



# Towards ending the animal cognition war: a three-dimensional model of causal cognition

Tobias Benjamin Starzak<sup>1</sup> · Russell David Gray<sup>2,3,4</sup>

Received: 2 October 2019 / Accepted: 20 January 2021  
© The Author(s) 2021

## Abstract

Debates in animal cognition are frequently polarized between the romantic view that some species have human-like causal understanding and the killjoy view that human causal reasoning is unique. These apparently endless debates are often characterized by conceptual confusions and accusations of straw-men positions. What is needed is an account of causal understanding that enables researchers to investigate both similarities and differences in cognitive abilities in an incremental evolutionary framework. Here we outline the ways in which a three-dimensional model of causal understanding fulfills these criteria. We describe how this approach clarifies what is at stake, illuminates recent experiments on both physical and social cognition, and plots a path for productive future research that avoids the romantic/killjoy dichotomy.

**Keywords** Comparative psychology · Causal understanding · Cognitive evolution · Animal cognition · Concept of understanding

## Introduction

Academic debates are a little like popular music. Some songs are rich, complex and infinitely rewarding. Others are dull and quickly forgotten. And a few have a strange ABBA-like “annoying-but-addictive” quality that makes them impossible to get out of your head. Debates in animal cognition often fall into this latter

---

✉ Tobias Benjamin Starzak  
tobias.starzak@rub.de

<sup>1</sup> Institut Für Philosophie II, Fakultät Für Philosophie und Erziehungswissenschaft, Ruhr-Universität Bochum, 44780 Bochum, Germany

<sup>2</sup> Max Planck Institute for the Science of Human History, 07743 Jena, Germany

<sup>3</sup> School of Psychology, University of Auckland, Auckland 1142, New Zealand

<sup>4</sup> Research School of the Social Sciences, Australian National University, Canberra 2601, Australia

category. Over and over again the familiar refrain is, “do animals have complex human-like cognitive abilities or can their behavior be explained in terms of simpler processes such as associative learning?” Regardless of whether the topic under discussion is language, culture, theory of mind, episodic memory, insight, physical cognition, numerical cognition, or mental time travel, the structure of the debate remains the same. On one side of the battle researchers, whom Dennett (1987) has dubbed “romantics”, joyfully trumpet that their animal has remarkably complex cognitive abilities. On the other side researchers whom Dennett (1987) labels “killjoys”, coolly reply that the apparently complex behavior can be underpinned by simpler cognitive mechanisms. The recent debate about “flexible planning” in ravens provides a good example of this. In an article in *Science*, Kabadayi and Osvath (2017) claim that, just like humans and other great apes, the ravens in their experiments showed flexible planning in tool selection and token exchange tasks. They claimed that the ravens performance was better than that of four year old children. However, their “killjoy” critics (Redshaw et al. 2017; Hampton 2019) quickly pointed out that, as the stones and bottle tops had previously been associated with rewards, their selection and exchange could also be explained by associative conditioning. A similar dynamic has played out in debates about “insightful” problem-solving in animals. While some claim that the spontaneous retrieval of a food item on a string shows that some birds have insight (e.g. Heinrich 2000), others have pointed out that success in the task requires the moment to moment perceptual-motor reinforcement of the food getting closer with each pull of the string (Taylor et al. 2012). When the birds are deprived of the feedback (for example with coiled string on a flat platform), they no longer succeed in retrieving the food. Lest we strain our readers’ patience too far, one more example should suffice to make the general point about the ubiquity of the romantic/killjoy dynamic in disputes about animal cognition. In an article published in *Nature*, Gentner et al. (2006) claimed that European starlings could recognize and classify acoustic sequences from recursive, centre-embedded grammars—a remarkable ability that had previously been thought to be the cornerstone of our unique human linguistic abilities. In response, Corballis (2007) pointed out that learning a simple counting rule seemed much more likely, especially since the starlings were given thousands of trials.

From an evolutionary point of view neither the romantic nor the killjoy position is satisfactory. As evolutionary scholars from Darwin to Dawkins have emphasized, adaptations are built up in a stepwise incremental fashion (Dawkins 1996). Both the romantic position with its focus on comparing a single species with humans, and the killjoy position leave completely unexplained how the evolutionary transition between standard cognitive mechanisms and the complexities of human, or “human-like”, cognition might occur. In this paper, we want to make a step towards ending the polarized debates on the topic of causal understanding. We will discuss the conceptual and practical problems the dichotomy between “simple” associative learning and “complex” causal understanding causes, and describe the ways in which a three-dimensional model of causal understanding cannot only provide greater conceptual clarity but also guide future empirical research.

## Dissecting disagreement

Animal cognition researchers disagree wildly about the question of animal causal understanding. While some researchers argue that some animals understand aspects of causality (e.g. Visalberghi and Tomasello 1998; Seed et al. 2006; Tebbich et al. 2007; Seed and Call 2009; Taylor et al. 2007, 2009a, b), others insist on the lack of evidence that animals possess anything remotely resembling full-fledged human-like causal understanding (e.g. Penn and Povinelli 2008, 2009, 2011). Why is it that researchers, who share access to the same body of experimental results and observational data develop into schools with such strong and stable disagreement? Is the data inconclusive? And if so, can the problem be solved if we come up with better experiments? In this section, we will briefly discuss the main factors that contribute to the deadlock. *First*, different researchers evaluate the available evidence in different ways, based on different principles of interpretation. *Second*, due to misconstructions of the respective positions, disagreement is sometimes overstated. *Third*, there is conceptual disagreement concerning the notion of causal understanding. In our view, this is the main reason for disagreement in this debate. Pointing to conceptual disagreement relativizes the extent of actual disagreement and instead raises a new question: rather than asking whether an animal fulfills the criteria for causal understanding, we should shift our attention to the question of how to *conceptualize causal understanding*. And rather than treating animal research in this domain as being relevant only for the question which organisms instantiate causal understanding, we should approach animals as informative concerning the question how to best think about causal understanding.

### Principles of interpretation

When different researchers evaluate the same body of evidence in very different ways, a natural thought is that they rely on different principles of interpretation. A principle of interpretation that has received a lot of attention in comparative cognition research is the idea that we should always prefer explanations in terms of less sophisticated cognitive abilities. This intuitive principle can be traced back to *Morgan's Canon*, (Morgan 1903), and it lies at the bottom of Dennett's claim that "behaviorism is the null hypothesis against which all cognitive accounts are [to be] tested" (Dennett 1983). It is obvious how different stances towards this principle could produce disagreement. Causal understanding is commonly understood as a more sophisticated (or, in Morgan's terms, a *higher*) psychological ability than associative reasoning. Hence, Morgan's Canon seems to support explanations in terms of associative processes over explanations in terms of causal cognition. It is therefore not surprising that killjoy researchers tend to endorse the principle as a special version of more general and widely accepted principles, like the law of parsimony, simplicity, or economy (Penn and Povinelli 2007; Karin-D'Arcy 2005; Shettleworth 2013). Romantics, on the other hand, apparently feel only loosely committed to the principle, since they often agree that simpler explanations in terms of

associative processes are possible, but deny that they are preferable. One possibility to support the romantic position is to deny that Morgan's Canon is a justified principle. But despite repeated criticism from the philosophy of science (Sober 1998; 2015; Fitzpatrick 2008; Meketa 2014; Starzak 2017), the principle still enjoys a lot of popularity in comparative psychology, including the romantics' camp. However, even if one does not criticize the principle per se, a plausible specification of the principle is that it applies only in those cases where *evidentially and explanatorily all things are equal*. It is widely agreed that every observed behavior is *compatible* with explanations in terms of associative reasoning. In many cases, however, other explanations seem possible as well, which leads to a problem of *underdetermination*, because the existing behavioral data appears to be "equally well confirmed by multiple incompatible cognitive mechanisms" (Mikhalevich et al. 2017, p. 5). For some cognitive abilities it has been suggested that underdetermination does not only reflect our current lack of knowledge, but constitutes a *logical problem*, i.e. one that cannot be resolved experimentally (Penn and Povinelli 2007).<sup>1</sup> It is not obvious, however, why causal cognition should face a similar *logical problem*.<sup>2</sup> Nevertheless, one might argue that this is a case where evidentially and explanatorily all things are equal, such that Morgan's Canon should be applied. Others however, argue that just because explanations in terms of associative reasoning cannot be ruled out this does not imply that they are always equally supported by the evidence, or that they score better on explanatory metrics (Fitzpatrick 2008). This corresponds to one line of criticism against deflationary accounts of physical-problem solving abilities in terms of associative learning: It is agreed upon that associative hypotheses can be constructed *post-hoc* for every experimental outcome. This general applicability implies that associative learning as the mechanism driving physical problem-solving abilities in animals does not produce clear behavioral predictions that can be falsified, which makes the hypothesis explanatorily weak (Hanus 2016). In every case, to the extent that disagreement is due to different evaluative principles (or their range of application), new experiments are unlikely to solve disagreement and researchers should engage more in theoretical debates concerning principles of evaluation.

## A big misunderstanding and the conceptual question

Romantics take the fact that simple associative learning cannot account very well for animal behavior as observed in many experiments as evidence in favor of causal understanding. In its strongest form, this amounts to a *mutual exclusion error*, i.e. the inference that if simple associative learning can be ruled out, the only other

---

<sup>1</sup> The logical problem refers, roughly, to the problem of finding evidence that supports a theory of mind (ToM) rather than behavior-reading in non-human animals. It is based on the assumption that for every mind-reading (or Theory of mind) hypothesis as explanation for animal behavior there is a *complementary* behavior reading hypothesis that is equally well supported by the evidence. For discussion see Lurz (2011) and Halina (2015).

<sup>2</sup> This should become clear in Sect. 3.1 when we introduce different kinds of causal information. While on some understandings, causal cognition involves the representation of unobservables (which would be analogous to having a ToM) other accounts can do without that (difference-making accounts).

possibility is that the animal has used human-like cognition. Moreover, this error leads them to interpret killjoy-positions as claiming that the best explanation for the behavior in question *is* simple associative learning. According to the traditional associationist view “associations between components are formed automatically, without any form of intentionality being involved” (Hanus 2016, p. 241). It has its roots in a behavioristic, non-mentalistic psychology, thus explanations in terms of associative processes seem to stand in stark contrast to explanations in terms of cognitive processes. It may be because of this understanding of associationism that researchers like Povinelli and Penn have been credited with views according to which chimpanzees are “inflexible” (Hare et al. 2006), or “arbitrary cue learners” (Seed and Call 2009), nothing more than “behavioral rule learners” (Tomasello and Call 2006; Call and Tomasello 2008), or that they only invoke “behavioristic principles of learning” (Tomasello and Call 2006). However, while the idea that explanations in terms of associative learning and explanations in terms of cognitive states are mutually exclusive is still very much alive, a growing number of theorists argues that they are not incompatible (Blaisdell 2008; Buckner 2011), and more recent theories of associative learning explicitly exhibit a hybrid character that combines both cognitivist aspects with components of instrumental learning theories (Heyes 2012; Hanus 2016). It is clearly in the context of these hybrid-theories where killjoys like Povinelli and Penn locate themselves (see Povinelli and Penn 2011), thus the inference from their view being associationist to it being non-cognitivist is mistaken, and as a consequence overstates the existing disagreement.

But distortions go both ways, and while romantics may set the competence bar for causal understanding too low, killjoys exhibit the opposite tendency to “tie the competence criteria for cognitive capacities to an exaggerated sense of typical human performance” (Buckner 2013). For instance, Penn and Povinelli’s concept of causal understanding as the capacity for *second order relational reasoning* is arguably a demanding one that not everyone shares. As a consequence, it is a mistake to interpret every claim about an animal’s causal understanding as a claim about its ability for second-order relational reasoning. When Penn and Povinelli criticize their romantic opponents’ claims that some animals *do* understand aspects of causality boldly as *irrational* (Penn and Povinelli 2009) or as *alchemy* (Penn and Povinelli 2011), they seem to do so under the assumption that they all share the same concept. Since researchers from both camps probably agree a good deal on which problems animals like rats, chimpanzees, or New Caledonian crows can and cannot solve, an important part of the disagreement has its roots in the conceptual question which processes merit the label *causal understanding*. This shows that putting more effort into understanding what exactly the respective claims concerning causal understanding entail is helpful to shed light on the real extent of disagreement. But more importantly, it also puts a spotlight on the normative conceptual question: How *should* we think about causal understanding? In the rest of this paper we will develop an approach to this question.

## The conceptual space of causal cognition

How should we think about causal understanding? To start with, we should break up the complex notion into its parts: what needs to be understood, and what does it mean to understand it. The approach we suggest combines ideas from James Woodward and Kim Sterelny. In a nutshell, according to Woodward (2011, p. 18), human causal cognition is not a single ability but consists of a bundle of distinct abilities. In normal adult human beings these abilities are relatively well integrated, but they are—at least conceptually—distinguishable. In a more general fashion, Sterelny argues that when we think about an animal’s cognitive system, “we need to think about the channels through which information flows to its mind, and about the flexibility with which it can use crucial information” (Sterelny 2003, p. 34). Combining the two, we’ll argue that in thinking about the nature of causal understanding we should think about the extent to which organisms can differ with respect to the kind of information they can pick up; with respect to the different sources of causal information they can exploit; with respect to the way they can process this information and integrate different types of information or information stemming from different sources; and with respect to the flexibility with which they can use this information to guide their behavior.

Approaching the question of animal causal understanding from this angle has various advantages. First of all, shifting the debate from *causal understanding* to a more general, gradualist and less normative notion of *causal cognition* somewhat brackets normative issues.<sup>3</sup> To the extent that conceptual disagreement concerning *understanding* does play a key role in the debate, dissecting the object of investigation into empirically tractable parts (the sum of which we refer to as causal cognition) promises to be acceptable for romantics and killjoys alike. The idea is to provide a framework in which everything an animal can do with causal information (broadly construed) can be situated, and to postpone the question of how we should label the underlying combination of abilities. Secondly, given that we find both similarities and differences between humans and non-human animals, this approach is neutral concerning special interests of researchers to highlight either the animal in humans, or what makes humans a special case. Rather than blurring or overemphasizing the differences between humans and animals, we will argue that our approach can both point out commonalities and differences in a much more fine-grained manner than any dualistic *either-or* approach. If there are major disparities in the causal cognition of animals this should emerge from our dimensional approach rather than being imposed a priori. This feature makes it a better fit to the overall aims of comparative cognition research. Finally, in grounding subsequent discussions concerning the criteria for *causal understanding* on empirical data, our framework also provides a key to solving the normative conceptual question. With this, it potentially contributes to put an end to the animal cognition war.

<sup>3</sup> These normative issues refer to the question which level of sophistication in the use of causal information is needed to qualify as *real causal understanding*, as discussed in Sect. 2.2.

In Sect. 3.1, we will start by giving a brief account of causal information. Following this we will turn to the notion of understanding and explore the space of conceptual possibilities of how the different parameters of understanding (sources, integration, and explicitness) relate to one another, and how they can dissociate (Sects. 3.2 and 3.3).

## Causal information

Although what causation is metaphysically and how it is represented psychologically are two different questions, they are closely related (Woodward 2011). When we want to know whether animals are sensitive to causal information and what they can do with this information, we need some grasp of what causal relations are and what the cues are by which they can be identified in nature. Concerning causal *understanding* it makes sense to distinguish the representation of relations that are in fact causal from the representation of these relations *as* causal. In the philosophical discussion, different accounts of how to best characterize causal relations can be found, and this difference concerning the metaphysics of causality is also mirrored in different psychological accounts of what it means to represent, reason about, or understand causal relations. This is obviously a possible source of confusion, as it introduces the possibility to evaluate judgments concerning causal understanding (or the lack thereof) according to different standards. In this section, we briefly introduce interventionist and geometrical–mechanical approaches to causal information and argue, following Woodward (2011), that rather than being exclusive alternatives these theories are better viewed as being about different kinds or aspects of causal information. While adult humans typically exploit both kinds of information, young children and non-human animals may be more constrained in this respect.

### ***Difference-making accounts of causality***

David Hume defined “a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, *if the first object had not been, the second never had existed*” (Hume 1995, p.87, our italics). While Hume did not work out the idea of causality as a difference-making relation (he rather understood causal relations in terms of regularity) the quote nicely captures the core idea of difference-making accounts of causality: “what it is for c to be a cause of effect e is for c to be something that if it had not occurred, e would not have occurred” either (Ney 2009, p. 738). The details of the right kind of difference which causes make for their effects have been spelled out in terms of probability (Eells 1991), counterfactuals (Lewis 1973), or more recently as invariance under intervention (Woodward 2003). For instance, according to interventionism, if a manipulation of one thing or event A (while controlling for other relevant factors/possible causes) changes the value of some other thing or event B, A makes the right kind of difference to B and the two are causally related.

Given this approach to the nature of causal relations, causal information—the information to which some agents are sensitive and which they can pick up to track causal relations or in terms of which they represent causal relations—can be thought of as *difference-making* information, i.e. as information about the contingency or co-variation of the right kind between cause and effect. Understanding or representing causal relations, according to this view, is about thinking about causes as “*handles* for manipulating or controlling their effects” (Woodward 2011, p. 25), or as Hohwy (2013, p. 6) puts it, about “being able to imagine what happens when the world is intervened upon in a controlled manner”.

Importantly, we have to distinguish how a subject represents causal relations from how she learns about these relations. First, as Woodward argues, psychological interventionism does not entail that the only way for a subject to learn about causal relationships is to actually perform an intervention oneself. We’ll discuss different sources of causal information in some detail in the following section. Secondly, while a natural thought may be that the only way to learn about difference-making information is via extracting statistical data from patterns of co-variation and contingency, psychological accounts of understanding causal relations along these lines are not committed to that claim. Background information as well as inferential biases may result in forming representations of relations as causal on the basis of a single observation. Moreover, as long as the content of a representation can be spelled out in terms of difference-making and intervention, it can be said to be about causal relations even in cases where a subject is mistaken. Finally, theories about the representation of difference-making information include associative accounts of causal learning and judgment (like Dickinson and Shanks (1995)). Thus, they have a wider scope and are in this respect less demanding concerning the mental requirements for causal learning than theories that contrast *real* causal learning with associative learning right from the get-go (as Penn and Povinelli’s *biased-association hypothesis*, (Penn and Povinelli 2000, 2009, 2011; but also Seed et al. 2011).

### **Geometrical–mechanical accounts**

A different approach towards causality are so-called *geometrical–mechanical* accounts. According to these theories, what is distinctive about a causal relation is that the relata are connected *in the right way*. This right way is spelled out as “the cause being spatiotemporally contiguous with the effect via a spatiotemporally contiguous process that *transfers energy*” (Woodward 2011, p. 24, our italics). Simply put, the idea is that A causes B via the transmission of momentum, spin, mass, or energy during contact, e.g. a moving billiard ball that causes a stationary ball to move after collision. Theories along these lines are intended to capture a wide range of phenomena involving mechanical interactions like pushing and pulling, breaking, support, and mechanical properties like rigidity, weight, or impenetrability.

Geometrical–mechanical accounts of causality seem to be what many comparative researchers have in mind when they investigate causal understanding. Much of the literature on animal causal understanding investigates the ability to exploit geometrical–mechanical cues in the context of tool-use, and to understand properties

like rigidity (Povinelli 2000; Vonk and Povinelli 2011), weight (Hanus and Call 2008, 2011), gravity (Povinelli 2001; Tomonaga et al. 2007), support (Spinozzi and Potì 1989), and mediating forces more generally (Visalberghi and Tomasello 1998).

Theories along these lines offer an explanation for *why* A causes B—i.e. via the transmission of energy/momentum or contact force—while difference-making theories only account for *whether* A causes B. Povinelli and Penn (2011, p. 72, our italics) argue that while some animals like chimpanzees can “grasp the causal nature of their own goal-directed actions, [they do so] only darkly: [they don’t] *understand* causality in a *diagnostic, theory-like manner*.” They never ask “why?”. This puts the bar for what counts as *understanding* causality very high and it is easy to see why many researchers are skeptical that any non-human animals have this ability. To be able to represent and reason about causal relations and interactions like collision, pushing, support or containment not only requires an agent to track geometrical–mechanical relations—they also need to possess the relevant concepts and theories, like core physical principles (Spelke et al. 1995), force transmission (Leslie 1995), or force dynamics (Wolff 2007), and they need to represent unobservable forces and higher-order relations (Penn and Povinelli 2000, 2009, 2011). But setting aside the question of full-blown understanding, it has been argued that exploiting geometrical–mechanical cues can also be less demanding than fine-grained representations of difference-making relationships that enable a subject to perform successful interventions. As Woodward (2011) notes, a chimpanzee that observes how another uses a hammer and an anvil to crack open nuts might understand that *contact* plays a crucial role while not being able to figure out how exactly to perform an intervention to bring about the result. Similarly, in another experiment, chimpanzees had to choose a tool with which to retrieve a food source that was some distance away (Penn and Povinelli 2011). The tools they could choose from differed with respect to their length, rigidity and with respect to whether they had hooks at their ends. The chimpanzee’s choices reflected that they understood that physical contact is of importance, but not how exactly this contact would bring about the intended result. “It is” Woodward (2011, p. 32) says “as though the primates grasp the idea that retrieval of the food requires that there be a causal process connecting their hands to the food (putting the stick in contact with the food constitutes such a process) but don’t get the idea that using the tool in a way that makes a difference for food retrieval requires something more”.<sup>4</sup>

### **Difference-making and geometrical–mechanical aspects of human concept of causation**

Both geometrical–mechanical aspects as well as difference-making aspects seem to be part of folk physics and the normal adult human concept of causation. While

<sup>4</sup> While this is an example for how different forms of causal cognition (exploiting difference-making or geometrical mechanical information) can achieve the same end, the example at the same time shows how these differences will be reflected in the kinds of problems an animal can solve and the kinds of mistakes it will probably make in trying to solve a problem.

some philosophers (e.g. Hall 2004) argue that humans simply have two different concepts of causation, Woodward points out the possibility that these may rather be two aspects of the same concept. However, even if they are integrated aspects of one and the same concept in humans, Woodward (2011, p. 35) argues that this need not necessarily be the case in all organisms:

“[T]here is nothing inevitable about the integration of the two concepts (or elements, or strands) in causal cognition. That is, it seems entirely possible that a creature might [...] be sensitive to some simple spatial or geometrical cues to causal relationships, but not other such cues and might not appreciate *how* [geometrical–mechanical perceptual cues] matter for the kind of difference-making associated with successful manipulation.”

The upshot is twofold. First, following Woodward it seems that difference-making information and geometrical–mechanical perceptual cues are *different kinds of causal information* by which agents can track causal relations in the real world. Secondly, organisms can differ with respect to which kinds of information they can exploit. Thus, even if normal adult human causal understanding involves sophisticated abilities including the possession of abstract concepts like invisible forces and mastery of theoretical principles, organisms that are more limited in this domain may not lack causal cognition altogether. Other organisms may be able to track causal relations and use causal information to guide their actions in a variety of different ways that are less demanding. Naturally, their causal cognition will differ from human causal cognition. But investigating *how* they differ rather than stating *that* they differ should be the main goal of comparative psychology.

## Understanding causality

Comparative researchers seem to agree that a useful concept of causal *understanding* needs to be anchored in human performance: we want to know how similar or different animals are in this respect to humans. Thus, spelling out what it means to *understand* causal relations is supposed to respect the intuitive difference between organisms that are merely sensitive to, and can track and act upon some causal relations on the one hand and organisms that explicitly represent these relations *as causal* on the other hand. For instance, Penn & Povinelli (2009) write:

“Whether a given organism behaves in a way that approximates a given rational model of causal reasoning is not the same question as whether a given organism actually represents and reasons about the entities variables and relationships posited by that model.”

The distinction between *functional level explanations* and *representational level explanations* is supposed to capture that some animals (like chimpanzees, corvids, or rats) are able to exploit causal cues (like weight) when solving some problems, without *understanding* the causal properties and structures involved. On Penn and Povinelli’s view, the representational level of real, human-like causal understanding essentially involves *second-order relational reasoning*: an

abstraction from the perceptual cues provided by the stimulus of a task. In contrast, they characterize the representational level of functionally similar behavior (i.e. successful problem-solving behavior that looks *as if* it involved such second-order relational reasoning) in terms of associative reasoning with a natural predisposition to “perceive certain clusters of features as more salient than others” (Penn and Povinelli 2007).

While the importance of the functional/representational level distinction is not controversial, the simple conceptual dichotomy between full-blown causal understanding, and processes that merely realize *as-if* behavior has been criticized by a number of authors, like Seed et al. (2011). Their criticism is in part empirical, and in part conceptual. The problem based on empirical findings is that the problem-solving behavior of some animals doesn't fit either category: it neither supports full-blown causal understanding, nor can it be appropriately accounted for in terms of associations of arbitrary stimuli (Seed et al. 2011). The conceptual problem is that Penn and Povinelli's conception of the abstraction-involving representational level is said to be too narrow: organisms can have a constrained ability to abstract from perceptual cues of a task in a way that the information represented is not reducible to the perceptual features, but does not involve symbolic knowledge either. They thus introduce a third, intermediate representational category. This intermediate level is characterized as second-order relational reasoning as well, but in contrast to Penn and Povinelli's account it is based on structural (as opposed to *symbolic*) knowledge.

Conceptually, we think that exploring the intermediate level between normal adult human causal cognitive skills and the association of arbitrary stimuli is a move in the right direction. Comparative researchers are not only interested in the question whether some animal's causal cognitive abilities equal those of adult humans, but also a) what the differences are, and b) how various species compare to each other in this respect. An account that posits only a single representational level for all animals that exhibit some ability that merely looks like causal reasoning does not give us insight into either of these questions. However, it is not so obvious whether Seed et al.'s suggestion is more successful.

First, an intermediate level between full-fledged causal understanding and simple associative learning is already present in Penn and Povinelli's account as well. After all, there is a clear difference between explaining behavior with the simple association of arbitrary stimuli, and explaining behavior with associative learning in which some cues that correlate with functionally important features are perceived as more salient than others. Secondly, the *structural knowledge account* is not significantly more explanatory. While it is true that the *biased associations hypothesis* leaves room for a good deal of species differences and thus cannot account for these differences, the structural knowledge approach is just as underdetermined in this respect. Seed et al. (2011, p. 107) are optimistic that future research will work out the details how to capture species differences in their framework (i.e. which species form which kind of abstract, multimodal representations of which structural properties of objects), but future research may also provide us with a more specific model of biased associations that has the resources to capture fine-grained species differences. Thus, until we get a better grasp on how to carve up this intermediate space more precisely and where to locate animals in this space, we should abstain from

judgments concerning the degree of similarity or difference between humans and animals.

In the rest of this section we'll propose a modified version of Woodward's approach to this question. We will discuss three parameters of causal cognition that can at least conceptually dissociate in various ways. The suggestion is to empirically investigate how various species fare concerning each of these parameters. This will give us a better overview over how the parameters involved in causal cognition can dissociate in biological organisms, and thus promises to give us insight into the nature and the evolution of causal cognition, and ground normative judgments about causal understanding. The intent of our approach is to abstract away from the fine-grained details of each species' ecology. However, we recognize still that there might be broad characterizations of niches that map on to our three dimensions. For example, extractive foraging might require more integration of different sources of information, as might living in groups. Importantly, this approach is open to the possibility that there are no clear lines between different representational levels in causal cognition, but rather a (multi-dimensional) continuum of different degrees of understanding.

## Parameters of causal cognition

### *a) Sources of causal information*

The first parameter concerns how agents acquire information about causal relations. Humans can exploit different sources of causal information. One source, that has already been mentioned, are an agent's own actions. In manipulating causes and observing the effects, an agent can gain insight into the causal relation of these two things. This kind of causal learning has sometimes been called *ego-centric* causal learning (Papineau 2003; Woodward 2011). But humans do not always need to perform interventions themselves to learn about causal relations. For instance, children can learn that pressing a switch turns the light on or off by observing that others press the switch. Furthermore, one can learn that shaking a tree will result in apples falling from the tree by observing that the same effect occurs when the wind shakes the tree. Hence, humans can also learn by observing the effects of the manipulations of others (Woodward calls this *agent causal learning*, we will use the label *social causal learning*), or by observing the right kind of natural co-variation (*observational causal learning*).

While normal adult human beings can exploit all three sources of causal information, conceptually there is nothing wrong with an organism that is more limited in this respect, i.e. an organism that can use only one or two sources, and every combination seems at least conceivable. Whether all these combinations are in fact possible is an empirical question. Furthermore, there could be a hierarchy of processes, i.e. the ability to exploit one source could be a prerequisite for the ability to exploit others. In this case, all organisms with source limitations would be limited in a similar way. There is some empirical support that ego-centric causal learning is much more widespread in the animal kingdom than forms of observational learning. For instance, it seems that many bird species are able to pick up causal information

as a consequence of their own actions. But they are often poor social and observational learners. New Caledonian crows, for example, can solve a diverse range of physical cognition problems, such as reasoning by exclusion (Jelbert et al. 2015), the trap tube/trap table task (Taylor et al. 2009a, b), meta-tool tasks (Taylor et al. 2007, 2010), and variants of Aesop's fable task (Logan et al. 2014). They are also able to learn about causal relations from observing natural variation. Jelbert et al. (2019) demonstrated that they could infer an object's weight by observing its movement in a breeze. However, they show no ability to imitate the actions of conspecifics (Logan et al. 2016), and they solve tasks designed to test for collaborative abilities in an individualistic manner with no understanding of cooperation (Jelbert et al. 2015). One reason for this could be that the outcome of one's own actions naturally draws attention to itself because of its more immediate (positively or negatively) reinforcing consequences. Another possibility is that learning from observation is more difficult, because it involves a translation from an observer perspective into one's own body schema/first person perspective (Whiten and Ham 1992).

Furthermore, not all causal cognition that is somehow mediated by living in social groups involves the ability to extract causal information via observation. Social groups can structure an individual's environment in a way that makes ego-centric causal learning more likely, even if individuals do not pay special attention to the actions of others. For instance, being surrounded by conspecifics that use a certain tool increases the probability for an animal to find out on its own what the tool can be used for. Sweet-potato washing in Japanese macaques is often cited as an example (Tomasello 1999), but it has also been argued that a great deal of tool-use in chimpanzees might be explainable along those lines (see Sterelny 2003).

Moreover, some animals like chimpanzees can extract information about causal relations between actions and outcomes by observing the actions of their conspecifics more directly (e.g. Horner and Whiten 2005). Extracting causal information from the observation of natural co-variation (natural causality) appears to be more demanding. Blaisdell et al. (2006) found that rats can extract causal information by observing natural co-variation, but do so more often when agents are involved. Attention could play a role here as well. In human ontogenetic development, the ability to exploit natural co-variation also seems to develop last: even 24 months old toddlers can use social causal learning to guide their actions, but there is evidence that they cannot extract the same causal information from settings which do not involve intentional agents (Bonawitz et al. 2010). This ability only develops later, or has to be scaffolded via intentional language. Findings like this further support Woodward's claim that dissociations between the components of full-blown causal understanding are not only conceptually possible but actual in human beings in different developmental stages.

#### *b) Integration*

The second parameter concerns the question how animals can combine different pieces of causal information, or information originating from different sources. More explicitly, the parameter *integration* refers to the holistic structure of information, i.e. the extent to which an organism can update, extend or combine one piece of information with other pieces of information. While human causal cognition is

characterized by a high degree of integration, there are different ways in which an organism may be limited in this respect.

First, an animal may be limited in its ability to combine perceptual cues of geometrical–mechanical aspects with difference-making information (Woodward 2003). Here the question is not only whether an animal is sensitive to both kinds of information, but also the extent to which an animal can put together these different kinds of information, and to which it can update representations of one kind of information in the light of representations of the other kind of information to predict outcomes or plan actions. There are two obvious ways in which an animal’s ability to integrate causal information in this respect could be limited. The less interesting one simply concerns an upper bound of complexity that can be computed by an organism. It is less interesting because there is nothing mysterious about limitations in this realm: humans also fail to grasp implications of different pieces of causal information they possess if things get complicated enough. Nevertheless, this could play a role in accounting for species differences concerning causal cognitive abilities. A more interesting possibility concerns architectural constraints. This could be the case if the representations of the different kinds of causal information were informationally encapsulated. Empirically, these possibilities could be distinguished by investigating whether animals lack this ability altogether (architectural constraints), or whether the threshold of complexity they can compute is simply lower than in the case of humans.

Furthermore, an animal may be limited in its ability to integrate causal information coming from different sources. Information that is acquired as a consequence of one’s own actions may be limited to guiding the animal’s own actions, while not being available to predict the outcome of the actions of conspecifics. Similarly, an organism may use causal information extracted from observing others, or natural causality for predicting what is about to happen, but fail to use it to inform her own actions. Again, there is some evidence that this is not only a conceptual possibility, but an empirical fact of human development, too. For example, in an experiment by Bonawitz et al. (2010), 24 months old toddlers showed sensitivity to causal structures by learning a predictive relationship between two physically connected events. However, they failed to use this predictive knowledge to initiate action themselves. This suggests a limitation to use information stemming from observational sources for one’s own interventions (see Woodward 2011). An interesting empirical question in that context is in how far these different aspects of integration are related, i.e. in how far, while conceptually separable, ability to integrate different kinds of causal information predicts ability to integrate information from different sources.

### *c) Explicitness*

The last parameter we would like to introduce to our model concerns the extent to which an organism’s representations of causal relations are explicit rather than implicit. Woodward relates explicitness to a representation’s availability to figure in conscious reasoning, the agent’s ability to report it, and to the agent’s ability to use it “in a variety of different sorts of reasoning and planning [...]” (Woodward 2011, p. 40). The first two criteria, however, are all but ideal to investigate animal cognition: the behavioral criteria for consciousness are notoriously unclear, and we

have to drop reportability for lack of language in animals. The last criterion is more promising, and it may best be understood in contrast to *integration*.

While integration is mainly concerned with the ability to update and combine causal information, we take explicitness to be about *what an animal can do with causal information*, i.e. how these representations can fuel flexible behavior. This is obviously related to the availability of representations for different sorts of reasoning or planning. Although some degree of integration seems necessary for the flexible use of information, it is not sufficient, as can be illustrated by the piping plover's (*Charadrius melodus*) *broken wing display*: these birds integrate a lot of information to update their classification of other organisms as predator (Ristau 1991), but once they identify something as a predator, their behavioral repertoire is limited to a single response (see Sterelny 2003, pp 27–29). Apparently, they cannot represent predators independently from how to react to them, i.e. their representations of predators are highly implicit. The behavioral flexibility that comes with the piping plover's ability to integrate information is restricted to adjusting the triggering conditions for a fixed behavioral routine, and representations of this kind are not available to a wide range of different sorts of reasoning.

In the philosophical literature, representations underlying this kind of behavior have been analyzed as *imperative*, combining informational aspects (indicating states of the world, what Searle (1983) called a *mind-to-world direction of fit*) with motivational aspects (*world-to-mind direction of fit*) in a way that ties them to specific behavioral reactions (Millikan 1996; 2006). Another way to put this is to say that these implicit representations fuse means and ends into a single representation. Given this, explicitness can be spelled out in terms of an animal's ability to represent means *decoupled* from ends, and, on a more fine-grained level, "the extent to which means themselves are decoupled into representations of more proximate and distal causes" (Woodward 2011, p. 44). Importantly, the degree of explicitness of representations can vary between different representations within a single agent such that an animal may represent some difference-making relationships as more explicit than others.

So how can we investigate explicitness? Since it is tied to flexible behavior, we should investigate the degree of flexibility with which an animal can use causal information. Woodward suggests, that the higher the degree of explicitness, the more representations of means and ends "incorporate detailed information about how to alter means in the face of changing circumstances to achieve the same goal" (Woodward 2011, p. 21), or about when to use the same means to achieve different goals.

Thus, three kinds of tasks seem particularly informative. *First*, one can test in how far an animal is able to adjust a learned solution to a similar task, by presenting a modified version of the problem to the animal, that requires some *behavioral modification* as compared to the original solution. This differs importantly from standard transfer tasks (what Heyes (1993) called *triangulation*), used to identify the perceptual cues an animal uses to track functional relationships: there the functional set-up remains similar or identical, and only the perceptual cues get modified.

*Secondly*, one can investigate whether competence in some tasks concerning a functional property (like weight, length or flexibility) or some difference-making

relationship accelerates their ability to learn to solve a novel task. Are animals that learned to avoid the trap in the trap-tube task able to use their knowledge about traps, when successful performance demands to *use rather than to avoid* the trap, e.g. in order to retrieve food (Seed et al. 2006) or, in meta-tool use tasks, to retrieve a tool (Taylor et al. 2007)? And how fast can they learn about the function of novel tools (Herrmann et al. 2008; Taylor et al. 2011)?

Finally, explicitness is also related to *insight learning*, since the ability to solve a problem never encountered before without extensive trial and error may be explainable by an agent's ability to exploit background knowledge about functional properties and difference-making relationships whose relevance for the novel problem one appreciates (however, for an alternative explanation for "insight-learning" in some tasks see Taylor et al. (2012)). The *floating peanut task* is a nice example for some sophistication in this domain in orangutans (Mendes et al. 2007) and chimpanzees (Tennie et al. 2010; Hanus et al. 2011).

## From causal cognition to causal understanding

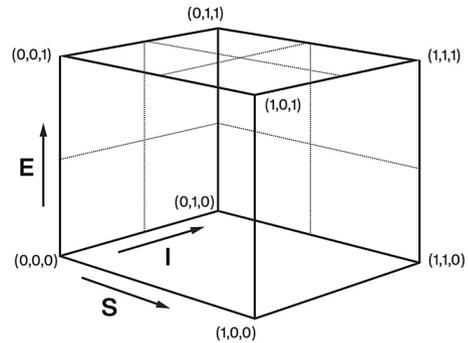
In the preceding section, we argued that organisms can vary considerably concerning their causal cognitive abilities. This variation not only concerns what is known, but also the sources of information they can exploit, their ability to integrate this information, and the flexibility with which they can use this information to guide their actions. So far, the discussion concerning the parameters of causal cognition and how they can dissociate has mostly been on a conceptual level: not all dissociations that are conceptually possible may be possible in actual biological organisms. Let us call this the *conceptual space of causal cognition* (CSCC). In this section, we will discuss how we can use CSCC to capture fine-grained differences and similarities between species, develop an empirically grounded notion of causal *understanding*, derive hypotheses about the evolution of causal cognition, and guide future empirical research.

### A three-dimensional model of causal cognition

One of the central problems in investigating causal understanding in animals is that there is ongoing controversy on what constitutes understanding. The solution we propose is to start with the more neutral and less normative notion of causal *cognition* instead, and work our way up from there. The idea is this: *First*, we start with CSCC. Since there are three parameters of how organisms can deal with causal information, we can think of it as a three-dimensional space (see Fig. 1).<sup>5</sup> *Secondly*, we can map empirical data of animal behavior onto this space. Which sources can

<sup>5</sup> Three-dimensional models are well-suited to represent gradations rather than binary differences. They have been used by philosophers productively in the past to account for (and represent) major evolutionary transitions. The model we use here, including the numerical numbering from zero (=ability not present) to (=ability developed without constraints) has been adopted from Godfrey-Smith's (2009) "Darwinian populations and natural selection".

**Fig. 1** The Three-dimensional space of causal cognition. The vectors: I = integration; E = explicitness; S = sources of causal information. The back top right corner (max. values 1,1,1) represents the highest degree of causal cognition along all parameters



an animal use? To which degree can an animal integrate different pieces of information? How flexible can it use this information? Following the three vectors gives us a specific point within CSCC. *Thirdly*, we can compare species by comparing their different locations. The model then represents similarities and differences between species as proximity or distance within the three-dimensional space of causal cognition. And it does so without any reference to causal *understanding*.

The main strength of this model is that it allows for a more fine-grained evaluation of causal cognitive abilities in animals than models that only distinguish between two or three coarse grained stages of causal cognition or understanding. Thus, while the latter are mostly limited to answering *whether* animals understand this or that aspect of causality, we can analyze much more precisely *how* they are different or similar to humans. Furthermore, in attempting to describe fine-grained differences in causal cognitive abilities, the model addresses a question that *every* comparative account of causal understanding faces: how to account for differences in species whose causal cognitive abilities fall short of adult human ability. Thus, starting with CSCC should be acceptable for romantics and killjoys alike. With this suggestion we do not claim to settle the debate, but rather propose to take a step back to a less controversial point: causal *cognition* (not understanding) involves abilities that are *conceptually* dissociable, while the extent to which they can actually dissociate is an empirical matter. While being less controversial though, this approach is more in line with the goals of comparative psychology, because it works on a more fine-grained level of analysis. Furthermore, any evaluative argument concerning where to draw the line (or the lines) between understanding and not-understanding causality should be evaluated in the light of how CSCC turns out once we mapped the data of various species on it (see Sect. 4.2).

### The evolution of causal cognition and the nature of causal understanding

Once the model includes enough data, it can be used to address further questions concerning the nature and the evolution of causal cognition. For instance, we may ask why we find some dissociations, but not others. Is there a reason that some

abilities (say, observational causal learning) never occur in the absence of others (say, ego-centric causal learning)? Is there a systematic relation between different vectors representing the different parameters? Looking at the cognitive designs we find tells us something about how brains realize causal cognition, and thinking about those designs which we do not find offers a window into what may not be possible to implement. Hence, the question why some things are possible and others apparently not produces interesting questions to be addressed in future research.

Similarly, the model may shed light on the evolution and the development of causal cognition once it contains data of many species. Comparing our close primate relatives can give us a clear picture of the trajectory in which human causal cognition evolved. Comparing the evolutionary trajectories from species that are only distantly related, like corvids and great apes, gives us insight into whether other evolutionary routes from simple to sophisticated causal cognition are possible. Likewise, mapping onto the model a single species in different developmental stages (say human children of various ages and adults) gives us insight into the ontogenetic development of causal cognitive abilities in that species.

Another feature of the model is that it sheds some light on the question whether some animals have causal *understanding* or not. While there is conceptual disagreement concerning the notion of *causal understanding* (see Sect. 2.2), there seems to be a broad consensus that human level performance counts as the paradigm case for causal understanding. One way to understand the normative question then is to ask *how similar animal causal cognition has to be to count as understanding causality*. Different researchers may have different views concerning that question. But for the question whether or not an animal has *real causal understanding* to carry some weight, there should be a *significant* difference between causal understanding and mere causal cognition. In other words, if the concept of causal understanding does not refer to an ability that is significantly different from other forms of causal cognition, the concept is not an interesting one for comparative causal cognition research. Given that, our model is informative concerning two questions. First, it leaves open where to locate humans in this model, thus making sure to avoid the mistake Buckner (2013) calls *anthropofabulation*: raising the bar too high for what counts as causal understanding because we tie competence criteria to an exaggerated sense of human cognitive ability. We should not just assume that humans occupy location (1,1,1). Secondly, since carving up the living world into creatures that understand causality and those that do not emphasizes difference, we can see whether the model confirms that there is a significant gap between some species and all other species.<sup>6</sup> A huge difference between causal cognitive abilities would be represented in the model as a noticeable empty space: is that what we find? And if so,

<sup>6</sup> This point needs to be qualified to some extent: since cognitive abilities are hard (if not impossible) to infer for extinct species, we are faced with a problem stemming from the incomplete material record. As a result, we cannot rule out that gaps in cognitive abilities between species are artifacts of missing data from extinct species. This is especially important in the hominin lineage. In other words, we should be wary to not jump to conclusions if we find such gaps. The problem of incomplete material record does not, however, affect the model's informative value in the case where we do not find such a significant gap. In this case, the model could falsify claims according to which there is a substantial gap between, say, human and non-human animal causal cognition. We would like to thank anonymous Reviewer # 2 for pointing this out.

where do we find it? Or do we find various gaps, indicating more than one stage of causal understanding? In that sense, the model could be used to develop an empirically grounded concept of causal understanding. Alternatively (if the model does not support the view of a significant gap), we should refrain from dualistic thinking concerning causal understanding. In other words, this would support the view that there is simply causal cognition to varying degrees. In this case, the notion of causal understanding does not add anything of importance in the context of comparative cognition research since specifying the degree of understanding we find in an animal would be equivalent to specifying its causal cognitive abilities in the model we propose. In any case, it may be not super important how we label a specific set of causal cognitive abilities. What matters the most is how to evaluate claims about understanding and not-understanding (or degrees of understanding) in terms of similarity and difference—and the model provides a key for such evaluations.

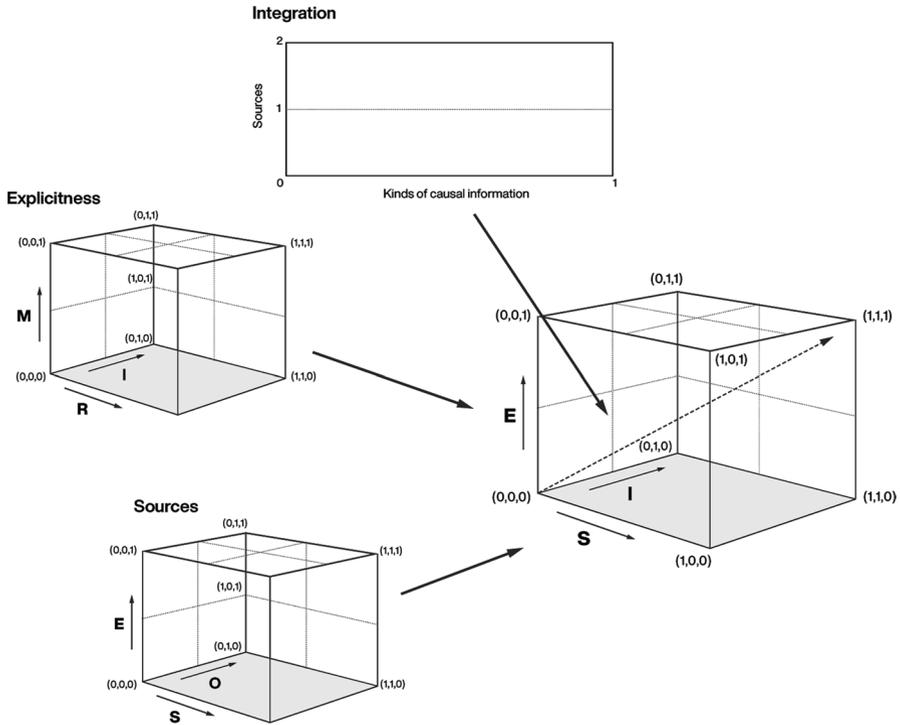
### The metrics of the model and future research

So far, we have simply assumed that it would be an easy task to map empirical data on the model, which unfortunately it is not. The main problem that needs to be solved is how to assign values to each parameter. We need a metric to translate behavioral data to a specific location on each axis. Furthermore, to make things even more complicated, we have to solve a complexity problem: the parameters themselves consist of various dimensions. This is most obvious for the *sources*-vector: it expresses the extent to which an agent is competent to exploit three different sources. But it also applies to the other parameters. The behavioral capacities related to *explicitness* are the ability to modify behavior appropriately in the face of changing circumstances, to realize that the same means can be appropriate for different goals, and to solve problems without prior experience (insight learning). The underlying hypothesis that binds these features together is that they all are expressions of the same architectural feature that allows them to represent means decoupled from ends. But it is conceivable that ability in one of these tasks is a poor predictor for ability in the other kinds of tasks. Similarly, the value for integration has been defined in terms of the ability to combine perceptual cues of geometrical–mechanical aspects with difference-making information, and in terms of the ability to integrate causal information originating from different sources. Again, there is no conceptually necessary link between these criteria.

The solution we propose to the complexity problem is to zoom-in to the model, such that the value of each vector is the mean value of separate two- or three-dimensional vector-spaces, as depicted in Fig. 2.<sup>7</sup>

However, one problem remains: Solving the complexity problem does not solve the problem of finding exact metrics that translate behavioral data into values for the vectors, such that we can compare the causal cognitive abilities in different species. Rather, the metrics-problem arises anew because our suggestion for the complexity

<sup>7</sup> Given what we know, we see no reason to assign more weight to anyone parameter in comparison to the others. If all parameters carry the same weight, taking the mean value seems plausible. However, whether this turns out to be the best way of calculating the values remains to be seen, and future research may provide reason to calculate these values differently.



**Fig. 2** Each vector value of CSCC is the product of a separate vector space. Top: the two-dimensional space constitutes the vector for integration ( $x$ -axis: integration of kinds of causal information;  $y$ -axis: integration of information originating from different sources). Left: three-dimensional space representing explicitness, with parameters  $i$ =insight learning;  $m$ =modified behavior to reach same goal;  $r$ =recognizing novel situation for same means to different goal. Bottom: three-dimensional space representing sources, with parameters  $e$ =ego-centric causal learning;  $s$ =social causal learning; and  $o$ =observational causal learning

problem *presupposes* that we can translate behavior into comparable values. We are confident that the metrics problem is tractable, by standardizing test scores for example, as is common in psychological research, but the full solution to this problem is beyond the scope of this paper and needs to be addressed in future research.

## Conclusion

In this paper, we proposed a step towards ending the animal cognition war. Ending a war is not very likely to be achieved by simply declaring a winner, and this is not the strategy we pursued. After identifying conceptual disagreement concerning the notion of causal understanding as one of the main forces driving the controversy, we suggested bracketing off normative issues concerning the notion of *understanding* and shifting the debate towards empirically tractable questions concerning the less contentious notion of *causal cognition*. Building on Woodward's idea that causal cognition is a complex feature that consists of various different abilities, we argued that the best way to advance comparative causal cognition research is to investigate the extent to which these conceptually dissociable abilities can dissociate in actual biological organisms. We argued that the three central parameters *sources*, *integration*, and *explicitness* span a three-dimensional conceptual space of causal cognition. Mapping the causal cognitive abilities of different species onto this geometrical model allows us to answer comparative questions in a much more fine-grained manner than dualistic approaches to causal understanding. Moreover, we outlined how using this model can help to investigate the nature of causal cognition, answer questions concerning the evolution and development of causal cognition within and across species, empirically ground the normative concept of causal understanding, and derive interesting questions for future research.

**Acknowledgements** We would like to thank Albert Newen for helpful discussion and Antonella Tramaccere for comments on an earlier draft of this manuscript. We would also like to thank Michelle O'Reilly for help with the figures.

**Funding** Open Access funding enabled and organized by Projekt DEAL..

## Compliance with ethical standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Blaisdell AP (2008) Cognitive dimension of operant learning. In: JH Byrne (ed) Learning and memory: a comprehensive reference, vol 1. Academic Press, Oxford UK, pp 173–195
- Blaisdell AP, Sawa K, Leising KJ, Waldmann MR (2006) Causal reasoning in rats. *Science* 311:1020–1022

- Bonawitz E, Ferranti D, Gopnik A, Meltzoff A, Woodward J, Schulz L (2010) Just do it? Toddlers' ability to integrate prediction, action and expectation about contact relations. *Cognition* 115(1):104–117
- Buckner C (2011) Two approaches to the distinction between cognition and mere association. *Int J Comparative Psychol* 24(1):1–35
- Buckner C (2013) Morgan's Canon, meet Hume's Dictum: avoiding anthropofabulation in cross-species comparisons. *Biol Philos* 28:853–871
- Corballis MC (2007) Recursion, language, and starlings. *Cogn Sci* 31:697–704
- Dawkins R (1996) *Climbing mount improbable*. Norton, New York
- Dennett D (1983) Intentional systems in cognitive ethology: the 'Panglossian paradigm' defended. *Behav Brain Sci* 6:343–390
- Dennett D (1987) *The intentional stance*. The MIT Press, Cambridge
- Dickinson A, Shanks D (1995) Instrumental action and causal representation. In: Sperber D, Premack D, Premack A (eds) *Causal cognition*. Oxford University Press, Oxford, pp 5–25
- Eells E (1991) *Probabilistic causality*. Cambridge University Press, Cambridge
- Fitzpatrick S (2008) Doing away with Morgan's Canon. *Mind Lang* 23(2):224–246
- Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. *Nature* 440(7088):1204–1207
- Godfrey-Smith P (2009) *Darwinian populations and natural selection*. Oxford University Press, Oxford
- Halina M (2015) There is no special problem of mindreading in non-human animals. *Philos Sci* 82(3):473–490
- Hall N (2004) Two concepts of causation. In: Collins J, Hall N, Paul L (eds) *Causation and counterfactuals*. MIT Press, Cambridge, pp 225–276
- Hampton R (2019) Parallel overinterpretation of behavior of apes and corvids. *Learn Behav* 47(2):105–106
- Hanus D (2016) Causal reasoning versus associative learning: a useful dichotomy or a strawman battle in comparative psychology. *J Comp Psychol* 130(3):241–248
- Hanus D, Call J (2008) Chimpanzees infer the location of a reward on the basis of the effect of its weight. *Curr Biol* 18(9):R370–R372
- Hanus D, Call J (2011) Chimpanzee problem-solving: contrasting the use of causal and arbitrary cues. *Anim Cogn* 14(6):871–878
- Hanus D, Mendes N, Tennie C, Call J (2011) Comparing the performances of apes (*Gorilla gorilla*, *Pan troglodytes*, *Pongo pygmaeus*) and human children (*Homo sapiens*) in the floating peanut task. *PLoS ONE* 6(6):e19555
- Hare B, Call J, Tomasello M (2006) Chimpanzees deceive a human by hiding. *Cognition* 101:495–514
- Heinrich B (2000) Testing insight in ravens. In: Heyes C, Huber L (eds) *The evolution of cognition*. Vienna series in theoretical biology. The MIT Press, Cambridge, MA
- Herrmann E, Wobber V, Call J (2008) Great apes' (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, *Pongo pygmaeus*) understanding of tool functional properties after limited experience. *J Comp Psychol* 122(2):220–230
- Heyes CM (1993) Anecdotes, training, trapping, and triangulating: do animals attribute mental states? *Anim Behav* 46:177–188
- Heyes CM (2012) Simple minds: a qualified defense of associative learning. *Philos Trans R Soc Lond B Biol Sci* 367:2695–2703
- Hohwy J (2013) *The predictive mind*. Oxford University Press, Oxford
- Horner V, Whiten A (2005) Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Anim Cogn* 8:164–181
- Hume D (1995) *An inquiry concerning human understanding*. Prentice Hall, Upper Saddle River
- Jelbert SA, Taylor AH, Gray RD (2015) Reasoning by exclusion in New Caledonian crows (*Corvus moneduloides*) cannot be explained by avoidance of empty containers. *J Comp Psychol* 129(3):283–290
- Jelbert SA, Miller R, Schiestl M, Boeckle M, Cheke LG, Gray RD, Taylor AH, Clayton NS (2019) New Caledonian crows infer the weight of objects from observing their movements in a breeze. *Proc R Soc B* 286:20182332
- Kabadayi C, Osvath M (2017) Ravens parallel great apes in flexible planning for tool-use and bartering. *Science* 357(6347):202–204
- Karin-D'Arcy MR (2005) The modern role of Morgan's Canon in comparative psychology. *Int J Comp Psychol* 18:179–201

- Leslie A (1995) A theory of agency. In: Premack D, Premack A, Sperber D (eds) Causal cognition, Oxford University Press, Oxford, pp 121–141
- Lewis D (1973) Causation. *J Philos* 70:556–567
- Logan CJ, Jelbert SA, Breen AJ, Gray RD, Taylor AH (2014) Modifications to the Aesop's Fable Paradigm change new Caledonian crow performances. *PLoS ONE* 9(7):e103049. <https://doi.org/10.1371/journal.pone.0103049>
- Lurz RW (2011) Mindreading animals. The debate over what animals know about other minds. MIT Press, Cambridge
- Meketa I (2014) A critique of the principle of cognitive simplicity in comparative cognition. *Biol Philos*. <https://doi.org/10.1007/s10539-014-9429-z>
- Mendes N, Hanus D, Call J (2007) Raising the level: orangutans use water as a tool. *Biol Lett* 3(5):453–455
- Mikhalevich I, Powell R, Logan C (2017) Is behavioral flexibility evidence of cognitive flexibility? How evolution can inform comparative cognition. *Interface Focus* 7:20160121
- Millikan R (1996) Pushmi-Pullyu representations. *Philos Perspect* 9:185–200
- Millikan R (2006) Styles of rationality. In: Hurley S, Nudds M (eds) Rational animals?. Oxford University Press, Oxford, pp 117–126
- Morgan CL (1903) Introduction to comparative psychology. Walter Scott, London
- Papineau D (2003) The evolution of means-end reasoning. In: Papineau D (ed) The roots of reason. Philosophical essays on rationality, evolution and probability. Clarendon, Oxford, pp 83–129
- Penn DC, Povinelli DJ (2007) On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind.' *Philos Trans Roy Soc B Biol Sci* 362(1480):731–744
- Penn DC, Povinelli DJ (2008) Darwin's mistake: explaining the discontinuity between human and non-human minds. *Behav Brain Sci* 31:109–178
- Penn DC, Povinelli DJ (2009) On becoming approximately rational: The relational reinterpretation hypothesis. In: Watanabe S, Huber L, Blaisdell A, Young A (eds) Rational Animals, irrational humans. Keio University Press, Tokyo
- Penn DC, Povinelli DJ (2011) Through a floppy tool darkly. Toward a conceptual overthrow of animal alchemy. In: McCormack T, Hoerl C, Butterfill S (eds) Tool use and causal cognition. Oxford University Press, Oxford, pp 69–88
- Povinelli DJ (2000) Folk physics for apes. The chimpanzee's theory of how the world works. Oxford University Press, Oxford
- Redshaw J, Taylor AH, Suddendorf T (2017) Flexible planning in Ravens? *Trends Cogn Sci* 21(11):821–822
- Ristau CA (1991) Aspects of the cognitive ethology of an injury-feigning bird, the Piping Plover. In: Ristau CA, Griffin DR (eds) Cognitive ethology: the minds of other animals: essays in honor of Donald R. Griffin. Hillsdale Erlbaum, pp 91–126
- Searle JR (1983) Intentionality. An essay in the philosophy of mind. Cambridge University Press, Cambridge
- Seed AM, Tebbich S, Emery NJ, Clayton NS (2006) Investigating physical cognition in rooks, *Corvus frugilegus*. *Curr Biol* 16:697–770
- Seed AM, Hanus D, Call J (2011) Causal knowledge in corvids, primates, and children: more than meets the eye? In: Hoerl C, Butterfill S (eds) McCormack T. Tool use and causal cognition, Oxford University Press, Oxford, pp 89–110
- Shettleworth SJ (2013) Fundamentals of comparative cognition. Oxford University Press, New York
- Sober E (1998) Morgan's Canon. In: Cummins D, Allen C (eds) The evolution of mind. Oxford University Press, Oxford, pp 224–242
- Sober E (2015) Ockham's razors. A user's manual. Cambridge University Press, Cambridge
- Spelke ES, Phillips A, Woodward A (1995) Infants' knowledge of object motion and human action. In: Sperber D, Premack D, Premack A (eds) Causal cognition: a multidisciplinary debate. Clarendon Press, Oxford, pp 44–78
- Spinozzi G, Potì P (1989) Causality I: The support problem. In: Antinucci F (ed) Cognitive structure and development in nonhuman primates. Lawrence Erlbaum Associates, Hillsdale, NJ, pp 113–119
- Starzak T (2017) Interpretations without justification. A general argument against Morgan's Canon. *Synthese* 194(5):1682–1701
- Sterelny K (2003) Thought in a Hostile World. The evolution of human cognition. Blackwell, Oxford
- Taylor AH, Hunt GR, Holzhaider JC, Gray RD (2007) Spontaneous metatool use by New Caledonian crows. *Curr Biol* 17:1504–1507

- Taylor AH, Hunt GR, Medina FS, Gray RD (2009a) Do New Caledonian crows solve physical problems through causal reasoning? *Proc Roy Soc Lond B* 276:247–254
- Taylor AH, Roberts R, Hunt G, Gray R (2009b) Causal reasoning in New Caledonian crows. *Commun Integr Biol* 2(4):311–312
- Taylor AH, Elliffe D, Hunt GR, Gray RD (2010) Complex cognition and behavioural innovation in New Caledonian crows. *Proc Roy Soc Lond B* 277:2637–2643. <https://doi.org/10.1098/rspb.2010.0285>
- Taylor AH, Hunt GR, Gray RD (2011) Context-dependent tool-use in New Caledonian crows. *Biol Lett* 8(2):205–208
- Taylor AH, Knaebe B, Gray RD (2012) An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. *Proc R Soc B* 279(1749):4977–4981
- Tennie C, Call J, Tomasello M (2010) Evidence for emulation in chimpanzees in social settings using the floating peanut task. *PLoS ONE* 5(5):e10544
- Tomasello M (1999) *The cultural origins of human cognition*. Harvard University Press, Harvard
- Tomasello M, Call J (2006) Do chimpanzees know what others see – or only what they are looking at? In: Hurley S, Nudds M (eds) *Rational animals?*. Oxford University Press, Oxford, pp 371–84
- Visalberghi E, Tomasello M (1998) Primate causal understanding in the physical and psychological domains. *Behav Process* 42:189–203
- Vonk J, Povinelli DJ (2011) Social and physical reasoning in human-reared chimpanzees: new data from a set of preliminary studies. In: Roessler J, Lerman H, Eilan N (eds) *Perception, causation, and objectivity. Issues in philosophy and psychology*. Oxford University Press, Oxford, pp 342–367
- Whiten A, Ham R (1992) On the nature and evolution of imitation in the animal kingdom. *Adv Study Anim Behav* 21:239–283
- Wolff P (2007) Representing causation. *J Exp Psychol* 136:82–111
- Woodward J (2003) *Making things happen: a theory of causal explanation*. Oxford University Press, New York
- Woodward J (2011) A philosopher looks at tool use and causal understanding. In: McCormack T, Hoerl C, Butterfill S (eds) *Tool use and causal cognition*. Oxford University Press, Oxford, pp 18–49

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.