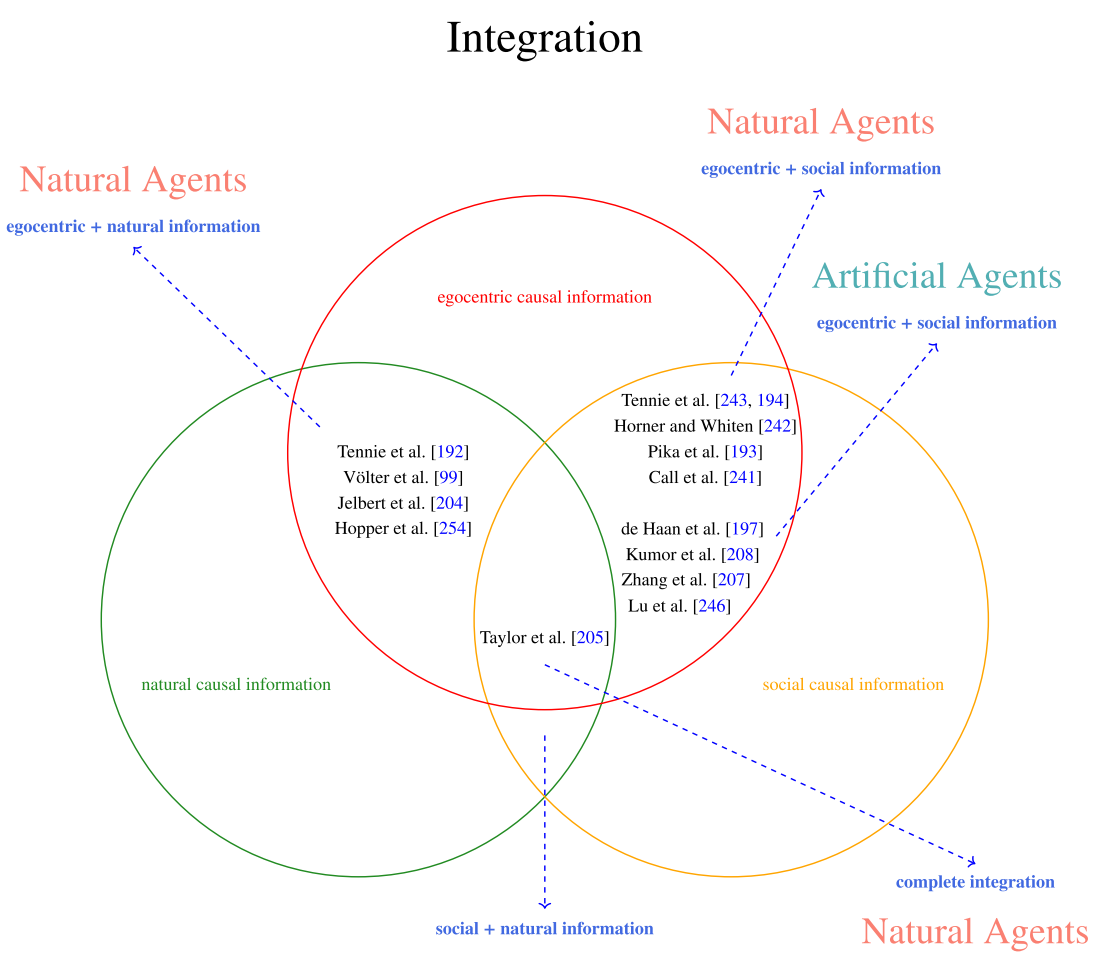


Fig. 9. Integration of causal information.

6.2.3. *Natural causal information*

Beyond the ability of learning from online interactions and social demonstrations, some (natural) agents also display a propensity for the acquisition of causal information from natural sources. Natural causal information is precisely information about the existence (or absence) of certain causal relationships or structures that is gleaned from observing the occurrence of natural events (see section 3.2).

Since the tree-branch thought experiment of Tomasello and Call [118], it seems that the general consensus on observational (or impersonal) causal learning being an exclusively human ability has not shifted [5]. However, there is empirical evidence coming from some **observational causal tasks**, in which key causal relations can only be inferred from observations, suggesting that observational causal cognition in some non-human animals might be more developed than what it has been normally thought. For instance, corvids have demonstrated an ability to take into consideration the potential effects of hidden causes, e.g. other agents or properties like the weight of an object, from observations alone [205,204] while chimpanzees have been shown to be capable of inferring the presence of causal relationships from patterns of covariation (with a blicket-like experiment) [99] and using temporal cues [192].

Similarly, **"ghost"-condition tasks**, showing an apparatus in a final desired state and/or how a mechanism works (by pulling invisible strings), used to study emulation learning, also suggest that non-human primates exploit observational causal information to guide subsequent successful behaviour [254,255,243].

Thus, arguably, despite not reaching the performance achieved by humans, some non-human animals appear to have the ability to learn about the causal structure of not only systems they interact with, but also of systems they can merely observe. This places them in a category beyond imitation (causal) learners, as they can make use of experience other than theirs, processing and capturing in causal terms events generated by external sources with a different body or physical configuration.

6.3. *Integration of causal information in natural and artificial agents*

Following section 5.3, integration can be seen, from our perspective, as the process of incorporating and fusing different domains of causal information since, regardless of its source, any experience can be stored in a replay buffer (see also Fig. 9). It is however important that different kinds of experience are integrated by concurrently taking into account their different roles, relevance for

the given task, and/or potential weights based on the identification of key causal relationships. The studies in animal cognition we examined so far can give us some clues, in the form of particular behavioural profiles, about what type of causal information integration happens in non-human animals. However, it is important to keep in mind that behavioural traits are here used as a rudimentary proxy for cognitive operations of integration that remain still largely unknown.

As we saw in section 6.2, different animal cognition studies involve social causal information, e.g. demonstrations of a desired behaviour by an expert [241–243,194,193]. Despite the existence of negative results (e.g. [224]), these studies provide supporting evidence for the claim that non-human animals are capable of **egocentric + social integration**, see Fig. 9. Within the same group, in RL, recent approaches have started to tease out the impact of certain causal relationships in imitation learning [207,208,197,246]. These works represent a first step towards a better understanding of what integrating egocentric and social causal information might mean and especially entail, e.g. looking for invariances, confounders, direct causes of an expert actions. Yet, it remains to be seen whether these approaches can be successfully extended and/or combined with methods that enable artificial agents to deal with high-dimensional, partially-observable scenarios (POMDPs), as their counterpart, natural causal learners, can integrate causal information starting from observations alone [241–243,194,193].

On the other hand, the fact that an animal's behaviour is influenced by the observation of certain causal relationships can be explained by invoking cognitive operations of integration that combine natural and egocentric causal information. Observational learning experiments suggest that basic forms of **egocentric + natural integration** are present in non-human animals [254,99,204,192]. For instance, inferring that the presence of a causal factor (e.g. the weight of an object) has an impact on what an action can accomplish (e.g. whether one can get food from a dispenser with a certain object or not), and behaving accordingly, can be considered as an example of this type of integration.

Conversely, there is almost no evidence for two other forms of integration in non-human animals, which are likely to require causal reasoning abilities about natural events and other agents that we know are present only in adult humans. Integrating social and natural causal information, **social + natural integration**, might entail scenarios where an agent of choice, say Agent 1, perceives a causal relationship present in the natural world as exploitable by *another* agent, say Agent 2 (for example, an expert in an imitation learning setup), for its own purposes but not for itself. In other words, Agent 2 can combine pieces of causal information involving social and natural facts but at the same time Agent 1 is unable to relate the finding to its own circumstances, preventing a further integration with egocentric causal information. What is hard to even conceive here is the very nature of this form of integration, by which effects on Agent 1 in some contexts are not directly and immediately measurable, because there is no behavioural evidence for this form of integration (this might come later when the agent is eventually able to capitalise on what it has observed and understood).

While in principle easier to detect for its benefits on a subject's egocentric perspective, integration of causal information from all sources, **complete integration**, can only be achieved by the most flexible and adaptive kind of behavioural responses and is thus hard to test in real setups. Work in this direction can be found for instance in [205], where crows can explore an environment and retrieve food from a box (egocentric causal information). The animal subjects learn through observation that the area near the box opening may be unsafe because it faces a curtain with a hole, from which a stick may appear and move, due to potentially natural forces in principle unknown to the subjects, to bother them (natural causal information). Crucially, the crows can also witness that at times a human being enters the testing room and goes behind the curtain, suggesting that the movement of the stick may be caused by another agent (social causal information), posing thus a different kind of threat. The question here is precisely whether the crows can "reason" about the opportunities of exploring the environment to get food based on the perceived risk of facing an aversive stimulus (the moving stick) in relation to the inferred presence/absence of a hidden agent. Since the crows were more hesitant to search the food box when an hidden agent was present, there is thus some evidence that the crows could integrate all those pieces of information (egocentric + social + observational) to tackle the task. It is however debatable whether the crows could effectively recognise the humans agents as the causes responsible for the stick's motion. In fact, it has been pointed out that non-human animals seem to lack a sophisticated understanding of causality in the psychological domain [97]: in this case that the hidden agents could have the intention to move the stick when in the room.

7. Discussion

Our comparative analysis so far has highlighted areas where animal cognition and causal reinforcement learning share some common ground in their otherwise different approaches to the study of causality in cognition and decision making. In this final section we look in more detail at some of the opportunities offered by our unifying formal account of causal cognition, showcasing ways to make a more synergistic use of its strengths, and speculating on areas we believe will be of particular interest for future explorations. At the very end, we will look back at human causal cognition, an area which we have largely overlooked to focus on lower-level causal skills typically found also in non-human animals, which we assumed throughout the work to provide enough of a challenge to face already for current AI models.

7.1. Computational interpretations of studies of natural agents

Modelling approaches in the literature of animal cognition are mainly concerned with capturing the cognitive and psychological processes of subjects exposed to tasks that are assumed to require an understanding of causal information [111,78,266]. However, simply providing evidence that subjects can learn complex causal structures from patterns of conditional (in)dependence shown to them and reason about interventions and counterfactuals in the world, leaves open the more fundamental question we asked in the introduction. Specifically, *how* does an adaptive agent interacting with and receiving feedback from an environment become sensitive

to certain causal information and process it in ways that are conducive to reaching its goals? To address this, we propose to use current models and algorithms developed in the fields of causal RL, and provide next some more specific examples.

7.1.1. Measuring explicitness in natural agents

In this work, we considered the question raised by Starzak and Gray [62] of how to define and measure explicitness in animal cognition studies and proposed to think of it as disentanglement (see section 5.1), roughly the degree of causal factorisation of a representation, to gain access to a relevant class of candidate metrics. While a widely-accepted measure of disentanglement is still missing, different proposals have been put forward, providing thus multiple options that could be considered for the modelling and testing of explicitness in animal cognition [267,121,126].

As a first step, we believe that a setup based on the AnimalAI Olympics framework [56,10,268] could be used to introduce an experimental pipeline involving training artificial agents on the same class of causal tasks used in the animal cognition literature, to compare their performance with that of animals. If the performance of two agents (artificial and natural) were to be comparable according to some appropriate success metric (e.g. solving a task, behavioural similarity), one could then measure the degree of disentanglement of the artificial agent's representation and use that to gain some understanding of a possible computational theory reflecting the natural agent's modelling capabilities. Furthermore, we believe that this approach has the potential to become a standard benchmark for causal AI research to test if, and what kind of, causal representations can unlock the necessary skills to tackle problems of different complexity, supporting causal learning abilities akin to those that appear to be present in natural agents [269,8].

7.1.2. Zero-shot learning for high(er) explicitness

A second example of the type of formalisation work afforded by causal RL implementations revolves around the fact that high explicitness is assigned to those animal agents that are capable of solving a task without much visual or sensorimotor feedback. Despite contrasting empirical evidence [191], there are in fact suggestions that some non-human animals are capable of finding solutions to a certain problem "in their head" [194,189], without the need for extended trial and error learning, which has been regarded as an indication of causal understanding [93].

Similarly, in RL, zero-shot, offline, meta, and continual learning approaches pursue models of learning agents capable of solving certain tasks with extremely limited training data. This is achieved thanks to a flexible and adaptable (causal) representation of tasks' requirements, derived from a process of *learning to learn* that facilitates the reuse and transfer of previously acquired causal knowledge in/to new situations [270–274,174]. These agents are at the forefront of RL research, with problem definitions, formalisations, and benchmarks in constant evolution [275–277] and we believe they are likely to provide another ideal baseline for the development of computational theories of causal reasoning in animals (in this respect, works on learning to learn in human subjects, chiefly the ability to learn abstract causal principles or *causal overhypotheses*, is also relevant [278–283], but see section 7.3). Furthermore, and more specifically, some work on continual or lifelong learning promises to overcome some of the limitations of the machine learning works discussed earlier, e.g. the fact of designing agents with fixed architectures or parametric models, by proposing instead methods that allow for the instantiations of new models or the composition of existing ones as the evolution of the data/task stream might demand [284–286].

7.1.3. Emulation as inverse RL

A large number of approaches in RL that make use of social causal information, i.e. causal information derived from observing the behaviour of other agents (see section 3.2), can be said to implement a form of imitation learning. Computationally, this form of learning can be described as behavioural cloning: trying to copy the policy of an expert agent as opposed to the outcomes of that policy [244]. There is however a growing interest in different approaches, including offline and inverse RL, that we believe have the potential to provide a computational account of emulation as opposed to imitation learning (see section 6.2.2), where the former describes natural agents that learn to reproduce outcomes of expert demonstrations, e.g. as in [194].

In a typical inverse RL setup for instance, the goal is to learn the reward map of another agent, say the expert. This (non-unique) map can in general entail different optimal action policies, and thus by learning the expert's goal itself, an agent is not bound to only mimic the actions of the expert. In turn, this could allow estimating intentions underlying other agents goal-directed behaviours based on a notion of causality in the psychological domain where intentions can be interpreted as causes of behaviour [65,97]. A first step in this direction could include, for example, testing RL agents in realistic settings with imitation vs. emulation tasks inspired by the animal cognition literature, such as floating-reward tasks [103,104,195,190,189], which could help to determine whether some degree of emulation is possible with current causal imitation learning techniques.

7.2. Causal cognition inspired RL

Looking then at Figs. 7, 8 and 9 we can also identify potential areas where current causal RL frameworks can take inspiration from ideas developed in animal cognition. In particular, we refer to areas where works and standard theories of causal understanding in animals have currently no counterpart in RL: causal insight in Fig. 7, learning from natural causal information in Fig. 8 and complete integration in Fig. 9.

7.2.1. Causal insight for causal RL

In section 6.1, causal insight was described as the (1) capacity to produce adaptive responses as a result of reorganising one's causal knowledge, encoded in (2) highly explicit causal representations that lead to (3) an innovative solution to a problem.

The first defining element (1), involving flexible reuse of information and past knowledge in new tasks and/or environments, is a long-standing challenge in AI research [287–291]. The second aspect (2), based on the relevance of causal representations for *strong generalisation*, i.e. the ability to generalise out-of-distribution, encompasses transfer/meta/multi-task learning paradigms and has been noted in several causal machine learning works, with consensus that disentangled, structured, modular, causal representations can provide several benefits [225,33,187,15,52,292,293]. The third feature (3), putting the emphasis on innovative solutions, in the sense that they give a new take on existing problems or can be used for a new and unseen problem, is not fully captured or does not clearly emerge from the generalisation approaches just mentioned. In fact, for instance, systematic studies of out-of-distribution learning, as currently found in the literature, are mostly limited to synthetic datasets and/or toy problems characterised by narrow task distributions, neglecting more realistic and ecological settings [294–296,48].

More specifically, these sorts of investigations have not been carried out by means of evaluation methods and benchmarking that can take advantage of work found in the animal cognition literature. As suggested in some recent works [10,269,297], using training and testing protocols from animal cognition experiments has the potential to improve current architectures towards the goal of reproducing common sense abilities of different non-human animals (e.g. understanding of everyday physical notions like objecthood, containers, obstructions, and the related sets of affordances). Since common sense abilities are deeply intertwined with an understanding of causality, studies of this kind could help to ascertain the extent to which causal RL can truly capture the manifestations of causal cognition in natural agents.

More generally, embracing a learning paradigm in which the central question is how an agent should gather and store key causal information for the purpose of subsequently extrapolating a strategy for tasks never seen before, seems *essential* for causal insight to develop in artificial agents.[20] Further to the requirement of different training procedures and/or more computational power and data, tackling these kinds of questions will help to overcome persisting limitations, e.g. by introducing a more formal notion of causal insight connected with strong generalisation.

7.2.2. Natural causal information in causal RL

Offline reinforcement learning appears to be one of the closest available formalisations of a decision-making problem like the one posed by the tree-branch example [118][21] where a successful learning agent ought to be able to conduct a causal analysis of the natural scene it is part of, and then proceed to shake a fruit-bearing branch, just on the basis of having witnessed a fruit falling due to the wind shaking the tree. One of the crucial aspects of this decision problem is the requirement of acting in an optimal manner immediately, based on the (natural) experience collected, i.e. without the ability to take advantage of further trial and error learning, which is precisely the setting of offline reinforcement learning.

However, while Tomasello and Call [118] highlight the importance of causal concepts to deal with this decision-making challenge, current approaches to deal with offline learning instead pursue strategies that try to mitigate the degree of distributional shift without any reference to causality. In a nutshell, some methods introduce constraints on the policy being learned so as to minimise its divergence from the policy that collected the transitions stored in the replay buffer, while others use uncertainty measures to learn more conservative value functions in order to avoid catastrophic mistakes due to distributional shifts, see [184] for a review.

It is also important to note that, while the state transitions stored in the replay buffer could come from any policy, e.g. even those employed by agents with different bodily configurations or "nature" itself, to the best of our knowledge there are still no techniques to infer trajectory information from visual data in offline RL. More specifically, we refer to the ability to process natural happenings, i.e. physical phenomena that don't involve any particular agent (the "ghost" conditions from the animal cognition literature, see section 6.2.3), through a causal lens. In practice, this would correspond to extracting state transitions tuples, $(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{a}_t, r_t)$, from high-dimensional visual input where a crucial step is to interpret some external (physical) event as an impersonal action, producing an environmental state transition (the tree shaken by the wind). Extending the offline framework in causal RL to include mechanisms to learn from nature has the potential to uncover further aspects of causality not yet understood and might spur a series of novel algorithmic solutions, getting closer to the design of an observational causal agent, where the "ghost" conditions from the animal cognition literature [254,255,243] could be used as a test bed for this new generation of agents.

7.2.3. Interventions in causal RL agents

The ability to reason about and perform interventions, which in a technical sense can be described as local perturbations of a system that set one or more causal factors/mechanisms to certain fixed values [112,32], has often been described as one of the hallmarks of causal reasoning agents [78,308,23,18,9] and has been used to draw a possible distinction between acting and intervening (see also [309–313] for works on active physical learning or experimentation in children and adult humans). In this view, the former only implies an appreciation of the consequences of one's bodily movements (e.g. locomotion, reaching, grasping) leading to changes in perception and conditions for achieving certain goals. The latter additionally involves an intentional modification of a certain aspect of the environment, exploiting an existing or induced causal relationship, to elicit a desired effect [5,98]. For instance, using a stick to make a fruit fall from a tree branch (intervening) is in many ways different from climbing the same tree to grab the fruit (acting), even though the outcome is ultimately the same (eating the fruit). This suggests that not all actions of an agent qualify as

[20] Among the main achievements of deep RL in the last decade, there are for instance successes again top-level players in Go and other online games [298,217,299,300] (see [301] for a review). Some of the moves in these games have been described as creative and insightful (moves that no human player would have made at the time). However, note that in these cases the artificial agents in question are overly specialised in a single domain so we cannot talk of causal insight.

[21] We speculate that transformer-based architectures [302], especially large language models and vision-language models for RL [178,303–307], could provide an equally interesting proposal, whose in-depth discussion is however left for future work.

interventions but all interventions are actions, whether realised or only imagined. Importantly, however, while an agent might be regarded as performing an intervention from the perspective of an external observer, it does not follow that the agent itself conceives of its actions as interventions.

By examining the animal cognition literature, at least two markers for intervention-aware agents can be identified. One is the ability of certain agents to infer that an environmental causal path cannot be activated if an action is directed at producing an effect along that path. If the tone is produced by the rat through a lever press, the cause (the light) that usually elicits the tone and makes the food available is likely absent and so will be the food [98,78]. The second one is the capacity to implement an innovative behavioural response simply following the exposure to certain observational patterns, i.e. what we called causal insight [100,101]. For instance, the primates solving floating-reward tasks seem to showcase a capacity to intervene, i.e. manipulating the environment (water level in the tube) to obtain a desired effect (reward is closer to the surface level) from observations alone, sometimes even without the need for visual feedback [189].

On the other hand, artificial RL agents appear to be exclusively engaged in acting (when the agent is embodied, either in a simulation or in the real world), rather than intervening, and it is still unclear what the most promising approach to develop some form of intervention-awareness even is. On a basic level, it might be crucial to revisit in causal terms some of the existing RL machinery. Recent works [314,315] for instance show how the advantage function (see eq. (25)) can be interpreted as the causal effect of an action on the return (see eq. (14)) that displays typical properties of a causal representation (e.g. disentanglement), and discuss ways to estimate such a function directly from experience. New architectures might also be needed, and among the necessary computational/algorithmic components of intervention-aware agents there might be an intervention model that converts low-level motor actions into interventions on causal factors [15]. Perhaps more drastically, causal counterparts of traditional machine learning (and/or RL) concepts will have to be developed and incorporated into the current theory and practice of RL, e.g. a causal/interventional version of the KL divergence [316,114].

7.3. Connections to causal cognition in humans

As mentioned at the very beginning of this work, the scientific study of causal cognition has a long history, one that is not only based on an extensive literature about non-human animals but that remains also intimately related to humans studies. It is thus useful at this final stage to look at what our current proposal can say about results in human causal cognition.

Early investigations tried to provide an account of human causal induction—roughly, the learning or judgement of certain relationships as causal—by relying on statistical or associationist frameworks [68,71,317,318,17].

With the advent of causal Bayesian networks [132,112], a flurry of research emerged on humans' inference and learning patterns in a variety of tasks involving reasoning about causal relationships, examined against that normative framework [319–323]. In cognitive psychology, this research crystallised into a view that identified on the ability to learn and operate with causal models the distinctive trait of the human mind [324,21,325].

More recently, several works have proposed computational models on how people induce causal structure, e.g. when they are allowed to intervene on static causal networks and observe the consequences [326,309,327], when they can intervene on a continuous-time dynamical system [328–330], or when they can only observe the unfolding of some events [331,332].

Another important line of research involves counterfactual reasoning, which is often recognised as one of the most sophisticated abilities a causal agent could master [9] (and one which was not explicitly considered in the present work due to our focus on lower-level cognition). Humans' ability to reason about counterfactuals has been investigated, for instance, in the context of the formulation of causal judgements about a state of affairs (e.g. whether two billiard balls are about to collide), to establish the extent to which they might depend on counterfactual simulations/evaluations on how some physical events are likely to unfold [333–337].

While this research in cognitive psychology shares some important themes and methods with the works on non-human animals we reviewed here, it is mainly motivated by the goal of how to characterise causal abilities in the realm of higher-level cognition, as exemplified by adult human subjects, which assumes the presence of some sort of causal representation or the capacity to acquire and manipulate one. In contrast, one of the main drives of the present work has been to understand how lower-level causal abilities can develop as part of a process of learning from (more or less) ecological interactions in biological and artificial systems of, we believe, comparable complexity. This implies that our focus was more towards causal cognition as emerging "from pixels" (as stated in section 1), or high-dimensional (observational) experience, as opposed to a starting point in more well defined (usually by a scientist) lower-dimensional model spaces (more common in the human causal cognition literature).

For instance, works on active physical learning in cognitive psychology examine the strategies used by children and adult humans to disclose some of the key (latent) causal properties of a system based on an intuitive physical understanding of the environment [309–313]. However, asking the question of how these kinds of inductive biases or intuitive theories are exploited is different from asking the question of how they could be learned (or revised) by the cognitive agent from experience in the first place (*cf.* [41,338], where the RL agent learns through active experimentation to categorise experience based on different physical features, such as friction, gravity or the shape of objects).

Similarly, works on "learning by program induction" consider the ability to operate on or tinker with symbolic structures that are analogues of software programs as a foundational cognitive strategy used by humans to learn and develop world models [339–342]. This line of research can be understood as a modern formulation of the language of thought hypothesis [343] and relies on a modelling approach whereby the learning agent has built-in domain-specific knowledge or (symbolic) primitives. In contrast, the machine learning perspective elaborated in the present work tends to assume less specific a-priori knowledge and is more concerned with how richer kinds of models can be learned when the agent faces certain tasks.

In other words, the present work puts the emphasis on an investigation into causal cognition that starts with the formal and computational principles underlying the capacities of embodied agents of solving comparative cognition tasks [269] as opposed to those supporting the higher-level abilities of linguistic agents like adult humans [8] (see also the reply of Botvinick et al. [344] to Lake et al. [8] for a similar perspective on what differentiates the two approaches). A natural extension of our current proposal could involve a deep (reinforcement) learning approach to higher-order skills. At the moment, it is however unclear if the three-dimension framework we use as inspiration [62] can effectively be used to account for more general aspects of causal cognition, e.g. language or counterfactual reasoning, and it may need some further additions to better accommodate the structure and manifestations of higher-level forms of causal information processing.

8. Conclusion

In this work, we introduced a unifying theoretical and computational framework for causal cognition, connecting different strands of research, from the classical literature on animal cognition to modern accounts of causal reinforcement learning in AI. While traditionally presented in antithesis to associative learning, causal cognition has more recently been taken to span a wider spectrum of cognitive abilities all the way from associative learning to complex tasks such as tool use, emulation/imitation and observational learning. A key aspect of this integrative view stems from recognising different levels of causal understanding as a key component of a framework for causal cognition.

At the same time, the lack of operational definitions for causal understanding, causal information and other similar notions has severely constrained the development of this field. Recognising the crucial role played by the concept of causal learning and understanding in several influential works [63–66] (see also [5] for a recent review), Starzak and Gray [62] have outlined a conceptual space for causal cognition characterised by three dimensions: explicitness, sources and integration of causal information. In the present work, we introduced a formal framework that provides more rigorous and clear underpinnings for those dimensions, and offers some precise coordinates to study various aspects of causal cognition.

More specifically, we defined levels of explicitness in terms of degrees of disentanglement [119,143,267,345,121], i.e. degrees of factorisation of a representation, and grouped approaches to disentanglement in macro categories that we introduced under the names of weak and strong disentanglement, based on how much causal structure they can take into account. We then classified sources of causal information in terms of where this information comes from, i.e. egocentric (an agent's own experience), social (from other agents), and natural sources (the "physics" of an agent's environment) [66,65]. Using the idea that in causal RL this information is usually stored on a replay buffer, or experience replay [6], we then operationalised integration as the ability to fuse pairs of different sources, or even all three of them at the same time.

Finally, we used this framework to conduct a comparative study of causal information processing as seen through the lenses of animal cognition and reinforcement learning research, with the former exploring areas that could inspire the latter, and the latter showcasing concrete proposals for computational and process theories of causal cognition missing in the former. In future work, we will aim to build both (1) new computational models for causal learning in animal cognition and (2) algorithms implementing more powerful forms of causal cognition in reinforcement learning, with the goal of showcasing the advancements of an integrated, unifying approach based on the work we presented here.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Part of this work involves research conducted by F.T. during his PhD, supported by a Leverhulme Doctoral Scholarship under the doctoral programme "From sensation and perception to awareness" at the University of Sussex (https://www.sussex.ac.uk/sensation/).

References

[1] Gopnik A, Schulz L. Causal learning: psychology, philosophy, and computation. Oxford University Press; 2007.
[2] Vallortigara G, Chiandetti C, Rugani R, Sovrano VA, Regolin L. Animal cognition. WIREs Cogn Sci 2010;1:882–93. https://doi.org/10.1002/wcs.75.
[3] McCormack T, Hoerl C, Butterfill S. Tool use and causal cognition, consciousness and self-consciousness. Oxford University Press; 2011.
[4] Sloman SA, Lagnado D. Causality in thought. Annu Rev Psychol 2015;66:223–47. https://doi.org/10.1146/annurev-psych-010814-015135.
[5] Goddu MK, Gopnik A. The development of human causal learning and reasoning. Nat Rev Psychol 2024:1–21. https://doi.org/10.1038/s44159-024-00300-5.
[6] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature 2015;518:529–33. https://doi.org/10.1038/nature14236.
[7] Reed S, Zolna K, Parisotto E, Colmenarejo SG, Novikov A, Barth-maron G, et al. Trans Mach Learn Res 2022.
[8] Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. Behav Brain Sci 2017;40:1–72. https://doi.org/10.1017/S0140525X16001837.
[9] Pearl J, Mackenzie D. The book of why: the new science of cause and effect. Allen Lane; 2018.
[10] Crosby M, Beyret B, Shanahan M, Hernández-Orallo J, Cheke L, Halina M. The animal-AI testbed and competition. In: Escalante HJ, Hadsell R, editors. Proceedings of the NeurIPS 2019 competition and demonstration track. Vancouver, CA: PMLR; 2020. p. 164–76.

[11] Shevlin H, Vold K, Crosby M, Halina M. The limits of machine intelligence. EMBO Rep 2019;20:1–5. https://doi.org/10.15252/embr.201949177.
[12] Schölkopf B. Artificial intelligence: learning to see and act. Nature 2015;518:486–7. https://doi.org/10.1038/518486a.
[13] Levine S. Understanding the world through action. In: Proceedings of the 5th Conference on Robot Learning. PMLR, vol. 164. 2022. p. 1752–7. https://doi.org/10.48550/arXiv.2110.12543. arXiv:2110.12543v1 [cs.LG], 2021.
[14] LeCun Y. A Path Towards Autonomous Machine Intelligence. 2022.
[15] Goyal A, Bengio Y. Inductive biases for deep learning of higher-level cognition. Proc R Soc A, Math Phys Eng Sci 2022;478:20210068. https://doi.org/10.1098/rspa.2021.0068.
[16] Gupta T, Gong W, Ma C, Pawlowski N, Hilmkil A, Scetbon M, et al. The essential role of causality in foundation world models for embodied AI. https://doi.org/10.48550/arXiv.2402.06665. arXiv:2402.06665v2 [cs.AI], 2024.
[17] Sperber D, Premack D, Premack AJ, editors. Causal cognition: a multidisciplinary debate. Clarendon Press; 1995.
[18] Gopnik A, Sobel DM, Danks D, Glymour C, Schulz LE, Kushnir T. A theory of causal learning in children: causal maps and Bayes nets. Psychol Rev 2004;111:3–32. https://doi.org/10.1037/0033-295X.111.1.3.
[19] Tenenbaum J, Griffiths T. Theory-based causal inference. Advances in neural information processing systems, vol. 15. The MIT Press; 2002.
[20] Griffiths TL, Tenenbaum JB. Structure and strength in causal induction. Cogn Psychol 2005;51:334–84. https://doi.org/10.1016/j.cogpsych.2005.05.004.
[21] Sloman S. Causal models: how people think about the world and its alternatives. Oxford, UK: Oxford University Press; 2005.
[22] Penn DC, Povinelli DJ. Causal cognition in human and nonhuman animals: a comparative, critical review. Annu Rev Psychol 2007;58:97–118. https://doi.org/10.1146/annurev.psych.58.110405.085555.
[23] Gopnik A, Sobel DM. Detecting blickets: how young children use information about novel causal powers in categorization and induction. Child Dev 2000;71:1205–22. https://doi.org/10.1111/1467-8624.00224.
[24] Gopnik A, Sobel DM, Schulz LE, Glymour C. Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Dev Psychol 2001;37:620–9. https://doi.org/10.1037/0012-1649.37.5.620.
[25] Griffiths TL, Tenenbaum JB. Theory-based causal induction. Psychol Rev 2009;116:661–716. https://doi.org/10.1037/a0017201.
[26] Sutton RS, Barto AG. Reinforcement learning: an introduction, adaptive computation and machine learning. The MIT Press; 2018.
[27] Bruineberg J, Dołęga K, Dewhurst J, Baltieri M. The emperor's new Markov blankets. Behav Brain Sci 2022;45:e183. https://doi.org/10.1017/S0140525X21002351.
[28] Pathak D, Agrawal P, Efros AA, Darrell T. Curiosity-driven exploration by self-supervised prediction. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2017.
[29] Hafner D, Lillicrap T, Fischer I, Villegas R, Ha D, Lee H, et al. Learning latent dynamics for planning from pixels. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97. PMLR; 2019. p. 2555–65.
[30] Hafner D, Lillicrap T, Ba J, Norouzi M. Dream to control: learning behaviors by latent imagination. In: International Conference on Learning Representations; 2020.
[31] Mnih V, Badia AP, Mirza L, Graves A, Harley T, Lillicrap TP, et al. Asynchronous methods for deep reinforcement learning. In: 33rd International Conference on Machine Learning, ICML 2016; 2016. p. 2850–69.
[32] Schölkopf B, von Kügelgen J. From statistical to causal learning. In: Proc Int Congr Math, vol. 7. 2022. p. 5540–93. https://doi.org/10.48550/ARXIV.2204.00607. arXiv:2204.00607v1 [cs.AI], 2022.
[33] Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, et al. Toward causal representation learning. Proc IEEE 2021;109:612–34. https://doi.org/10.1109/JPROC.2021.3058954.
[34] Peters J, Janzing D, Schölkopf B. Elements of causal inference: foundations and learning algorithms (adaptive computation and machine learning series). The MIT Press; 2017.
[35] Peters J, Bühlmann P, Meinshausen N. Causal inference by using invariant prediction: identification and confidence intervals. J R Stat Soc, Ser B, Stat Methodol 2016;78:947–1012. https://doi.org/10.1111/rssb.12167.
[36] Annadani Y, Rothfuss J, Lacoste A, Scherrer N, Goyal A, Bengio Y, et al. Variational causal networks: approximate Bayesian inference over causal structures. pp. 1–14. arXiv:2106.07635v1 [cs.LG], 2021.
[37] Faria GRA, Martins A, Figueiredo MAT. Differentiable causal discovery under latent interventions. In: Proceedings of machine learning research; 2022. p. 1–22.
[38] Lorch L, Sussex S, Rothfuss J, Krause A, Schölkopf B. Amortized inference for causal structure learning. In: Advances in neural information processing systems, vol. 35. 2022. https://doi.org/10.48550/ARXIV.2205.12934. arXiv:2205.12934v1 [cs.LG], 2022.
[39] Löwe S, Madras D, Zemel R, Welling M. Amortized causal discovery: learning to infer causal graphs from time-series data. In: Proceedings of Machine Learning Research, vol. 140. 2022. p. 1–24.
[40] Ke NR, Wang JX, Mitrovic J, Szummer M, Rezende DJ. Amortized learning of neural causal representations. pp. 1–9. arXiv:2008.09301v1 [stat.ML], 2020.
[41] Sontakke SA, Mehrjou A, Itti L, Schölkopf B. Causal curiosity: RL agents discovering self-supervised experiments for causal representation learning. In: Meila M, Zhang T, editors. Proceedings of the 38th International Conference on Machine Learning, vol. 139. PMLR; 2021. p. 9848–58.
[42] Seitzer M, Schölkopf B, Martius G. Causal influence detection for improving efficiency in reinforcement learning. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021. p. 22905–18.
[43] Rezende DJ, Danihelka I, Papamakarios G, Ke NR, Jiang R, Weber T, et al. Causally correct partial models for reinforcement learning. pp. 1–28. arXiv:2002.02836v1 [cs.LG], 2020.
[44] Huang B, Lu C, Leqi L, Hernandez-Lobato JM, Glymour C, Schölkopf B, et al. Action-sufficient state representation learning for control with structural constraints. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, editors. Proceedings of the 39th international conference on machine learning. Proceedings of machine learning research, vol. 162. PMLR; 2022. p. 9260–79.
[45] Zholus A, Ivchenkov Y, Panov A. Factorized world models for learning causal relationships. In: Workshop on the elements of reasoning: objects, structure and causality; 2022. p. 1–9.
[46] Li M, Yang M, Liu F, Chen X, Chen Z, Wang J. Causal world models by unsupervised deconfounding of physical dynamics. https://arxiv.org/abs/2012.14228, 2020.
[47] Lei A, Schölkopf B, Posner I. Variational causal dynamics: discovering modular world models from interventions. arXiv:2206.11131v1 [cs.LG], 2022.
[48] Goyal A, Lamb A, Hoffmann J, Sodhani S, Levine S, Bengio Y, et al. Recurrent independent mechanisms. https://doi.org/10.48550/arXiv.1909.10893. arXiv:1909.10893v6 [cs.LG], 2020.
[49] Javed K, White M, Bengio Y. Learning causal models online. arXiv:2006.07461, 2020.
[50] Brawer J, Qin M, Scassellati B. A causal approach to tool affordance learning. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2021. p. 8394–9.
[51] Hellström T. The relevance of causation in robotics: a review, categorization, and analysis. Paladyn 2021;12:238–55. https://doi.org/10.1515/pjbr-2021-0017.
[52] Ahmed O, Träuble F, Goyal A, Neitz A, Bengio Y, Schölkopf B, et al. CausalWorld: a robotic manipulation benchmark for causal structure and transfer learning. pp. 1–18. arXiv:2010.04296v2 [cs.RO], 2020.
[53] Weichwald S, Mogensen SW, Lee TE, Baumann D, Kroemer O, Guyon I, et al. Learning by doing: controlling a dynamical system using causality, control, and reinforcement learning. In: Proceedings of the NeurIPS 2021 competitions and demonstrations track. PMLR; 2022. p. 246–58.

[54] Liu Y, Alahi A, Russell C, Horn M, Zietlow D, Schölkopf B, et al. Causal triplet: an open challenge for intervention-centric causal representation learning. In: Proceedings of the second conference on causal learning and reasoning. PMLR, vol. 213. 2023. p. 553–73. https://doi.org/10.48550/arXiv.2301.05169. arXiv:2301.05169v2 [cs.LG], 2023.
[55] Beyret B, Hernández-Orallo J, Cheke L, Halina M, Shanahan M, Crosby M. The animal-AI environment: training and testing animal-like artificial cognition. pp. 1–14. arXiv:1909.07483v2 [cs.LG], 2019.
[56] Crosby M, Beyret B, Halina M. The animal-AI olympics. Nat Mach Intell 2019;1:257. https://doi.org/10.1038/s42256-019-0050-3.
[57] Nalmpantis A, Lippe P, Magliacane S. Hierarchical causal representation learning. In: Causal representation learning workshop at NeurIPS 2023; 2023.
[58] Talon D, Lippe P, James S, Bue AD, Magliacane S. Towards the reusability and compositionality of causal representations. In: Proceedings of the third conference on causal learning and reasoning. PMLR, vol. 236. 2024. p. 296–324. https://doi.org/10.48550/arXiv.2403.09830. arXiv:2403.09830v1 [cs.LG], 2024.
[59] Whittington JCR, McCaffary D, Bakermans JJW, Behrens TEJ. How to build a cognitive map. Nat Neurosci 2022;25:1257–72. https://doi.org/10.1038/s41593-022-01153-y.
[60] Whittington JCR, Dorrell W, Ganguli S, Behrens T. Disentanglement with biological constraints: a theory of functional cell types. In: The eleventh International Conference on Learning Representations; 2023.
[61] Courellis HS, Minxha J, Cardenas AR, Kimmel DL, Reed CM, Valiante TA, et al. Abstract representations emerge in human hippocampal neurons during inference. Nature 2024;632:841–9. https://doi.org/10.1038/s41586-024-07799-x.
[62] Starzak TB, Gray RD. Towards ending the animal cognition war: a three-dimensional model of causal cognition. Biol Philos 2021;36:1–24. https://doi.org/10.1007/s10539-021-09779-1.
[63] Woodward J. Causation with a human face: normative theory and descriptive psychology. Oxford studies in philosophy of science. Oxford University Press; 2021.
[64] Woodward J. Causation: interactions between philosophical theories and psychological research. Philos Sci 2012;79:961–72. https://doi.org/10.1086/667850.
[65] Woodward J. A philosopher looks at tool use and causal understanding. In: McCormack T, Hoerl C, Butterfill S, editors. Tool use and causal cognition. Consciousness and self-consciousness. Oxford, UK: Oxford University Press; 2011. p. 18–50.
[66] Woodward J. Interventionist theories of causation in psychological perspective. In: Gopnik A, Schulz L, editors. Causal learning: psychology, philosophy, and computation. Oxford, UK: Oxford University Press; 2007. p. 19–36.
[67] Woodward J. Making things happen: a theory of causal explanation. Oxford studies in philosophy of science. Oxford University Press; 2005.
[68] Kelley HH. The processes of causal attribution. Am Psychol 1973;28:107–28. https://doi.org/10.1037/h0034225.
[69] Cheng PW. From covariation to causation: a causal power theory. Psychol Rev 1997;104:367–405. https://doi.org/10.1037/0033-295X.104.2.367.
[70] Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black A, Prokasy W, editors. Classical conditioning II: current research and theory. New York: Appleton Century Crofts; 1972. p. 64–99.
[71] Shanks DR, Dickinson A. Associative accounts of causality judgment. The psychology of learning and motivation: advances in research and theory, vol. 21. San Diego, CA, US: Academic Press; 1987. p. 229–61.
[72] Shanks DR. Associationism and cognition: human contingency learning at 25. Q J Exp Psychol 2007;60:291–309. https://doi.org/10.1080/17470210601000581.
[73] Dickinson A. Causal learning: association versus computation. Curr Dir Psychol Sci 2001;10:127–32.
[74] Dickinson A. The 28th Bartlett memorial lecture causal learning: an associative analysis. Q J Exp Psychol Sect B 2001;54:3–25. https://doi.org/10.1080/713932741.
[75] Waldmann MR, Holyoak KJ. Can causal induction be reduced to associative learning? In: Proceedings of the twelfth annual conference of the cognitive science society. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1990. p. 190–7.
[76] Waldmann MR, Holyoak KJ. Predictive and diagnostic learning within causal models: asymmetries in cue competition. J Exp Psychol Gen 1992;121:222–36. https://doi.org/10.1037/0096-3445.121.2.222.
[77] Waldmann MR, Holyoak KJ, Fratianne A. Causal models and the acquisition of category structure. J Exp Psychol Gen 1995;124:181–206. https://doi.org/10.1037/0096-3445.124.2.181.
[78] Blaisdell AP, Sawa K, Leising KJ, Waldmann MR. Causal reasoning in rats. Science 2006;311:1020–2. https://doi.org/10.1126/science.1121872.
[79] Dickinson A. Associative learning and animal cognition. Philos Trans R Soc Lond B, Biol Sci 2012;367:2733–42. https://doi.org/10.1098/rstb.2012.0220.
[80] Buckner C. Two approaches to the distinction between cognition and 'mere association'. Int J Comp Psychol 2011;24. https://doi.org/10.46867/ijcp.2011.24.04.06.
[81] Heyes C. Simple minds: a qualified defence of associative learning. Philos Trans R Soc Lond B, Biol Sci 2012;367:2695–703. https://doi.org/10.1098/rstb.2012.0217.
[82] Hanus D. Causal reasoning versus associative learning: a useful dichotomy or a strawman battle in comparative psychology? J Comp Psychol 2016;130:241–8. https://doi.org/10.1037/a0040235.
[83] Lyon P. Of what is "minimal cognition" the half-baked version? Adapt Behav 2020;28:407–24. https://doi.org/10.1177/1059712319871360.
[84] Baluška F, Levin M. On having no head: cognition throughout biological systems. Front Psychol 2016;7:1–19. https://doi.org/10.3389/fpsyg.2016.00902.
[85] Barandiaran X, Moreno A. On what makes certain dynamical systems cognitive: a minimally cognitive organization program. Adapt Behav 2006;14:171–85. https://doi.org/10.1177/105971230601400208.
[86] Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 2013;36:181–204. https://doi.org/10.1017/S0140525X12000477.
[87] Hohwy J. The predictive mind. Oxford, UK: Oxford University Press; 2013.
[88] Chater N, Oaksford M, Hahn U, Heit E. Bayesian models of cognition. Wiley Interdiscip Rev Cogn Sci 2010;1:811–23. https://doi.org/10.1002/wcs.79.
[89] Visalberghi E, Trinca L. Tool use in capuchin monkeys: distinguishing between performing and understanding. Primates 1989;30:511–21. https://doi.org/10.1007/BF02380877.
[90] Visalberghi E, Limongelli L. Tool use in capuchins (Cebus apella): is there an understanding of the cause-effect relationship? Ethol Ecol Evol 1993;5:419–20. https://doi.org/10.1080/08927014.1993.9523092.
[91] Visalberghi E, Limongelli L. Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (Cebus apella). J Comp Psychol 1994;108:15–22. https://doi.org/10.1037/0735-7036.108.1.15.
[92] Visalberghi E, Limongelli L. Acting and understanding: tool use revisited through the minds of capuchin monkeys. In: Russon AE, Bard KA, Parker ST, editors. Reaching into thought: the minds of the great apes. Cambridge, UK: Cambridge University Press; 1996.
[93] Limongelli L, Boysen ST, Visalberghi E. Comprehension of cause-effect relations in a tool-using task by chimpanzees (Pan Troglodytes). J Comp Psychol 1995;109:18–26. https://doi.org/10.1037/0735-7036.109.1.18.
[94] Mulcahy NJ, Call J. How great apes perform on a modified trap-tube task. Anim Cogn 2006;9:193–9. https://doi.org/10.1007/s10071-006-0019-6.
[95] Seed AM, Call J, Emery NJ, Clayton NS. Chimpanzees solve the trap problem when the confound of tool-use is removed. J Exp Psychol, Anim Behav Processes 2009;35:23–34. https://doi.org/10.1037/a0012925.
[96] Martin-Ordas G, Call J, Colmenares F. Tubes, tables and traps: great apes solve two functionally equivalent trap tasks but show no evidence of transfer across tasks. Anim Cogn 2008;11:423–30. https://doi.org/10.1007/s10071-007-0132-1.
[97] Visalberghi E, Tomasello M. Primate causal understanding in the physical and psychological domains. Behav Process 1998;42:189–203. https://doi.org/10.1016/S0376-6357(97)00076-4.

[98] Leising KJ, Wong J, Waldmann MR, Blaisdell AP. The special status of actions in causal reasoning in rats. J Exp Psychol Gen 2008;137:514–27. https://doi.org/10.1037/0096-3445.137.3.514.
[99] Völter CJ, Sentís I, Call J. Great apes and children infer causal relations from patterns of variation and covariation. Cognition 2016;155:30–43. https://doi.org/10.1016/j.cognition.2016.06.009.
[100] Taylor AH, Cheke LG, Waismeyer A, Meltzoff AN, Miller R, Gopnik A, et al. Of babies and birds: complex tool behaviours are not sufficient for the evolution of the ability to create a novel causal intervention. Proc R Soc Lond B, Biol Sci 2014;281:20140837. https://doi.org/10.1098/rspb.2014.0837.
[101] Jacobs IF, von Bayern A, Martin-Ordas G, Rat-Fischer L, Osvath M. Corvids create novel causal interventions after all. Proc R Soc Lond B, Biol Sci 2015;282:20142504. https://doi.org/10.1098/rspb.2014.2504.
[102] Taylor A, Hunt G, Medina F, Gray R. Do New Caledonian crows solve physical problems through causal reasoning? Proc R Soc Lond B, Biol Sci 2009;276:247–54. https://doi.org/10.1098/rspb.2008.1107.
[103] Jelbert SA, Taylor AH, Cheke LG, Clayton NS, Gray RD. Using the Aesop's fable paradigm to investigate causal understanding of water displacement by new caledonian crows. PLoS ONE 2014;9. https://doi.org/10.1371/journal.pone.0092895.
[104] Logan CJ, Jelbert SA, Breen AJ, Gray RD, Taylor AH. Modifications to the Aesop's fable paradigm change New Caledonian crow performances. PLoS ONE 2014;9:e103049. https://doi.org/10.1371/journal.pone.0103049.
[105] Miller R, Jelbert SA, Taylor AH, Cheke LG, Gray RD, Loissel E, et al. Performance in object-choice Aesop's fable tasks are influenced by object biases in New Caledonian crows but not in human children. PLoS ONE 2016;11:e0168056. https://doi.org/10.1371/journal.pone.0168056.
[106] Chappell J. Avian cognition: understanding tool use. Curr Biol 2006;16:R244–5. https://doi.org/10.1016/j.cub.2006.03.019.
[107] Hennefield L, Hwang HG, Weston SJ, Povinelli DJ. Meta-analytic techniques reveal that corvid causal reasoning in the Aesop's Fable paradigm is driven by trial-and-error learning. Anim Cogn 2018;21:735–48. https://doi.org/10.1007/s10071-018-1206-y.
[108] Hennefield L, Hwang HG, Povinelli DJ. Going meta: retelling the scientific retelling of Aesop's the crow and the pitcher. J Folk Res 2019;56:45–69. https://doi.org/10.2979/jfolkrese.56.2_3.04.
[109] Seed AM, Hanus D, Call J. Causal knowledge in corvids, primates, and children: more than meets the eye? In: McCormack T, Hoerl C, Butterfill S, editors. Tool use and causal cognition. Consciousness and self-consciousness. Oxford, UK: Oxford University Press; 2011. p. 89–110.
[110] Povinelli DJ, Penn DC. Through a floppy tool darkly: toward a conceptual overthrow of animal alchemy. In: McCormack T, Hoerl C, Butterfill S, editors. Tool use and causal cognition. Consciousness and self-consciousness. Oxford, UK: Oxford University Press; 2011. p. 69–88.
[111] Dickinson A, Balleine BW. Causal cognition and goal-directed action. In: Heyes C, Huber L, editors. The evolution of cognition. The MIT Press; 2000. p. 185–204.
[112] Pearl J. Causality: models, reasoning, and inference. 2 ed. Cambridge, UK: Cambridge University Press; 2009.
[113] Piccinini G, Scarantino A. Information processing, computation, and cognition. J Biol Phys 2011;37:1–38. https://doi.org/10.1007/s10867-010-9195-3.
[114] Simoes FNFQ, Dastani M, van Ommen T. Causal entropy and information gain for measuring causal control. In: Nowaczyk S, Biecek P, Chung NC, Vallati M, Skruch P, Jaworek-Korjakowska J, et al., editors. Artificial intelligence. ECAI 2023 international workshops. Switzerland, Cham: Springer Nature; 2024. p. 216–31.
[115] Simoes FNFQ, Dastani M, van Ommen T. Fundamental properties of causal entropy and information gain. In: Proceedings of the third conference on causal cearning and reasoning. PMLR; 2024. p. 188–208.
[116] Mascalzoni E, Regolin L, Vallortigara G. Innate sensitivity for self-propelled causal agency in newly hatched chicks. Proc Natl Acad Sci 2010;107:4483–5. https://doi.org/10.1073/pnas.0908792107.
[117] Lemaire BS, Vallortigara G. Life is in motion (through a chick's eye). Anim Cogn 2023;26:129–40. https://doi.org/10.1007/s10071-022-01703-8.
[118] Tomasello M, Call J. Primate cognition. Oxford University Press; 1997.
[119] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35:1798–828. https://doi.org/10.1109/TPAMI.2013.50.
[120] Wang X, Chen H, Tang S, Wu Z, Zhu W. Disentangled representation learning. IEEE Trans Pattern Anal Mach Intell 2024:1–20. https://doi.org/10.1109/TPAMI.2024.3420937.
[121] Zhang Y, Sugiyama M. A category-theoretical meta-analysis of definitions of disentanglement. https://doi.org/10.48550/arXiv.2305.06886. pp. 1–17. arXiv:2305.06886v2 [cs.LG], 2023.
[122] Mac Lane S. Categories for the working mathematician. 2 ed. Graduate texts in mathematics, vol. 5. Springer; 2013.
[123] Suter R, Miladinović Đ, Schölkopf B, Bauer S. Robustly disentangled causal mechanisms: validating deep representations for interventional robustness. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th International Conference on Machine Learning. Proceedings of machine learning research, vol. 97. 2019. p. 6056–65.
[124] Zhang Y, Sugiyama M. Enriching disentanglement: from logical definitions to quantitative metrics. https://doi.org/10.48550/arXiv.2305.11512. arXiv:2305.11512v2 [cs.LG], 2024.
[125] Perrone P. Markov categories and entropy. IEEE Trans Inf Theory 2024;70:1671–92. https://doi.org/10.1109/TIT.2023.3328825.
[126] Wang Y, Jordan MI. Desiderata for representation learning: a causal perspective. J Mach Learn Res 2024;25(275):1–65. arXiv:2109.03795v2 [stat.ML], 2022. arXiv:2109.03795v2.
[127] Garrabrant S. Temporal inference with finite factored sets. https://doi.org/10.48550/arXiv.2109.11513. arXiv:2109.11513, 2021.
[128] Kaddour J, Lynch A, Liu Q, Kusner MJ, Silva R. Causal machine learning: a survey and open problems. pp. 1–191. https://doi.org/10.48550/arXiv.2206.15475. arXiv:2206.15475v2 [cs.LG], 2022.
[129] Tibshirani R. The elements of statistical learning: data mining, inference, and prediction. second edition. Springer series in statistics. New York, NY: Springer; 2009.
[130] Hernán MA, Robins JM. Causal inference: what if. Boca Raton: Chapman & Hall/CRC; 2020.
[131] Berrevoets J, Kacprzyk K, Qian Z, van der Schaar M. Causal deep learning. pp. 1–31. https://doi.org/10.48550/arXiv.2303.02186. arXiv:2303.02186v1 [cs.LG], 2023.
[132] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. Cambridge, Massachusetts: The MIT Press; 2000.
[133] Zhang K, Hyvärinen A. On the identifiability of the post-nonlinear causal model. In: Proceedings of the 25th conference on uncertainty in artificial intelligence, UAI '09. Arlington, Virginia, USA: AUAI Press; 2009. p. 647–55.
[134] Zhang K, Wang Z, Zhang J, Schölkopf B. On estimation of functional causal models: general results and application to the post-nonlinear causal model. ACM Trans Intell Syst Technol 2015;7:13:1–1322:. https://doi.org/10.1145/2700476.
[135] Goudet O, Kalainathan D, Caillou P, Guyon I, Lopez-Paz D, Sebag M. Learning functional causal models with generative neural networks. In: Escalante HJ, Escalera S, Guyon I, Baró X, Güçlütürk Y, Güçlü U, et al., editors. Explainable and interpretable models in computer vision and machine learning. Springer; 2018. p. 39–80.
[136] Gresele L, von Kügelgen J, Stimper V, Schölkopf B, Besserve M. Independent mechanism analysis, a new concept? In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021. p. 28233–48.
[137] Hedges J, Sakamoto RR. Reinforcement learning in categorical cybernetics. https://doi.org/10.48550/arXiv.2404.02688, 2024. arXiv:2404.02688v1 [cs.LG].
[138] Gershman SJ, Niv Y. Learning latent structure: carving nature at its joints. Cogn Neurosci 2010;20:251–6. https://doi.org/10.1016/j.conb.2010.02.008.
[139] Gershman SJ, Norman KA, Niv Y. Discovering latent causes in reinforcement learning. Curr Opin Behav Sci 2015;5:43–50. https://doi.org/10.1016/j.cobeha.2015.07.007.

[140] Gershman SJ. Reinforcement learning and causal models. In: Waldmann MR, editor. The Oxford handbook of causal reasoning. Oxford University Press; 2017.
[141] Kingma DP, Welling M. Auto-encoding variational Bayes. pp. 1–14. https://doi.org/10.48550/arXiv.1312.6114. arXiv:1312.6114v11 [stat.ML], 2022.
[142] Kingma DP, Welling M. An introduction to variational autoencoders. Found Trends Mach Learn 2019;12:307–92. https://doi.org/10.1561/2200000056.
[143] Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, et al. β-VAE: learning basic visual concepts with a constrained variational framework. In: International conference on learning representations; 2017. p. 1–22.
[144] Doersch C. Tutorial on variational autoencoders. pp. 1–23. https://doi.org/10.48550/arXiv.1606.05908. arXiv:1606.05908v3 [stat.ML], 2021.
[145] Burgess CP, Higgins I, Pal A, Matthey L, Watters N, Desjardins G, et al. Understanding disentangling in β-VAE. pp. 1–11. https://doi.org/10.48550/ARXIV.1804.03599. arXiv:1804.03599v1 [stat.ML], 2018.
[146] Kim H, Mnih A. Disentangling by factorising. In: Dy J, Krause A, editors. Proceedings of the 35th international conference on machine learning. Proceedings of machine learning research, vol. 80. PMLR; 2018. p. 2649–58.
[147] Chen RTQ, Li X, Grosse R, Duvenaud D. Isolating sources of disentanglement in variational autoencoders. In: Advances in neural information processing systems (NeurIPS 2018), vol. 31. 2018. https://doi.org/10.48550/arXiv.1802.04942. arXiv:1802.04942v5 [cs.LG], 2019.
[148] Rubenstein PK, Schoelkopf B, Tolstikhin I. Learning disentangled representations with Wasserstein auto-encoders. In: Workshop at the 6th international conference on learning representations (ICLR); 2018.
[149] Ridgeway K, Mozer MC. Learning deep disentangled embeddings with the F-statistic loss. Advances in neural information processing systems, vol. 31. Curran Associates, Inc.; 2018.
[150] Eastwood C, Williams CKI. A framework for the quantitative evaluation of disentangled representations. In: International conference on learning representations; 2018.
[151] Locatello F, Bauer S, Lucic M, Rätsch G, Gelly S, Schölkopf B, et al. Challenging common assumptions in the unsupervised learning of disentangled representations. In: Proceedings of the 36th international conference on machine learning, vol. 97. 2019. p. 4114–24.
[152] Locatello F, Poole B, Rätsch G, Schölkopf B, Bachem O, Tschannen M. Weakly-supervised disentanglement without compromises. In: D. III H, Singh A, editors. Proceedings of the 37th international conference on machine learning. Proceedings of machine learning research, vol. 119. PMLR; 2020. p. 6348–59.
[153] Träuble F, Creager E, Kilbertus N, Locatello F, Dittadi A, Goyal A, et al. On disentangled representations learned from correlated data. In: Proceedings of the 38th international conference on machine learning, PMLR, vol. 139. 2021. p. 10401–12. arXiv:2006.07886v3 [cs.LG], 2021.
[154] Shu R, Chen Y, Kumar A, Ermon S, Poole B. Weakly supervised disentanglement with guarantees. In: International conference on learning representations; 2020. p. 1–36.
[155] Khemakhem I, Kingma D, Monti R, Hyvarinen A. Variational autoencoders and nonlinear ICA: a unifying framework. In: Chiappa S, Calandra R, editors. Proceedings of the twenty third international conference on artificial intelligence and statistics. Proceedings of machine learning research, vol. 108. PMLR; 2020. p. 2207–17.
[156] Sepliarskaia A, Kiseleva J, de Rijke M. How to not measure disentanglement. https://doi.org/10.48550/arXiv.1910.05587. arXiv:1910.05587v3 [cs.LG], 2021.
[157] Do K, Tran T. Theory and evaluation metrics for learning disentangled representations. https://doi.org/10.48550/arXiv.1908.09961. arXiv:1908.09961v3 [cs.LG], 2021.
[158] Tishby N, Pereira FC, Bialek W. The information bottleneck method. pp. 1–16. https://doi.org/10.48550/arXiv.physics/0004057. arXiv:physics/0004057v1 [physics.data-an], 2000.
[159] Tishby N, Zaslavsky N. Deep learning and the information bottleneck principle. In: 2015 IEEE information theory workshop (ITW); 2015. p. 1–5.
[160] Still S. Thermodynamic cost and benefit of memory. Phys Rev Lett 2020;124:050601. https://doi.org/10.1103/PhysRevLett.124.050601.
[161] Daimer D, Still S. Thermodynamically rational decision making under uncertainty. https://doi.org/10.48550/arXiv.2309.10476. arXiv:2309.10476v2 [physics.data-an], 2023.
[162] Alemi AA, Fischer I, Dillon JV, Murphy K. Deep variational information bottleneck. In: International conference on learning representations; 2017. p. 1–19.
[163] Lemeire J, Dirkx E. Causal models as minimal descriptions of multivariate systems. 2012. p. 1–16.
[164] Hyvärinen A, Karhunen J, Oja E. Independent component analysis. John Wiley & Sons; 2001.
[165] Hyvärinen A. Independent component analysis: recent advances. Philos Trans R Soc A, Math Phys Eng Sci 2013;371:20110534. https://doi.org/10.1098/rsta.2011.0534.
[166] Wendong L, Kekić A, von Kügelgen J, Buchholz S, Besserve M, Gresele L, et al. Causal component analysis. In: Advances in neural information processing systems, vol. 36. 2023. p. 32481–520.
[167] Xu D, Yao D, Lachapelle S, Taslakian P, von Kügelgen J, Locatello F, et al. A sparsity principle for partially observable causal representation learning. In: Proceedings of the 41st international conference on machine learning. PMLR, vol. 235. 2024. p. 55389–433. https://doi.org/10.48550/arXiv.2403.08335. arXiv:2403.08335v2 [cs.LG], 2024.
[168] Yang M, Liu F, Chen Z, Shen X, Hao J, Wang J. CausalVAE: disentangled representation learning via neural structural causal models. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Nashville, TN, USA: IEEE; 2021. p. 9588–97.
[169] Yao W, Sun Y, Ho A, Sun C, Zhang K. Learning temporally causal latent processes from general temporal data. In: International conference on learning representations; 2022.
[170] Lin L-J. Programming robots using reinforcement learning and teaching. In: Proceedings of the ninth national conference on artificial intelligence - volume 2, AAAI'91. Anaheim, California: AAAI Press; 1991. p. 781–6.
[171] Schulman J, Moritz P, Levine S, Jordan M, Abbeel P. High-dimensional continuous control using generalized advantage estimation. https://doi.org/10.48550/arXiv.1506.02438. arXiv:1506.02438v6 [cs.LG], 2018.
[172] Gu S, Lillicrap T, Ghahramani Z, Turner RE, Levine S. Q-prop: sample-efficient policy gradient with an off-policy critic. In: International conference on learning representations; 2017.
[173] Degris T, White M, Sutton RS. Off-policy actor-critic. In: Proceedings of the 29th international conference on machine learning, ICML'12. Madison, WI, USA: Omnipress; 2012. p. 179–86.
[174] Beck J, Vuorio R, Liu EZ, Xiong Z, Zintgraf L, Finn C, et al. A survey of meta-reinforcement learning. pp. 1–53. https://doi.org/10.48550/ARXIV.2301.08028. arXiv:2301.08028v1 [cs.LG], 2023.
[175] Fu J, Kumar A, Nachum O, Tucker G, Levine S. D4RL: datasets for deep data-driven reinforcement learning. https://doi.org/10.48550/arXiv.2004.07219. arXiv:2004.07219v4 [cs.LG], 2021.
[176] Gulcehre C, Wang Z, Novikov A, Paine T, Gómez S, Zolna K, et al. A suite of benchmarks for offline reinforcement learning. Advances in neural information processing systems, vol. 33. Curran Associates, Inc.; 2020. p. 7248–59.
[177] Yarats D, Brandfonbrener D, Liu H, Laskin M, Abbeel P, Lazaric A, et al. Don't change the algorithm, change the data: exploratory data for offline reinforcement learning. https://doi.org/10.48550/arXiv.2201.13425. arXiv:2201.13425v3 [cs.LG], 2022.
[178] Zhou G, Ke L, Srinivasa S, Gupta A, Rajeswaran A, Kumar V. Real world offline reinforcement learning with realistic data source. In: 2023 IEEE international conference on robotics and automation (ICRA); 2023. p. 7176–83.
[179] Kahn H. Use of different Monte Carlo sampling techniques. Santa Monica, CA: RAND Corporation; 1955.
[180] Peshkin L, Shelton CR. Learning from scarce experience. In: Proceedings of the nineteenth international conference on machine learning, ICML '02. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2002. p. 498–505.
[181] Precup D, Sutton RS, Singh SP. Eligibility traces for off-policy policy evaluation. In: Proceedings of the seventeenth international conference on machine learning, ICML '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2000. p. 759–66.

[182] Koller D, Friedman N. Probabilistic graphical models: principles and techniques. Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press; 2009.
[183] Jie T, Abbeel P. On a connection between importance sampling and the likelihood ratio policy gradient. Advances in neural information processing systems, vol. 23. Curran Associates, Inc.; 2010.
[184] Levine S, Kumar A, Tucker G, Fu J. Offline reinforcement learning: tutorial, review, and perspectives on open problems. pp. 1–43. https://doi.org/10.48550/arXiv.2005.01643. arXiv:2005.01643v3 [cs.LG], 2020.
[185] Lorch L, Rothfuss J, Schölkopf B, Krause A. DiBS: differentiable Bayesian structure learning. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021.
[186] Ke NR, Chiappa S, Wang J, Bornschein J, Weber T, Goyal A, et al. Learning to induce causal structure. In: The eleventh international conference on learning representations, ICLR 2023; 2023. https://doi.org/10.48550/ARXIV.2204.04875. arXiv:2204.04875v1 [stat.ML], 2022.
[187] Scherrer N, Goyal A, Bauer S, Bengio Y, Ke NR. On the generalization and adaption performance of causal models. In: ICML 2022: Workshop on spurious correlations, invariance and stability, SCIS 2022; 2022. pp. 1–25. https://doi.org/10.48550/ARXIV.2206.04620. arXiv:2206.04620v1 [cs.LG], 2022.
[188] Deng Z, Jiang J, Long G, Zhang C. Causal reinforcement learning: a survey. Trans Mach Learn Res 2023.
[189] Sebastián-Enesco C, Amezcua-Valmala N, Colmenares F, Mendes N, Call J. Raising the level: orangutans solve the floating peanut task without visual feedback. Primates 2022;63:33–9. https://doi.org/10.1007/s10329-021-00952-4.
[190] Hanus D, Mendes N, Tennie C, Call J. Comparing the performances of apes (Gorilla gorilla, Pan troglodytes, Pongo pygmaeus) and human children (Homo sapiens) in the floating peanut task. PLoS ONE 2011;6:e19555. https://doi.org/10.1371/journal.pone.0019555.
[191] Ebel SJ, Schmelz M, Herrmann E, Call J. Innovative problem solving in great apes: the role of visual feedback in the floating peanut task. Anim Cogn 2019;22:791–805. https://doi.org/10.1007/s10071-019-01275-0.
[192] Tennie C, Völter CJ, Vonau V, Hanus D, Call J, Tomasello M. Chimpanzees use observed temporal directionality to learn novel causal relations. Primates 2019;60:517–24. https://doi.org/10.1007/s10329-019-00754-9.
[193] Pika S, Sima MJ, Blum CR, Herrmann E, Mundry R. Ravens parallel great apes in physical and social cognitive skills. Sci Rep 2020;10:20617. https://doi.org/10.1038/s41598-020-77060-8.
[194] Tennie C, Call J, Tomasello M. Evidence for emulation in chimpanzees in social settings using the floating peanut task. PLoS ONE 2010;5:e10544. https://doi.org/10.1371/journal.pone.0010544.
[195] Mendes N, Hanus D, Call J. Raising the level: orangutans use water as a tool. Biol Lett 2007;3:453–5. https://doi.org/10.1098/rsbl.2007.0198.
[196] Zhang A, Lipton ZC, Pineda L, Azizzadenesheli K, Anandkumar A, Itti L, et al. Learning causal state representations of partially observable environments. pp. 1–35. https://doi.org/10.48550/arXiv.1906.10437. arXiv:1906.10437v2 [cs.LG], 2021.
[197] de Haan P, Jayaraman D, Levine S. Causal confusion in imitation learning. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, editors. Advances in neural information processing systems, vol. 32. Curran Associates, Inc.; 2019.
[198] Wang Z, Xiao X, Xu Z, Zhu Y, Stone P. Causal dynamics learning for task-independent state abstraction. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, editors. Proceedings of the 39th international conference on machine learning. Proceedings of machine learning research, vol. 162. PMLR; 2022. p. 23151–80.
[199] Mutti M, Santi RD, Rossi E, Calderon JF, Bronstein M, Restelli M. Provably efficient causal model-based reinforcement learning for systematic generalization. Proc AAAI Conf Artif Intell 2023;37:9251–9. https://doi.org/10.1609/aaai.v37i8.26109.
[200] Taylor AH, Hunt GR, Holzhaider JC, Gray RD. Spontaneous metatool use by New Caledonian crows. Curr Biol 2007;17:1504–7. https://doi.org/10.1016/j.cub.2007.07.057.
[201] Taylor AH, Hunt GR, Gray RD. Context-dependent tool use in New Caledonian crows. Biol Lett 2012;8:205–7. https://doi.org/10.1098/rsbl.2011.0782.
[202] Taylor AH, Knaebe B, Gray RD. An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. Proc Biol Sci 2012;279:4977–81. https://doi.org/10.1098/rspb.2012.1998.
[203] Seed AM, Tebbich S, Emery NJ, Clayton NS. Investigating physical cognition in rooks, Corvus frugilegus. Curr Biol 2006;16:697–701. https://doi.org/10.1016/j.cub.2006.02.066.
[204] Jelbert SA, Miller R, Schiestl M, Boeckle M, Cheke LG, Gray RD, et al. New Caledonian crows infer the weight of objects from observing their movements in a breeze. Proc R Soc Lond B, Biol Sci 2019;286:20182332. https://doi.org/10.1098/rspb.2018.2332.
[205] Taylor AH, Miller R, Gray RD. New Caledonian crows reason about hidden causal agents. Proc Natl Acad Sci 2012;109:16389–91. https://doi.org/10.1073/pnas.1208724109.
[206] Buesing L, Weber T, Zwols Y, Racanière S, Guez A, Lespiau JB, et al. Woulda, coulda, shoulda: counterfactually-guided policy search. In: International conference on learning representations; 2019. p. 1–15.
[207] Zhang J, Kumor D, Bareinboim E. Causal imitation learning with unobserved confounders. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in neural information processing systems, vol. 33. Vancouver, Canada: Curran Associates, Inc.; 2020. p. 12263–74.
[208] Kumor D, Zhang J, Bareinboim E. Sequential causal imitation learning with unobserved confounders. In: Advances in neural information processing systems, vol. 34. Curran Associates Inc.; 2021. p. 14669–80.
[209] Wang L, Yang Z, Wang Z. Provably efficient causal reinforcement learning with confounded observational data. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021. p. 21164–75.
[210] Thomas V, Bengio E, Fedus W, Pondard J, Beaudoin P, Larochelle H, et al. Disentangling the independently controllable factors of variation by interacting with the world. pp. 1–9. https://doi.org/10.48550/arXiv.1802.09484. arXiv:1802.09484v1 [stat.ML], 2018.
[211] Wulfmeier M, Byravan A, Hertweck T, Higgins Gupta I, Kulkarni T, Reynolds M, et al. Representation matters: improving perception and exploration for robotics. In: 2021 IEEE international conference on robotics and automation (ICRA); 2021. p. 6512–9.
[212] Tomar M, Zhang A, Calandra R, Taylor ME, Pineau J. Model-invariant state abstractions for model-based reinforcement learning. In: Self-supervision for reinforcement learning workshop; 2021.
[213] Achille A, Eccles T, Matthey L, Burgess CP, Watters N, Lerchner A, et al. Life-long disentangled representation learning with cross-domain latent homologies. pp. 1–15. https://doi.org/10.48550/ARXIV.1808.06508. arXiv:1808.06508v1 [cs.LG], 2018.
[214] Laversanne-Finot A, Pere A, Oudeyer P-Y. Curiosity driven exploration of learned disentangled goal spaces. In: Billard A, Dragan A, Peters J, Morimoto J, editors. Proceedings of the 2nd conference on robot learning. Proceedings of machine learning research, vol. 87. PMLR; 2018. p. 487–504.
[215] Zhang A, Lyle C, Sodhani S, Filos A, Kwiatkowska M, Pineau J, et al. Invariant causal prediction for block MDPs. Proceedings of the 37th international conference on machine learning, vol. 119. PMLR; 2020.
[216] Higgins I, Pal A, Rusu A, Matthey L, Burgess C, Pritzel A, et al. DARLA: improving zero-shot transfer in reinforcement learning. In: Precup D, Teh YW, editors. Proceedings of the 34th international conference on machine learning. Proceedings of machine learning research, vol. 70. PMLR; 2017. p. 1480–90.
[217] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science 2018;362:1140–4. https://doi.org/10.1126/science.aar6404.
[218] Ke NR, Singh A, Touati A, Goyal A, Bengio Y, Parikh D, et al. Learning dynamics model in reinforcement learning by incorporating the long term future. https://doi.org/10.48550/ARXIV.1903.01599. arXiv:1903.01599v2 [stat.ML], 2019.
[219] Watters N, Matthey L, Bosnjak M, Burgess CP, Lerchner A. COBRA: data-efficient model-based RL through unsupervised object discovery and curiosity-driven exploration. pp. 1–24. https://doi.org/10.48550/arXiv.1905.09275. arXiv:1905.09275v2 [cs.LG], 2019.

[220] Mendonca R, Rybkin O, Daniilidis K, Hafner D, Pathak D. Discovering and achieving goals via world models. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021. p. 24379–91.
[221] Ha D, Schmidhuber J. World models. https://doi.org/10.5281/zenodo.1207631, 2018.
[222] Ebert F, Finn C, Dasari S, Xie A, Lee A, Levine S. Visual foresight: model-based deep reinforcement learning for vision-based robotic control. https://doi.org/10.48550/arXiv.1812.00568. arXiv:1812.00568v1 [cs.RO], 2018.
[223] Renner E, Abramo AM, Karen Hambright M, Phillips KA. Insightful problem solving and emulation in Brown capuchin monkeys. Anim Cogn 2017;20:531–6. https://doi.org/10.1007/s10071-017-1080-z.
[224] Renner E, Kean D, Atkinson M, Caldwell CA. The use of individual, social, and animated cue information by capuchin monkeys and children in a touchscreen task. Sci Rep 2021;11:1043. https://doi.org/10.1038/s41598-020-80221-4.
[225] Arjovsky M, Bottou L, Gulrajani I, Lopez-Paz D. Invariant risk minimization. https://doi.org/10.48550/arXiv.1907.02893. arXiv:1907.02893v3 [stat.ML], 2020.
[226] Choe YJ, Ham J, Park K. An empirical study of invariant risk minimization. pp. 1–8. https://doi.org/10.48550/ARXIV.2004.05007. arXiv:2004.05007v2 [stat.ML], 2020.
[227] Kamath P, Tangella A, Sutherland D, Srebro N. Does invariant risk minimization capture invariance? In: Banerjee A, Fukumizu K, editors. Proceedings of the 24th international conference on artificial intelligence and statistics. Proceedings of machine learning research, vol. 130. PMLR; 2021. p. 4069–77.
[228] Rosenfeld E, Ravikumar PK, Risteski A. The risks of invariant risk minimization. In: International conference on learning representations; 2021.
[229] Bica I, Jarrett D, van der Schaar M. Invariant causal imitation learning for generalizable policies. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021. p. 3952–64.
[230] Sonar A, Pacelli V, Majumdar A. Invariant policy optimization: towards stronger generalization in reinforcement learning. In: Jadbabaie A, Lygeros J, Pappas GJ, Parrilo PA, Recht B, Tomlin CJ, et al., editors. Proceedings of the 3rd conference on learning for dynamics and control. Proceedings of machine learning research, vol. 144. PMLR; 2021. p. 21–33.
[231] Stojanov P, Li Z, Gong M, Cai R, Carbonell J, Zhang K. Domain adaptation with invariant representation learning: what transformations to learn? In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021. p. 24791–803.
[232] Zhang A, McAllister RT, Calandra R, Gal Y, Levine S. Learning invariant representations for reinforcement learning without reconstruction. In: International conference on learning representations; 2021.
[233] Lu C, Wu Y, Hernández-Lobato JM, Schölkopf B. Invariant causal representation learning for out-of-distribution generalization. In: International conference on learning representations; 2022. p. 1–32.
[234] Lu Y, Meisami A, Tewari A. Efficient reinforcement learning with prior causal knowledge. In: Proceedings of machine learning research, vol. 140. 2022. p. 1–27.
[235] Li L, Walsh TJ, Littman ML. Towards a unified theory of state abstraction for MDPs. In: International symposium on artificial intelligence and mathematics, AI&Math 2006; 2006. p. 4–6.
[236] Shalizi CR, Crutchfield JP. Computational mechanics: pattern and prediction, structure and simplicity. J Stat Phys 2001;104:817–79. https://doi.org/10.1023/A:1010388907793.
[237] Thorpe WH. Learning and instinct in animals, learning and instinct in animals. Cambridge, MA, US: Harvard University Press; 1956.
[238] Kounios J, Beeman M. The cognitive neuroscience of insight. Annu Rev Psychol 2014;65:71–93. https://doi.org/10.1146/annurev-psych-010213-115154.
[239] Shupe E. The irreconcilability of insight. Anim Cogn 2024;27:16. https://doi.org/10.1007/s10071-024-01844-y.
[240] Lind J, Ghirlanda S, Enquist M. Insight learning or shaping? Proc Natl Acad Sci 2009;106:E76. https://doi.org/10.1073/pnas.0906120106.
[241] Call J, Carpenter M, Tomasello M. Copying results and copying actions in the process of social learning: chimpanzees (Pan troglodytes) and human children (Homo sapiens). Anim Cogn 2005. https://doi.org/10.1007/s10071-004-0237-8.
[242] Horner V, Whiten A. Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). Anim Cogn 2005. https://doi.org/10.1007/s10071-004-0239-6.
[243] Tennie C, Call J, Tomasello M. Push or pull: imitation vs emulation in great apes and human children. Ethology 2006. https://doi.org/10.1111/j.1439-0310.2006.01269.x.
[244] Ross S, Bagnell D. Efficient reductions for imitation learning. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR workshop and conference proceedings; 2010. p. 661–8.
[245] Ho J, Ermon S. Generative adversarial imitation learning. In: Proceedings of the 30th international conference on neural information processing systems. Curran Associates Inc.; 2016. p. 4572–80. https://doi.org/10.5555/3157382.3157608.
[246] Lu C, Hernández-Lobato JM, Schölkopf B. Invariant causal representation learning for generalization in imitation and reinforcement learning. In: Workshop on the elements of reasoning: objects, structure and causality; 2022. p. 1–27.
[247] Li Y, Song J, Ermon S. InfoGAIL: interpretable imitation learning from visual demonstrations. In: Advances in neural information processing systems 2017-December; 2017. p. 3813–23.
[248] Kumar A, Zhou A, Tucker G, Levine S. Conservative Q-learning for offline reinforcement learning. Advances in neural information processing systems, vol. 33. Curran Associates, Inc.; 2020. p. 1179–91.
[249] Ghosh D, Ajay A, Agrawal P, Levine S. Offline RL policies should be trained to be adaptive. In: Proceedings of the 39th international conference on machine learning. PMLR; 2022. p. 7513–30.
[250] Ziebart BD, Maas A, Bagnell JA, Dey AK. Maximum entropy inverse reinforcement learning. In: Proceedings of the 23rd national conference on artificial intelligence - volume 3, AAAI'08. Chicago, Illinois: AAAI Press; 2008. p. 1433–8.
[251] Wulfmeier M, Ondruska P, Posner I. Maximum entropy deep inverse reinforcement learning. https://doi.org/10.48550/arXiv.1507.04888. arXiv:1507.04888v3 [cs.LG], 2016.
[252] Abbeel P, Ng AY. Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the twenty-first international conference on machine learning, ICML '04. New York, NY, USA: Association for Computing Machinery; 2004. p. 1.
[253] Fu J, Luo K, Levine S. Learning robust rewards with adversarial inverse reinforcement learning. https://doi.org/10.48550/arXiv.1710.11248. arXiv:1710.11248v2 [cs.LG], 2018.
[254] Hopper LM, Lambeth SP, Schapiro SJ, Whiten A. Observational learning in chimpanzees and children studied through 'ghost' conditions. Proc R Soc Lond B, Biol Sci 2008. https://doi.org/10.1098/rspb.2007.1542.
[255] Hopper LM. 'Ghost' experiments and the dissection of social learning in humans and animals. Biol Rev 2010. https://doi.org/10.1111/j.1469-185x.2010.00120.x.
[256] Igl M, Zintgraf L, Le TA, Wood F, Whiteson S. 2018. Deep variational reinforcement learning for POMDPs.
[257] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. pp. 1–12. https://doi.org/10.48550/arXiv.1707.06347. arXiv:1707.06347v2 [cs.LG], 2017.
[258] Haber N, Mrowca D, Wang S, Fei-Fei LF, Yamins DL. Learning to play with intrinsically-motivated, self-aware agents. Advances in neural information processing systems, vol. 31. Curran Associates, Inc.; 2018.
[259] Andrychowicz M, Wolski F, Ray A, Schneider J, Fong R, Welinder P, et al. Hindsight experience replay. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in neural information processing systems, vol. 30. Curran Associates, Inc.; 2017.
[260] Hussein A, Gaber MM, Elyan E, Jayne C. Imitation learning: a survey of learning methods. ACM Comput Surv 2017;50:21. https://doi.org/10.1145/3054912.
[261] Bannon J, Windsor B, Song W, Li T. Causality and batch reinforcement learning: complementary approaches to planning in unknown domains. https://doi.org/10.48550/arXiv.2006.02579. arXiv:2006.02579v1.

[262] Tomasello M. Cultural transmission in the tool use and communicatory signaling of chimpanzees? In: Gibson KR, Parker ST, editors. 'Language' and intelligence in monkeys and apes: comparative developmental perspectives. Cambridge: Cambridge University Press; 1990. p. 274–311.
[263] Tomasello M. Emulation learning and cultural learning. Behav Brain Sci 1998;21:703–4. https://doi.org/10.1017/S0140525X98441748.
[264] Whiten A, McGuigan N, Marshall-Pescini S, Hopper LM. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. Philos Trans R Soc Lond B, Biol Sci 2009;364:2417–28. https://doi.org/10.1098/rstb.2009.0069.
[265] Zentall TR. Mechanisms of copying, social learning, and imitation in animals. Psychol Learn Motiv 2022;80:101844. https://doi.org/10.1016/j.lmot.2022.101844.
[266] Waldmann MR, Cheng PW, Hagmayer Y, Blaisdell AP. Causal learning in rats and humans: a minimal rational model. In: Chater N, Oaksford M, editors. The probabilistic mind: prospects for Bayesian cognitive science. Oxford, UK: Oxford University Press; 2008. p. 453–84.
[267] Higgins I, Amos D, Pfau D, Racaniere S, Matthey L, Rezende D, et al. Towards a definition of disentangled representations. https://doi.org/10.48550/ARXIV.1812.02230. arXiv:1812.02230v1 [cs.LG], 2018.
[268] Voudouris K, Alhas I, Schellaert W, Crosby M, Holmes J, Burden J, et al. Animal-AI 3: what's new & why you should care. https://doi.org/10.48550/arXiv.2312.11414. arXiv:2312.11414, 2023.
[269] Crosby M. Building thinking machines by solving animal cognition tasks. Minds Mach 2020. https://doi.org/10.1007/s11023-020-09535-6.
[270] Wang JX, Kurth-Nelson Z, Tirumala D, Soyer H, Leibo JZ, Munos R, et al. Learning to reinforcement learn. pp. 1–17. https://doi.org/10.48550/arXiv.1611.05763. arXiv:1611.05763v3 [cs.LG], 2016.
[271] Duan Y, Schulman J, Chen X, Bartlett PL, Sutskever I, Abbeel P. RL^2: Fast reinforcement learning via slow reinforcement learning. pp. 1–14. https://doi.org/10.48550/arXiv.1611.02779. arXiv:1611.02779v2 [cs.AI], 2016.
[272] Nagabandi A, Clavera I, Liu S, Fearing RS, Abbeel P, Levine S, et al. Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. In: International conference on learning representations; 2019.
[273] Dasgupta I, Wang J, Chiappa S, Mitrovic J, Ortega P, Raposo D, et al. Causal reasoning from meta-reinforcement learning. https://doi.org/10.48550/arXiv.1901.08162. arXiv:1901.08162, 2019.
[274] Kirk R, Zhang A, Grefenstette E, Rocktäschel T. A survey of zero-shot generalisation in deep reinforcement learning. J Artif Intell Res 2023;76. https://doi.org/10.1613/jair.1.14174.
[275] Touati A, Rapin J, Ollivier Y. Does zero-shot reinforcement learning exist? In: The eleventh international conference on learning representations; 2023.
[276] Khetarpal K, Riemer M, Rish I, Precup D. Towards continual reinforcement learning: a review and perspectives. J Artif Intell Res 2022;75:1401–76. https://doi.org/10.1613/jair.1.13673.
[277] Abel D, Barreto A, Van Roy B, Precup D, van Hasselt HP, Singh S. A definition of continual reinforcement learning. In: Advances in neural information processing systems 36 (NeurIPS 2023); 2023. p. 50377–407.
[278] Kemp C, Perfors A, Tenenbaum JB. Learning overhypotheses with hierarchical Bayesian models. Dev Sci 2007;10:307–21. https://doi.org/10.1111/j.1467-7687.2007.00585.x.
[279] Kemp C, Goodman ND, Tenenbaum JB. Learning to learn causal models. Cogn Sci 2010;34:1185–243. https://doi.org/10.1111/j.1551-6709.2010.01128.x.
[280] Lucas CG, Bridgers S, Griffiths TL, Gopnik A. When children are better (or at least more open-minded) learners than adults: developmental differences in learning the forms of causal relationships. Cognition 2014;131:284–99. https://doi.org/10.1016/j.cognition.2013.12.010.
[281] Dasgupta I, Schulz E, Tenenbaum JB, Gershman SJ. A theory of learning to infer. Psychol Rev 2020;127:412–41. https://doi.org/10.1037/rev0000178.
[282] Kosoy E, Liu A, Collins JL, Chan D, Hamrick JB, Ke NR, et al. Learning causal overhypotheses through exploration in children and computational models. In: First conference on causal learning and reasoning; 2022. p. 1–17.
[283] Jiang C, Lucas CG. Actively learning to learn causal relationships. Comput Brain Behav 2024;7:80–105. https://doi.org/10.1007/s42113-023-00195-0.
[284] Nagabandi A, Finn C, Levine S. Deep online learning via meta-learning: continual adaptation for model-based RL. In: International conference on learning representations; 2019. https://openreview.net/forum?id=HyxAfnA5tm.
[285] Lee S, Ha J, Zhang D, Kim G. A neural Dirichlet process mixture model for task-free continual learning. In: International conference on learning representations; 2020. https://openreview.net/forum?id=SJxSOJStPr.
[286] Mendez JA, Eaton E. Lifelong learning of compositional structures. In: International conference on learning representations; 2021. https://openreview.net/forum?id=ADWd4TJO13G.
[287] Schmidhuber J, Zhao J, Wiering MA. Simple principles of meta-learning. Technical Report 69-96. Lugano, Switzerland: IDSIA; 1996.
[288] Thrun S. Is learning the n-th thing any easier than learning the first? In: Touretzky D, Mozer MC, Hasselmo M, editors. Advances in neural information processing systems, vol. 8. MIT Press; 1996.
[289] Caruana R. Multitask learning. Mach Learn 1997;28:41–75. https://doi.org/10.1023/A:1007379606734.
[290] Yu T, Quillen D, He Z, Julian R, Hausman K, Finn C, et al. Meta-world: a benchmark and evaluation for multi-task and meta reinforcement learning. In: Kaelbling LP, Kragic D, Sugiura K, editors. Proceedings of the conference on robot learning. Proceedings of machine learning research, PMLR, vol. 100. 2020. p. 1094–100.
[291] Geisa A, Mehta R, Helm HS, Dey J, Eaton E, Dick J, et al. Towards a theory of out-of-distribution learning. https://doi.org/10.48550/ARXIV.2109.14501. arXiv:2109.14501v4, 2022.
[292] Ahuja K, Caballero E, Zhang D, Gagnon-Audet J-C, Bengio Y, Mitliagkas I, et al. Invariance principle meets information bottleneck for out-of-distribution generalization. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021. p. 3438–50.
[293] Wenzel F, Dittadi A, Gehler PV, Simon-Gabriel C-J, Horn M, Zietlow D, et al. Assaying out-of-distribution generalization in transfer learning. https://doi.org/10.48550/ARXIV.2207.09239. arXiv:2207.09239v1 [cs.LG], 2022.
[294] Dittadi A, Träuble F, Locatello F, Wuthrich M, Agrawal V, Winther O, et al. On the transfer of disentangled representations in realistic settings. In: International conference on learning representations; 2021.
[295] Träuble F, Dittadi A, Wuthrich M, Widmaier F, Gehler PV, Winther O, et al. The role of pretrained representations for the OOD generalization of RL agents. In: International conference on learning representations; 2022. p. 1–20.
[296] Ke NR, Didolkar A, Mittal S, Goyal A, Lajoie G, Bauer S, et al. Systematic evaluation of causal discovery in visual model based reinforcement learning. pp. 1–38. arXiv:2107.00848v1 [stat.ML], 2021.
[297] Shanahan M, Crosby M, Beyret B, Cheke L. Artificial intelligence and the common sense of animals. Trends Cogn Sci 2020;24:862–72. https://doi.org/10.1016/j.tics.2020.09.002.
[298] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484–9. https://doi.org/10.1038/nature16961.
[299] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of go without human knowledge. Nature 2017;550:354–9. https://doi.org/10.1038/nature24270.
[300] Vinyals O, Babuschkin I, Czarnecki WM, Mathieu M, Dudzik A, Chung J, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature 2019;575:350–4. https://doi.org/10.1038/s41586-019-1724-z.
[301] Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. A brief survey of deep reinforcement learning. IEEE Signal Process Mag 2017;34:26–38. https://doi.org/10.1109/msp.2017.2743240.

[302] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems, vol. 30. Curran Associates, Inc.; 2017. p. 1–15.
[303] Open-X Embodiment Collaboration, O'Neill A, Rehman A, Maddukuri A, Gupta A, Padalkar A, Lee A, et al. Open X-embodiment: robotic learning datasets and RT-x models. https://doi.org/10.48550/arXiv.2310.08864. arXiv:2310.08864v7 [cs.RO], 2024.
[304] Wang G, Xie Y, Jiang Y, Mandlekar A, Xiao C, Zhu Y, et al. Voyager: an open-ended embodied agent with large language models. https://doi.org/10.48550/arXiv.2305.16291. arXiv:2305.16291v2 [cs.AI], 2023.
[305] Gupta A, Fan L, Ganguli S, Fei-Fei L. MetaMorph: learning universal controllers with transformers. In: International conference on learning representations; 2022.
[306] Fan L, Wang G, Jiang Y, Mandlekar A, Yang Y, Zhu H, et al. MineDojo: building open-ended embodied agents with Internet-scale knowledge. In: Thirty-sixth conference on neural information processing systems datasets and benchmarks track; 2022.
[307] Ghosh D, Walke H, Pertsch K, Black K, Mees O, Dasari S, et al. Octo: an open-source generalist robot policy. https://doi.org/10.48550/arXiv.2405.12213. arXiv:2405.12213v2 [cs.RO], 2024.
[308] Hagmayer Y, Sloman SA. Decision makers conceive of their choices as interventions. J Exp Psychol Gen 2009;138:22–38. https://doi.org/10.1037/a0014585.
[309] Coenen A, Rehder B, Gureckis TM. Strategies to intervene on causal systems are adaptively selected. Cogn Psychol 2015;79:102–33. https://doi.org/10.1016/j.cogpsych.2015.02.004.
[310] Coenen A, Bramley N, Ruggeri A, Gureckis T. Beliefs about sparsity affect causal experimentation. In: CogSci 2017; 2017. p. 1788–93.
[311] McCormack T, Bramley N, Frosch C, Patrick F, Lagnado D. Children's use of interventions to learn causal structure. J Exp Child Psychol 2016;141:1–22. https://doi.org/10.1016/j.jecp.2015.06.017.
[312] Bramley NR, Gerstenberg T, Tenenbaum JB, Gureckis TM. Intuitive experimentation in the physical world. Cogn Psychol 2018;105:9–38. https://doi.org/10.1016/j.cogpsych.2018.05.001.
[313] Bramley NR, Ruggeri A. Children's active physical learning is as effective and goal-targeted as adults'. Dev Psychol 2022;58:2310–21. https://doi.org/10.1037/dev0001435.
[314] Pan H-R, Gürtler N, Neitz A, Schölkopf B. Direct advantage estimation. In: Advances in neural information processing systems 35 (NeurIPS 2022); 2022. p. 11869–80.
[315] Pan H-R, Schölkopf B. Skill or luck? Return decomposition via advantage functions. In: Proceedings of the twelfth international conference on learning representations (ICLR); 2024.
[316] Wildberger JB, Guo S, Bhattacharyya A, Schölkopf B. On the interventional Kullback-Leibler divergence. In: van der Schaar M, Zhang C, Janzing D, editors. Proceedings of the second conference on causal learning and reasoning. Proceedings of machine learning research, PMLR, vol. 213. 2023. p. 328–49.
[317] Cheng PW, Novick LR. A probabilistic contrast model of causal induction. J Pers Soc Psychol 1990;58:545–67. https://doi.org/10.1037/0022-3514.58.4.545.
[318] Cheng PW, Novick LR. Covariation in natural causal induction. Psychol Rev 1992;99:365–82. https://doi.org/10.1037/0033-295X.99.2.365.
[319] Glymour C. The mind's arrows: Bayes nets and graphical causal models in psychology. The MIT Press; 2001.
[320] Holyoak KJ, Cheng PW. Causal learning and inference as a rational process: the new synthesis. Annu Rev Psychol 2011;62:135–63. https://doi.org/10.1146/annurev.psych.121208.131634.
[321] Rottman BM, Hastie R. Reasoning about causal relationships: inferences on causal networks. Psychol Bull 2014;140:109–39. https://doi.org/10.1037/a0031903.
[322] Hagmayer Y. Causal Bayes nets as psychological theories of causal reasoning: evidence from psychological research. Synthese 2016;193:1107–26. https://doi.org/10.1007/s11229-015-0734-0.
[323] Rottman BM. The acquisition and use of causal structure knowledge. In: Waldmann MR, editor. The Oxford handbook of causal reasoning. Oxford, UK: Oxford University Press; 2017. p. 86–114.
[324] Glymour C. Learning, prediction and causal Bayes nets. Trends Cogn Sci 2003;7:43–8. https://doi.org/10.1016/S1364-6613(02)00009-8.
[325] Danks D. Unifying the mind: cognitive representations as graphical models. Cambridge, Massachusetts: The MIT Press; 2014.
[326] Steyvers M. Inferring causal networks from observations and interventions. Cogn Sci 2003;27:453–89. https://doi.org/10.1016/S0364-0213(03)00010-7.
[327] Bramley NR, Dayan P, Griffiths TL, Lagnado DA. Formalizing Neurath's ship: approximate algorithms for online causal learning. Psychol Rev 2017;124:301–38. https://doi.org/10.1037/rev0000061.
[328] Bramley N, Mayrhofer R, Gerstenberg T, Lagnado DA. Causal learning from interventions and dynamics in continuous time. In: CogSci 2017; 2017. p. 150–5.
[329] Davis ZJ, Bramley NR, Rehder B. Causal structure learning in continuous systems. Front Psychol 2020;11:1–16. https://doi.org/10.3389/fpsyg.2020.00244.
[330] Gong T, Gerstenberg T, Mayrhofer R, Bramley NR. Active causal structure learning in continuous time. Cogn Psychol 2023;140:101542. https://doi.org/10.1016/j.cogpsych.2022.101542.
[331] Rothe A, Deverett B, Mayrhofer R, Kemp C. Successful structure learning from observational data. Cognition 2018;179:266–97. https://doi.org/10.1016/j.cognition.2018.06.003.
[332] Valentin S, Bramley NR, Lucas CG. Discovering common hidden causes in sequences of events. Comput Brain Behav 2023;6:377–99. https://doi.org/10.1007/s42113-022-00156-z.
[333] Weisberg DS, Gopnik A. Pretense, counterfactuals, and Bayesian causal models: why what is not real really matters. Cogn Sci 2013;37:1368–81. https://doi.org/10.1111/cogs.12069.
[334] Gerstenberg T, Goodman ND, Lagnado DA, Tenenbaum JB. From counterfactual simulation to causal judgment. In: Proceedings of the 36th annual conference of the cognitive science society. Austin, TX: Cognitive Science Society; 2014. p. 523–8.
[335] Gerstenberg T, Goodman ND, Lagnado DA, Tenenbaum JB. A counterfactual simulation model of causal judgments for physical events. Psychol Rev 2021;128:936–75. https://doi.org/10.1037/rev0000281.
[336] Gerstenberg T. What would have happened? Counterfactuals, hypotheticals and causal judgements. Philos Trans R Soc Lond B, Biol Sci 2022;377:20210339. https://doi.org/10.1098/rstb.2021.0339.
[337] Quillien T, Lucas CG. Counterfactuals and the logic of causal selection. Psychol Rev 2023. https://doi.org/10.1037/rev0000428.
[338] Sontakke SA, Iota S, Hu Z, Mehrjou A, Itti L, Schölkopf B. GalilAI: out-of-task distribution detection using causal active experimentation for safe transfer RL. In: Proceedings of the 25th international conference on artificial intelligence and statistics. PMLR; 2022. p. 7518–30.
[339] Chater N, Oaksford M. Programs as causal models: speculations on mental programs and mental representation. Cogn Sci 2013;37:1171–91. https://doi.org/10.1111/cogs.12062.
[340] Bramley NR, Zhao B, Quillien T, Lucas CG. Local search and the evolution of world models. Top Cogn Sci 2023. https://doi.org/10.1111/tops.12703.
[341] Piantadosi ST. The computational origin of representation. Minds Mach 2021;31:1–58. https://doi.org/10.1007/s11023-020-09540-9.
[342] Rule JS, Tenenbaum JB, Piantadosi ST. The child as hacker. Trends Cogn Sci 2020;24:900–15. https://doi.org/10.1016/j.tics.2020.07.005.
[343] Fodor JA. The language of thought. Harvard University Press; 1975.
[344] Botvinick M, Barrett DGT, Battaglia P, de Freitas N, Kumaran D, Leibo JZ, et al. Building machines that learn and think for themselves. Behav Brain Sci 2017;40:e255. https://doi.org/10.1017/S0140525X17000048.
[345] Burgess CP, Matthey L, Watters N, Kabra R, Higgins I, Botvinick M, et al. Unsupervised scene decomposition and representation. pp. 1–22. https://doi.org/10.48550/ARXIV.1901.11390. arXiv:1901.11390v1 [cs.CV], 2019.