The Journal of Mind and Behavior
Summer and Autumn 2015, Volume 36, Numbers 3 and 4
Pages 121–138
ISSN 0271–0137

Detecting Animal Deception

Shane D. Courtland

University of Minnesota, Duluth

By witnessing displays of deception, experimenters may be able to determine whether contemporary (alive today) non-human animals possess a theory of mind. Higher-order deception, in particular, requires the deceiver to be capable of second-order intentionality. The purpose of this essay is to demonstrate the inherent difficulty in scientifically determining whether animals employ higher-order deception. After examining such difficulties, I provide an explication of higher-order deception as "cause–causation with a mental state." This explication will allow us to better determine whether animals are capable of employing such deceit.

Keywords: deception, theory of mind, animal cognition

The use of deception seems to have played a significant role in the evolution of the mind. The animal that is better able to deceive those in his social group may also be better able to reproduce and gain resources. Thus, the smart individuals that detect the deceit (and perhaps employ counter-deception) will gain a competitive advantage. These individuals will be able to end their former exploitation, thereby furthering their own genetic success, perhaps through deceptive acts of their own. This struggle between deception, detection, and counter-deception may have engendered a "cognitive arms race," in which a powerful trait for genetic success is to be slightly more intelligent than the conspecifics that share your lot (Tomasello and Call, 1997).

I do not doubt that the ability to deceive (and to detect deception) would engendered such an "arms race." What is suspect, however, is the extent to which contemporary (alive today) non-human animals are able to deceive. This, as we will see, is a terribly tricky thing to determine. There is an inherent difficulty in scientifically establishing whether animals employ higher-order deception.

I would like to thank Robert Schroer, Radu Bogdan, Raymond Russ, and two anonymous referees for their helpful comments on earlier versions of this essay. Correspondence concerning this article should be addressed to Shane D. Courtland, Ph.D., University of Minnesota, Duluth, Department of Philosophy, 1121 University Drive, Duluth, Minnesota 55812. Email: scourtla@d.umn.edu

Why concern ourselves with higher-order deception? The goal, *ultimately*, is to find evidence that animals possess a theory of mind. According to Heyes, an "animal with a theory of mind believes that mental states play a causal role in generating behavior and infers the presence of mental states in others by observing their appearance and behavior under various circumstances" (1998, p. 102). Currently, there is a significant and intransigent divide: some theorists claim that animals possess this ability, while others remain skeptical.[1]

The goal of this essay is not to defuse or overcome this divide. Instead, it is to examine the difficulties that are incumbent upon one particular method of ascertaining if animals possess a theory of mind: establishing whether they employ higher-order deception. Since higher-order deception presupposes a theory of mind, evidence of such deception would, ipso facto, constitute evidence for an animal theory of mind. As this essay will attest, there are significant conceptual quagmires that must be bypassed in order to successfully employ such a method.

Before proceeding, it is important to acknowledge two caveats to this work. First, this essay is primarily concerned with the conceptual issues that arise when determining whether animals are able to employ deception. It will not focus on the actual, empirical studies that have been conducted. This does not, however, render this work irrelevant or superfluous. Any study that has been conducted, or will be conducted in the future, should address the arguments in this essay. Second, I am not claiming that the detection of higher-order deception is the only way of ascertaining whether animals possess a theory of mind; rather, it is one of many. As Heyes (1998) writes, "The most commonly cited evidence in support of this view comes from studies of imitation, self-recognition, social relationships, *deception*, role-taking (or empathy), and perspective-taking" (p. 102, emphasis added). It is quite possible that we might determine that animals possess a theory of mind through other means (e.g., studies of imitation).

---

[1]Robert Lurz (2011) emphasizes this divide. He writes the following:

> For over thirty years now, there has been a sustained and heated debate over the question of whether any species of nonhuman animal is capable of attributing mental states. The field is presently divided between those who claim that there is empirical evidence to support the hypothesis that some animals are mindreaders and those who deny this. . . . [T]he debate has stalled in that both sides are content with a handful of arguments that, I believe, fail to support the rather strong claims that they are used to make. Contrary to what either side holds, the question of whether animals are mindreaders is in no way a settled matter. (pp.1–2)

Likewise, in Heyes (1998), this divide is heavily stressed. She writes, "In spite of nearly 20 years of research effort, there is still no convincing evidence of theory of mind in primates" (p. 102). Just as there have been many who have claimed that animals can possess some semblance of a theory of mind (see, for example, Braüer et al., 2004; Byrne, 1994; Call and Tomasello, 2008; Cheney and Seyfarth, 1990, 1992; de Waal, 1991; Emery and Clayton, 2009; Gallup, 1982; Hauser and Wood, 2010; Jolly, 1991; Whiten and Byrne, 1991); others remain skeptical that animals have satisfactorily demonstrated this ability (see, for example, Penn, Holyoak, and Povinelli, 2008; Penn and Povinelli, 2007; Povinelli, 1996; Povinelli and Eddy, 1996; Povinelli and Vonk, 2006; Premack, 1988; Vonk and Povinelli, 2006).

The Concept of Deception

Strategic Deception

Robert Mitchell (1986) provides a clear and comprehensive method for describing different notions of animal deception. He divides deception into four levels.[2] Each consecutive level assimilates the characteristics of the level before it.

Mitchell's first level of deception applies to some animals' ability to mimic the appearance or actions of other animals. Defining this level, he writes, "At this level, the actions of the receiver have no influence on the deceptive actions of the sender, at least during the lifetime of the sender. However, the actions of the receiver may have an effect on the population of the sender's species such that more deceitful senders will have greater representation in subsequent populations than will less deceitful senders" (1986, pp. 21–22). There is a certain rigidity in this notion of deception; the animals that employ it seem to be "programmed" to "always do p" or "always look like p." In other words, the deceptive act never changes and thus remains constant for the duration of the animal's life. Changes would arise via selective pressures. The species might alter (over a period of many generations) its deceptive act or appearance due to deleterious selective pressures and/or beneficial adaptations (Mitchell, 1986).

There are many examples of this level of deception in nature. First, there are butterflies with wingtips that look like the heads of their predators. Many predators are deceived by such an appearance, a mistake that allows the butterfly to escape predation. Second, there are some plants whose parts mimic the appearance of the reproductive organs of a particular type of bee. A confused male bee might mistake this part of the plant for a female. When the bee attempts to reproduce with what it believes to be a female, the bee instead pollinates the plant (Mitchell, 1986).

The behavior associated with Mitchell's second level of deception is more flexible than the first. Describing this level, he writes, "At this level, an organism's acts can still be described as programmed, but they appear programmed to the organism's registration of acts of another organism" (p. 24). The organism is still programmed by selective pressures to elicit a particular appearance. However, we get a conditional structure: "'do p given that q is the case' (where p and q are actions of the sender and receiver, respectively)" [p. 24].

At the second level, the animal's actions are more flexible than they were at the first level because the animal is not forced to constantly engage in deception. However, once the organism is stimulated in a certain fashion (q), it invariably conducts the deceptive act (p). Take, for example, the angler fish. This fish has a worm-like lure protruding from between its eyes. However, the lure only moves

[2]In a later work, Mitchell (1993) employs different levels. Instead of four levels, his new account contains six. In addition, some of the contents of the various levels have been altered.

once the fish's prey is present. The prey is tricked by the lure, and the angler fish gets a meal. This is not a maneuver that the angler fish learns; it is a behavior that all members, *ceteris paribus*, of the species possess (Mitchell, 1986).

Complex Deception

The third level's key difference is that it requires learning on the part of the deceptive organism. Mitchell (1986) writes that the "organism's actions are the result of an open program, but this program, unlike that of level two, can be modified by the results of the actions of, and by the observations of, the organism" (p. 25). As with the second level, this level requires the registration of some outside event (e.g., observance of another's action). In the third level, however, the organism's program is more flexible than it was in previous levels. Here, the organism acts under a program such as the following: "do any action p given that this p has resulted in some desired consequence q in our past" (p. 25). Essentially, the organism learns (via repeated trials) that a particular deceptive act yields a desirable outcome.

Since the third level involves learning, it must, unlike the previous levels, contain a time component. In order to learn (via trial and error, observation, instruments, etc.), the organism must be able to recognize the connection between the particular act (p) and the desired effect (q). Granted, some organisms are able to learn faster than others. Thus, for these organisms, the time it takes to learn the connection is shorter. An important rule of thumb, however, is that if an organism takes a sufficiently long period of time to adopt a particular deceptive act (perhaps months, years, etc.), it is probably employing trial-and-error learning; via differential reinforcement, the organism comes to understand that some acts "pay" better than others (Hauser, 2000; Tomasello and Call, 1997).

There are many examples of the third level. First, after a period of time, a dog might learn that it will be pet more often if it appears to have a hurt foot. When around humans, such a dog will limp and then gain the reward for its deceptive act (Mitchell, 1986). Second, a certain primate may get attacked every time it attempts to mate while a rival is in sight. After some time, the primate could develop a strategy in which it would only attempt to mate while the rival is not in sight or if the rival's line of sight is precluded by a rock or some other object (Tomasello and Call, 1997).

An important point to understand about the third level is that it does not require that the deceptive animals (senders) have any beliefs about the belief states (second-order beliefs) of those they are attempting to deceive (receivers). The sender is merely attempting to bring about a state of affairs in the world; it wants to elicit a desirable effect by performing the deceptive act (Mitchell, 1986). It does not, however, understand that it is manipulating the mental states of the deceived (receiver), nor does it intend to.

Animals that are merely conducting level-three deception do not have to be described as second-order intentional systems. Dennett (1978) defines a second-order intentional system as "one to which we ascribe not only simple beliefs, desires and other intentions, but beliefs, desires, and other intentions about beliefs, desires, and other intentions" (p. 273). To think that a particular animal possesses second-order intentionality, we must ascribe to it mental states like "S believes that T desires that p, S hopes that T fears that q, and reflexive cases like S believes that S desires that p" (Dennett, 1978, p. 273). The behavior that is witnessed with level-three deception does not warrant a second-order ascription.

The problem, then, with third-level deception is that the animals who engage in it view the receiver merely as an object. Granted, this object (as viewed by such animals) is quite complex: it is animate and behaves in a complicated fashion. However, there is nothing about these animals that forces us to ascribe beliefs about the receiver's mental states to their behavior. A deceptive act cannot be classified as higher-order deception until the sender deliberately attempts to manipulate the mental states of the receiver.

The fourth level of deception, what I refer to as higher-order deception, is separated from the other levels by two prominent features: *metaprogramming* and the *intention to deceive*.[3] Mitchell (1986) writes the following:

> The fourth level of deception involves an open program which is capable of programming and reprogramming itself based upon the past and present actions of the organism being deceived. . . . [A]t this level of deception the sender actually *intends to deceive* the receiver. When the sender has a deceptive strategy which is malleable with respect to the known current behavior of the receiver and which takes into account the receiver's past behavior, and when the sender's actions appear intelligible only if the animal is described as having planned the deception prior to its actions (in contradistinction to its having planned to manipulate, as may be true for level-three deceivers), then one has an example of intended (or self-programmed) deception. (p. 26)

Like the third level, the fourth level contains an open program. What is added, however, is metaprogramming (thinking or planning): the ability to have representations about representations. In other words, the sender is able to "mentally rehearse" various plans of action and, in light of this rehearsal, is then able to decide which plan to enact (Bogdan, 2000, p. 15). Another important feature is that the sender can represent the receiver as having particular mental states

[3]As mentioned above, Mitchell (1993) provides a modified list of these levels in a later work. In that list, there is a level that precedes higher-order deception. At this new level, the animal is described as possessing the ability to employ pretense and imaginative planning. Mitchell writes, "Given that imaginal pretense is necessary for intentional deception, organisms should exhibit evidence of pretending to be another before we can assume that their deceptions are intentionally deceptive" (pp. 72–73). I agree that pretense/planning is a necessary condition for higher-order deception (intentional deception). As I will demonstrate in the section discussing deception and cause–causation, I view level-three deception as a "broad church." It can include, in my view, pretense/planning and even some instances of causal knowledge.

(Mitchell, 1986). The receiver's states, then, become the target of the sender's manipulation. The sender can thus be described as intending (by means of a deceptive act) to instill false beliefs (via playing upon desires, playing upon hopes, etc.) into the receiver. Therefore, with this level of deception, we refer to the sender as having "intentionally deceived."

Level-four deceptive acts lack the rigidity found in lower levels. In the first two levels, the deception is invariably conducted by every member of the species. These animals seem to be reflexively conducting particular deceptive acts. Fourth-level deception, on the other hand, requires that the sender "planned the deception prior to its actions" (Mitchell, 1986, p. 26). In other words, fourth-level senders are "free" to engage in various forms of deception. They, however, have the innate *capacity* (via metaprogramming) to engage in deception. These deceivers can only be described, in a parsimonious fashion, as having "chosen" a particular deceptive act as the most efficacious (according to the deceivers) method to meet the deceivers' ends.

Unlike those in the previous three levels, fourth-level senders are able to represent the mental states of the receiver. Level one, two, and three senders treat the receiver as a complex object that is to be manipulated in order to achieve the sender's ends. In essence, the sender does not appear to treat the receiver as though it has any mental states at all. With fourth-level deception, the sender is directly trying (intending to deceive) to manipulate the mental states of the receiver. The senders, on this level, are described (at the very least) as being second-order intentional systems.

Humans, most would contend, are the paradigmatic example of entities that engage in level-four deception. An open question, however, remains as to whether any other animal is capable of this kind of deception. It is this question that the remainder of this essay will focus on. I will not, however, answer the question via the presentation of new empirical data. Instead, I will discuss the difficulties that any such answer must avoid.

The Problem

Anecdotal Evidence

Often, researchers who attempt to demonstrate that particular animals can engage in level-four deception employ anecdotal evidence of deceptive, albeit infrequently observed, behavior. Instead of directly testing for level-four deception, these researchers often test other hypotheses (or, while the observation was occurring, not directly testing anything at all). They then, perhaps by chance, stumble upon behavior that seems to fit (in a *prima facie* fashion) a level-four description. Since the particular observation(s) seems to fit such a description, these experimenters then cite the observation as evidence.

Tomasello and Call (1997) cite a particular methodology, in which the central form of evidence is anecdotal. Tomasello and Call (1997) write the following:

> Whiten and Byrne . . . were concerned that many primatologists had observed instances of tactical deception, *but because they were relatively isolated examples*, they were not published. A compendium of such observations might reveal some important patterns. Consequently, the two investigators contacted a large number of primatologists (mostly through membership roles in scientific organizations) and asked them for any observations they might have made over the years that implied tactical deception. They then sifted through the resulting reports (that met the basic functional definition of deception in that one animal did something and another perceived a situation wrongly as a result) and classified them in various ways with an eye to the cognitive mechanisms involved. (p. 234, emphasis added)

In this case, Whiten and Byrne were directly concerned with deception (tactical deception as opposed to primarily level-four deception) and conducted their research with that particular goal in mind. However, the actual evidence (the particular observations) used to further their research is anecdotal. The vast majority of their observations come from primatologists who were not primarily concerned with the notions of tactical deception. In fact, Whiten and Byrne were worried that these observations were so "isolated" in nature, that they might not have ever been published (Tomasello and Call, 1997, p. 234).

Irwin Bernstein (1988) harshly criticizes the anecdotal method Whiten and Byrne used to conduct their research. He writes, "It is indeed disturbing to find Whiten [and] Byrne dismissing studies in which an independent variable is manipulated. To suggest that deliberate manipulations make results suspect is to dismiss most of science without examination. . . . The plural of 'anecdote' is not 'data'" (p. 247). Byrne, however, defends this type of methodology. He writes, "Nearly all of the available data on primate deception are anecdotal, in the sense that they were collected during the course of other studies, ad lib, with no easy means of evaluating frequency against any proper control. However, they are not anecdotal in the pejorative sense; they are not the casual observations of inexperienced observers, embellished by multiple retelling and rife with implicit interpretation" (1998, p. 135). Byrne then cites the following examples of acceptable anecdotal evidence (in the "non-pejorative" sense): "Goodall, 1971; Menzel, 1974; de Waal, 1982, 1986; Byrne [and] Whiten, 1985; and the records of many observers collated in Whiten [and] Byrne, 1986, and in Byrne [and] Whiten, 1990" (1998, p. 135).

Nonetheless, the use of anecdotal evidence is troubling. Of particular interest to this essay is why Byrne feels he must resort to mere anecdotal evidence. He writes, "Conversely, the rarity, spontaneity, and subtlety of tactical deception have hindered experimental analysis of the phenomenon, and little advance has been made in this direction since the pioneering work of Menzel (1974) and papers in Mitchell and Thompson (1986)" [1998, p. 136]. In essence, Byrne claims that the use of anecdotal evidence must be accepted due to the difficulty of observing

deceptive behavior. By explicating what is required in such deception (provided it truly reaches level-four deception), this essay should also show why, as Byrne claims, higher-order deception is difficult to study.

Frequent Observations

The evidence for higher-order deception may come in forms other than infrequent anecdotal observations — via numerous observations, for example. Even if this is the case, however, there are still some potential problems. For instance, if only one deceptive act is both witnessed constantly and present in every member of a particular species, it is most likely not level-four deception. When every member of a species displays the exact same deceptive behavior, it is probably an adaptation, thus most likely a level-one or level-two deception. Since I doubt that this is contentious, I will not discuss it further.

What *is* contentious is whether certain deceptive acts merit a level-four description — as opposed to a mere level-three description. These acts are frequently observed, but not every member of the species practices them; they, then, are most likely not innate. Also, due to the complexity of this type of deception, these acts seem to merit a level-four description. In essence, experimenters can describe this type of behavior as belonging to (at least) a second-order intentional system that is attempting to manipulate the mental states of another animal. What is problematic, however, is that experimenters have *not* precluded a plausible level-three description.

Dennett (1978) presents an example of a dog that "deceives" her master:

> One evening I was sitting in a chair at my home, the *only* chair my dog is allowed to sleep in. The dog was lying in front of me, whimpering. She was getting nowhere in her trying to "convince" me to give up the chair to her. Her next move is the most interesting, nay, the *only* interesting part of the story. She stood up, and went to the front door where I could still easily see her. She scratched the door, giving me the impression that she had given up trying to get the chair and had decided to go out. However as soon as I reached the door to let her out, she ran back across the room and climbed into her chair, the chair she had "forced" me to leave. (pp. 274–275)

It seems (at least *prima facie*) that the dog is engaged in level-four deception. The dog could be attributing mental states to her owner, and then attempting to manipulate such states. We could ascribe to the dog the intention that her "master *believe* she *wants* to go out" (Dennett 1978, p. 275). By intentionally manipulating her master's belief, the dog employs level-four deception and is able to sit in the chair.

To ensure that it is not merely a random confluence of events, suppose that this particular behavior is frequently observed. Of course, the first time it happened, the owner was truly deceived. But after that, the owner allowed the same act to occur over and over again. He went on to document the dog's behavior. Is the

owner, provided this evidence, justified in asserting that the behavior is more than mere level-three deception?

Dennett argues, and rightfully so, that the dog ought not to be viewed as rising above level-three deception (though he does not use this term). He writes, "[The dog] may simply believe, as a good behaviorist, that she has conditioned [the owner] to go to the door when she scratches. So she applies the usual stimulus, gets the usual response, and that's that. [The owner's] case succumbs . . . [to] the more modest hypothesis . . . that the dog believes her master is conditioned to go to the door when she scratches" (1978, pp. 275–276). Dennett is offering a killjoy explanation. This deception does not require that the sender ascribe any mental states to the receiver. Desiring to get into the chair, the dog merely engages in behavior that will allow her to achieve her desired end. The dog, in carrying out this behavior, never has to ascribe beliefs to the owner.

The point I am advancing is that in order to ensure that the witnessed behavior is more than level-three deception (and thus level-four), experimenters must preclude the likelihood that the act was *merely learned* over a period of time (via trial and error). Certain animals, such as the dog, may have learned over time that certain behaviors yield certain responses. When desiring the response, these animals then participate in the behavior that yields that response. They would *never* have to attribute mental states to others.

To determine that an action is not merely a result of trial-and-error learning, experimenters must witness a significant variation in the animal's behavior. For example, the dog must engage in significantly variable behavior in order to accomplish the same result. Instead of only scratching the door, the dog would have to do something like "running to the window and looking out, growling suspiciously" (Dennett, 1978, p. 276). Variation is required because it is the only thing that precludes a level-three description (or less). In order for a behavior to count as evidence of level-four deception, the only likely explanation for that behavior must be that the animal planned (ahead of time) to manipulate the mental states of those it deceived.[4] If the behavior is not significantly variable (such as the normal behavior of an animal that is frequently observed), then a level-three description has not been properly excluded.

We should now be able to see the problem that any research regarding level-four deception must avoid. Significantly variable behavior must be witnessed in order to ensure that the animal is not merely engaging in level-three deception.

---

[4] Although humans are well known for engaging in planned acts of deception, I am not making the stronger claim that they only engage in planned deception (nor would I require that of animals). In sports, for example, players often employ "deceptive techniques [that] are likely to be scripted" (Mitchell, 1996, p. 833). They learn various feints and ruses that become second nature when practiced over time. We even assign names to such common forms of deception (in basketball, for example, "head fake," "stutter step," "the cross over," etc.). In fact, when engaging in such frequent and repetitive acts of deception, athletes may not be giving much thought to engaging in deception — their behavior is, in a sense, automatic (Mawby and Mitchell, 1986).

However, from the perspective of the animal, significantly variable behavior may not always be in pursuit of a deceptive goal. This presents a problem: significantly variable behavior may be just a random confluence of events. The animal in question may be acting in a variant and deceptive fashion only due to happenstance. The researcher, then, must ensure that the witnessed behavior is more than merely a string of events associated by chance.

In order to provide the above necessities, the researcher must employ a controlled experiment that would allow for frequent observation of the variable (but connected) behavior. If the animal frequently performs a particular deceptive act (which would preclude a "chance" description), a likely explanation would be that the animal learned the deceptive act via trial and error, thus casting into doubt the attribution of anything higher than level-three deception. To warrant a higher ascription, we need an experiment that shows that the animal engaged in significantly variable behavior, all the while in pursuit of a particular deceptive end.

Variant Deceptive Behavior: Cause–Causation with Mental States

Cause–Causation — Recognizing Causal Relationships

To determine if a behavior fits a level-four deception, experimenters need to observe an animal frequently employing "cause–causation" while it is representing, and subsequently manipulating, a particular mental state of the animal it is deceiving. In other words, studies need to show that the sender has causal knowledge of the receiver's mental states and that the sender is intentionally manipulating such states.

Let us examine what sort of behavior warrants an ascription of causal knowledge of an object to an animal. This object could be anything, and "knowledge" could vary from the causal knowledge of a stick to the casual knowledge of a conspecific's mental states. What kind of observed behavior, then, provides the observer with a sufficient justification that a particular animal is able to recognize the causal properties of an object?

Radu Bogdan (2000) provides three different types of causal behaviors:[5]

(A) Agent acts → outcome

(B) Agent acts with implement → outcome (p. 45)

[C] Agent acts on → [object → desired outcome] (p. 47)
Cause 1 → Cause 2

[5] " → " denotes cause; "[. . .]" denotes "represented by the agent."

Causal behavior (A) is present in all of the animal kingdom. According to this script, an animal moves its body, and, as a result, there is an outcome. Observation of (A) merely requires a researcher to watch an animal pursue a goal in a causal fashion. Due to its ubiquitous nature, I will not cite any examples. All that the reader has to understand is that the thing acting can be described in a "goal-pursuing" fashion (it is an agent) and that its actions (while pursuing such goals) entail outcomes (Bogdan, 2000).

It is implausible to think that every time an animal moves its body, as in (A), the animal knows that it is employing a causal relationship. Granted, an animal's action might entail an outcome. This, however, does not provide evidence that the animal recognizes or understands that the action is connected to such an outcome. Bogdan (2000) writes, "Most species do not recognize action–outcome relations as causal. They simply act, outcomes follow causally, and the successful pattern is learned or naturally selected and wired in" (pp. 46–47). There are other plausible explanations (e.g., conditioning and hardwiring via natural selection) that account for the observed behavior. Due to the ubiquity of script (A) and the likelihood of other plausible explanations, one cannot claim that (A) provides a sufficient justification for the attribution of causal understanding (Bogdan, 2000).

Causal behavior (B) occurs when an animal uses something other than its body to bring about a certain outcome; it moves its body in conjunction with an instrument that helps it achieve its goals. For example, a rat may be trained to press a certain lever in order to get food. The instrument, in this case, is the lever. The rat acts *with* the lever in order to gain its desired outcome, food (Bogdan, 2000). Fewer animals are observed employing (B) than (A).

Although (B) escapes the problem of ubiquity, it still does not warrant the ascription of causal understanding. Bogdan (2000), referring to the example of the rat, writes the following:

> The rat need not recognize that its action causally relates to an object (the lever) that causally relates to still another object (the food). The rat simply acquires a new instrumental-action pattern that causes culinary satisfaction. An instrumental action is more complex than a simple action, yet it remains tightly attached to its implement, as an extension of it. It is the rat's action with an implement causing an outcome that is trained by the experiment, not the rat's *recognition that* an implement acted upon causes something else. (p. 46)

In essence, the "implement" in (B) might be merely viewed by the agent (the rat) as an extension of its body (an extension of its paw). The rat may just be "habituated" to new actions that happen to involve an instrument (a lever). Since this plausible explanation remains, we are not warranted to ascribe to the rat an understanding of the causal connection between the lever and the food (Bogdan, 2000).

Causal behavior [C], on the other hand, is qualitatively different from the other two forms of causal behavior. Only [C] requires causal knowledge on the part of the acting agent. This script requires the agent to represent (in its mind) the second cause. Bogdan (2000) writes, "$Cause_1$ is causation produced by acting, $cause_2$ is the causation caused by the action in question. . . . Only [C] requires causal knowledge, because it requires the agent to represent distinctly the causation that her action causes. Most organisms are causal agents, simply because they act, but not causal knowers" (p. 47). Script [C] agents act deliberately "*to cause a further causation*" (p. 47, emphasis in original). The agent understands (recognizes or represents) the causal properties of the object that it is acting upon. In order to achieve a goal, the agent acts (cause 1) on an object that it recognizes (or represents) as causally related (cause 2) to such a goal. Type [C] is the causation we are interested in; it is *cause–causation* (Bogdan, 2000).

In order to gain a clearer understanding of [C], we will look at an example. Say I desire to ram another bumper car with the bumper car I am driving. It can be parsed as follows:

Agent acts on pedal → [Car 1 → Car 2 is rammed]

According to the above script, I recognize (or represent) Car 1 as causally related to Car 2. My desired outcome is to have Car 1 ram Car 2. I understand that the relationship between Car 1 and Car 2 can be enacted by pressing the gas pedal of Car 1. I press the pedal and ram the second car. The important part is the section enclosed in brackets. This denotes the point at which I recognize (or represent) the causal connection between Car 1 and Car 2.

Bogdan, in addition, presents two conditions that are intended to further distinguish [C] from (B). The first condition is *generality* — the agent is able to use the particular implement in a variety of fashions and in a variety of contexts. For example, in one instance, I may use a stick (implement) to "fish for termites"; in another instance, I may use the stick to beat a conspecific. In yet another instance, I might use the stick to lift a heavy rock. Such a variety of uses and contexts seems to imply that I *recognize* some of the causal properties of the stick. I can use the stick in different manners because I understand what the stick "does." To demonstrate the satisfaction of this condition, we must witness the agent use the implement (e.g., a stick) in many different fashions/contexts (Bogdan, 2000).

The second condition is *separability*. When describing this condition, Bogdan (2000) writes, "The [causal] knowledge is . . . separate from the specific action utilized (e.g., which hand or finger is used) or particular type of action undertaken (say hand versus nose movement, if the hand is not available)" [p. 48]. The agent must demonstrate that it understands that the implement has causal properties that are distinct from the particular action(s) that activates such properties.

Pretend that I, similar to a chimpanzee, am termite fishing with a particular stick. To possess an understanding of the causal properties of the stick, I must know how the stick interacts with the world in a manner that is separate from me merely "bobbing" the stick with my hand. I must understand that the stick will behave in the same manner (have the same causal properties) regardless of what bobs the stick (e.g., a hand, a foot, someone else's foot, etc.). To demonstrate the satisfaction of this condition, then, we must witness, in a variety of manners, the agent causing the particular implement to bring about the same effect (Bogdan, 2000).

Higher-Order Deception as Cause–Causation

To recap, in order for an animal to be believed to have a casual understanding of an object, it must interact with the object in a cause–causation manner (script [C]). In other words, observation of the animal's behavior must warrant the claim that it satisfies two conditions: generality and separability. For the former, the agent is able to use a particular implement in a variety of fashions and in a variety of contexts. For the latter, the agent must have knowledge of the causal properties of the implement, separate from it being acted upon (Bogdan, 2000).

Next, I will apply cause–causation to deception, particularly higher-order deception, in which the sender is intending to deceive the receiver. The sender can represent the receiver as having particular mental states (Mitchell, 1986). These states become the target of the sender's manipulation. The sender, then, can be described as intending (with the deceptive act) to instill false beliefs (playing upon desires, playing upon hopes, etc.) into the receiver.

The connection between cause–causation and higher-order deception can now be articulated. When particular outcomes are desired, the sender intentionally manipulates the mental state of the receiver, entailing these outcomes:

[D] Agent acts on → 1 [Mental State X → 2 desired outcome]

The brackets indicate the mental representation of the sender. The sender is modeling the mental state of the receiver (a particular representation of the receiver) and the causal connection of such a state to a particular outcome. To gain the desired outcome, the sender acts in a certain fashion (cause 1) to bring about Mental State X. The sender does this action (cause 1) because it understands the connection between X and the desired outcome (the representation of cause 2). The sender, then, can be described as intending (with the deceptive act, cause 1) to instill misrepresentations (playing upon desires, playing upon hopes, etc.) into the receiver. This is level-four deception.

Of course, to reach [D], an agent would have to possess the ability to employ some form of planning, problem solving, event anticipation, etc. At the core of this is the ability to employ imagination. According to Bogdan (2000), imagination is

the "ability to envisage alternative, nonfactual situations, whatever the manner of envisaging — motor or mental images, symbolic representations, whatnot" (p. 58). The ability to employ imagination is a necessary condition for [D]. It is not, however, a sufficient condition. An animal could have this ability yet not be a causal knower. Thus, the behavior in question could fail to reach [C] and, ipso facto, [D].

The ability both to employ imagination and to be a causal knower, though, is still not sufficient to reach script [D]. To see this, it is important to understand the shortcomings of the following script:

[E] Agent acts on → 1 [Autonomous Agent X → 2 desired outcome]

In [E], the sender has a causal understanding of the receiver as an autonomous agent. In other words, the sender recognizes that the receiver is a goal pursuer and an autonomous originator of causes. This recognition (and subsequent causal exploitation), however, does not warrant a level-four description. In order to reach level four, the sender must recognize more than the mere fact that the receiver is a self-caused agent. In addition, the sender must further recognize that the receiver is an autonomous agent that has representations about the world. It is the recognition and manipulation of such representations that defines level four. [E], at its best, would provide for level-three deception.[6]

Of course, in order for an animal's behavior to be described via script [D], its behavior must indicate that it satisfies the generality and separability conditions. This is important and I cannot stress it enough: the observation of the fulfillment of the generality and separability conditions is only required to resolve an epistemic problem. This is meant to establish that a particular animal (whose capabilities are in doubt) can, in fact, engage in higher-order deception. It is possible, then, that an animal can engage in higher-order deception, all the while its behavior not showing that it has met the generality and separability conditions. We are not warranted, however, in ascribing the ability of higher-order deception to an animal until there is some record of these conditions being met.

One may be tempted to object by claiming that I am employing a double standard. After all, I would not require the observation of the generality and separability conditions in order to justifiably ascribe higher-order deception to a particular human — yet I would require the satisfaction of such conditions in the case of an animal. This double standard is justified, however, as the cases are quite different. For animals, we are looking for a genuine case of (possibly) the

[6]Level-three deception requires that the entity is capable of learning from past behavior. This ability can come in different forms, and learning it can involve planning and imaginative pretense. With more cognitive complexity, the ability to plan (and to learn via such plans) will be enhanced. Eventually, we may be warranted in claiming that a particular entity is capable of understanding causal properties, as in scripts [C] and [E]. However, an entity, in my view, has not reached level four (higher-order deception) until it is capable of a particular sort of causal understanding, via script [D].

first ascription of higher-order deception. The satisfaction of such conditions, then, is required to defuse killjoy objections. With humans, on the other hand, we have access to other lines of evidence (e.g., inferences from introspection). These other lines allow us to relax the requirement that our observations satisfy the generality and separability conditions.

Tomasello and Call (1997) provide an example of how the generality and separability conditions might be expressed in animals:

> To take a hypothetical example: Suppose that a subordinate individual has been excluded from a watering hole by more dominant individuals. Now suppose, by chance, a rock rolls down a cliff leading to the water's edge and scares away the drinkers. The subordinate, far enough away not to be frightened by the rock, then precedes to drink in peace. Perhaps the same scene repeats itself a few times over the course of several weeks. From these experiences the observing subordinate may come to understand that the falling rock reliably predicts the scattering of others, but it is questionable whether it understands that the rock leads to fear in others, which then leads to their scattering. Most observers of nonhuman primates would be astounded, we believe, if one day the subordinate individual climbed the cliff and deliberately rolled the rock down toward the others to get them to scatter, since the rolling rock was not something originally produced but observed. . . . It would be even more astounding if the subordinate individual found some novel way to make its groupmates fearful and so scatter, for example, by making some other noise from behind the cliff. A creative strategy such as this would imply that it had understood that the fear of the others was the "mediating variable" (to use Whiten's, 1993, term) and that manipulating that variable in some way (not necessarily involving the rock that had led to the result previously) would also be effective. (p. 386)

The situation would fail to elicit a level-four description had the subordinate merely pushed the rock down the cliff. At best, such behavior warrants a level-three description. The animal might lack understanding of the "mediating variable." In order to warrant a level-four description, the subordinate must manipulate his conspecifics in a significantly variable fashion. If his behavior is sufficiently variable, then he probably understands the causal properties of the mediating variable (fear).

The variable behavior in the above example is an application of the separability condition. Essentially, the subordinate is demonstrating that it has knowledge of the causal properties of the implement (e.g., mediating variable, Mental State X, fear, etc.) separate from his knowledge of the implement being acted upon. The subordinate demonstrates such knowledge by activating the same outcome in different manners. Since the subordinate understands the causal properties of fear as well as the fact that fear plays a role in the behavior of the scattering conspecifics, the subordinate is able to activate the causal properties of the implement (fear → scattering) in a diverse fashion (making a noise behind the cliff as opposed to merely rolling the rock).

It would also be simple to imagine a situation in which the aforementioned subordinate satisfied the generality condition. Essentially, the subordinate

would have to use the particular implement (mediating variable, Mental State X, fear, etc.) in a variety of fashions and contexts.[7] Instead of merely manipulating the mediating variable to achieve an uninhibited drink of water, the subordinate would have to use the mediating variable to achieve different things at different times. For example, he might activate the mediating variable (fear) in order to protect a food source, gain access to a female, or escape from a violent conspecific. Such a variety (of uses and contexts) would imply that the subordinate recognizes some of the causal properties of fear — in other words, that he understands what fear does.[8]

Once these conditions have been satisfied by the subordinate, we can then plausibly describe its behavior via script [D]. The subordinate could be described as possessing a causal understanding of its conspecifics' mental states. This is level-four deception as cause–causation with a mental state.

Notice that by describing higher-order deception in this manner, I am better able to explicate what researchers ought to be focusing on in experiments. Essentially, reaserchers should aim to frequently observe animal behavior that indicates the same mental state is being manipulated in a variety of fashions and contexts. This would be a combination of two components: the frequently inferred mediating variable and the observed variations.

The variability comes via the observed satisfaction of the generality and separability conditions. In order to fulfill the former, we need to witness the agent use the particular implement (e.g., stick or mental state) in many different fashions and contexts. In order to satisfy the latter, we must witness the agent use the particular implement in a variety of manners to bring about the same effect (Bogdan, 2000). The observation of the behavior that satisfies the aforementioned conditions allows the researcher to describe the animal's behavior as higher-order deception. Of course, what I have just described is not meant to be taken as a formal experiment. It is merely meant to provide a basic idea of what phenomena we are trying to witness.

The satisfaction of these conditions requires the observation of variable behavior; however, this does not mean that such behavior is random. It is not

[7]Mitchell (1993) seems to recognize the generality condition. He writes, "But by the time one is an adult, one's interpretation of the pretense is in terms of the other's thoughts and beliefs, and the fakeout sequences can become highly variable. This finding suggests that the gorilla who faked window-slams enjoyed her game because she wanted to see if her fakes could *startle* people, not just because she wanted to see if her fakes could make the people *move* in a particular way" (p. 81).

[8] Of course, this is simplifying matters quite a bit. It may turn out that animals cognize the internal states of conspecifics in different manners than humans do. Fear, for example, may not be a unified concept — thus, for a particular animal, there may be many types/variations of fearful states. All that is required to reach [D] is that an animal possesses a causal understanding of the mental states of its intended deceptive target. The animal would, then, be employing a theory of mind to deceive. However, the content of that theory of mind might very well be different from our own. If an animal's theory of mind is too different, however, we might not be warranted in claiming that it does, in fact, possess a theory of mind. Thanks to an anonymous commentator for bringing these concerns to my attention.

as if an animal merely does things outside of its normal behavior and then researchers claim that the animal is able to commit higher-order deception. Instead, the variable behavior that is a part of the two aforementioned conditions must have a reoccurring theme. Essentially, the animal must manipulate the *same* implement in a variety of fashions, a behavior that must be constantly witnessed in order to describe it as level-four deception.

Researchers ought to observe animals for their frequent and variable manipulation of the same implement. A researcher, thus, would also frequently observe the animal manipulating the same implement to achieve its desired ends. After observing sufficient variation of such manipulations, eventually the only plausible explanation will be that the animal has a causal understanding of the particular object it is manipulating. Since this variation (regarding the same implement) would be frequently observed, it would most likely not be a mere random confluence of events.

I realize that there are additional complexities that remain regarding the study of level-four deception in animals. My discussion was not meant to resolve all of these complexities. Instead, my concern was to show (conceptually) what would be required to warrant an ascription of higher-order deception to an animal. By viewing level-four deception as cause–causation with a mental state, I have presented a first approximation of such a solution.

References

Bernstein, I. (1988). Metaphor, cognitive belief, and science. *Behavioral and Brain Sciences, 11,* 247–248.

Bogdan, R. (2000). *Minding minds.* Cambridge, Massachusetts: MIT Press.

Braüer, J., Call, J., and Tomasello, M. (2004). Visual perspective taking in dogs *(Canis familiaris)* in the presence of barriers. *Applied Animal Behaviour Science, 88,* 299–317.

Byrne, R. (1994). The evolution of intelligence. In P. Slater and T. Halliday (Eds.), *Behaviour and evolution* (pp. 223–265). Cambridge, Massachusetts: Cambridge University Press.

Byrne, R. (1998). What's the use of anecdotes? Distinguishing psychological mechanisms in primate tactical deception. In N. Thompson, L. Miles, and R. Mitchell (Eds.), *Anthropomorphism, anecdotes, and animals* (pp. 134–150). Albany, New York: SUNY Press.

Byrne, R., and Whiten, A. (1985). Tactical deception of familiar individuals in baboons *(Papio ursinus). Animal Behaviour, 33,* 669–673.

Byrne, R., and Whiten, A. (1990). Tactical deception in primates: The 1990 data-base. *Primate Report, 27,* 1–101.

Call, J., and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12,* 187–192.

Cheney, D., and Seyfarth, R. (1990). *How monkeys see the world.* Chicago: University of Chicago Press.

Cheney, D., and Seyfarth, R. (1992). Precis of how monkeys see the world. *Behavioral and Brain Sciences, 15,* 135–182.

Dennett, D. (1978). *Brainstorms.* Montgomery, Alabama: Bradford Books.

de Waal, F. (1982). *Chimpanzee politics.* London: Jonathan Cape.

de Waal, F. (1986). Deception in the natural communication of chimpanzees. In R. Mitchell and N. Thompson (Eds.), *Deception: Perspectives on human and nonhuman deceit* (pp. 221–244). Albany, New York: State University of New York Press.

de Waal, F. (1991). Complementary methods and convergent evidence in the study of primate social cognition. *Behaviour, 118,* 297–320.

Emery, N., and Clayton, N. (2009). Comparative social cognition. *Annual Review of Psychology, 60,* 87–113.

Gallup, G. Jr. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology, 2,* 237–248.

Goodall, J. (1971). *In the shadow of man.* Boston: Houghton.

Hauser, M. (2000). *Wild minds.* New York: Henry Holt.

Hauser, M., and Wood, J. (2010). Evolving the capacity to understand actions, intentions, and goals. *Annual Review of Psychology, 61,* 303–324.

Heyes, C.M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences, 21,* 101–148.

Jolly, A. (1991). Conscious chimpanzees? A review of recent literature. In C. R. Ristau (Ed.), *Cognitive ethology: The minds of other animals* (pp. 231–253). Hillsdale, New Jersey: Lawrence Erlbaum.

Lurz, R. (2011). *Mindreading animals: The debate over what animals know about other minds.* Cambridge, Massachusetts: MIT Press.

Mawby, R., and Mitchell, R. (1986). Feints and ruses: An analysis of deception in sports. In R. Mitchell and N. Thompson (Eds.), *Deception: Perspectives on human and nonhuman deceit* (pp. 313–322). Albany, New York: SUNY Press.

Menzel, E.W. Jr. (1974). A group of young chimpanzees in a one-acre field. In A. Schrier and F. Stollnitz (Eds.), *Behavior of nonhuman primates* (pp. 83–153). New York: Academic Press.

Mitchell, R. (1986). A framework for discussing deception. In R. Mitchell and N.S. Thompson (Eds.), *Deception: Perspectives on human and nonhuman deceit* (pp. 3–40). Albany, New York: State University of New York Press.

Mitchell, R. (1993). Animals as liars: The human face of nonhuman duplicity. In M. Lewis and C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 59–89). New York: Guilford Press.

Mitchell, R. (1996). The psychology of human deception. *Social Research, 63,* 819–861.

Mitchell, R., and Thompson, N.S. (1986). *Deception: Perspectives on human and nonhuman deceit.* New York: SUNY Press.

Mitchell, R., Thompson, N.S., and Miles, L. (Eds.). (1997). *Anthropomorphism, anecdotes, and animals.* Albany, New York: SUNY Press.

Penn, D., Holyoak, K., and Povinelli, D. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109–178.

Penn, D., and Povinelli, D. (2007). On the lack of evidence that nonhuman animals possess anything remotely resembling a "theory of mind." *Philosophical Transactions of the Royal Society B, 362,* 731–744.

Povinelli, D. (1996). Chimpanzee theory of mind? The long road to strong inference. In P. Carruthers and P. Smith (Eds.), *Theories of theories of mind* (pp. 293–329). Cambridge, Massachusetts: Cambridge University Press.

Povinelli, D., and Eddy, T. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research on Child Development, 61,* 1–189.

Povinelli, D., and Vonk, J. (2006). We don't need a microscope to explore the chimpanzee's mind. In S. Hurley and M. Nudds (Eds.), *Rational animals?* (pp. 385–412). Oxford: Oxford University Press.

Premack, D. (1988). Does the chimpanzee have a theory of mind? Revisited. In R.W. Byrne and A. Whiten (Eds.), *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans* (pp. 160–179). Oxford: Oxford University Press.

Tomasello, M., and Call, J. (1997). *Primate cognition.* New York: Oxford University Press.

Vonk, J., and Povinelli, D. (2006). Similarity and difference in the conceptual system of primates: The unobservability hypothesis. In E. Wasserman and T. Zentall (Eds.), *Comparative cognition: Experimental exploration of animal intelligence* (pp. 363–387). New York: Oxford University Press.

Whiten, A. (1993). Human enculturation, chimpanzee enculturation and the nature of imitation. *Behavioral and Brain Sciences, 16,* 538–539.

Whiten, A., and Byrne, R. (1986). The St. Andrews catalogue of tactical deception in primates. *St. Andrews Psychological Reports, No. 10.*

Whiten, A., and Byrne, R (1991). The emergence of metarepresentation in human ontogeny and primate phylogeny. In A. Whiten (Ed.), *Natural theories of mind* (pp. 267–281). Oxford: Basil Blackwell.