

Deception: a functional account

Marc Artiga, Cédric Paternotte

(Penultimate draft, February 2017 – do not cite without permission)

Abstract: Deception has recently received a significant amount of attention. One of main reasons is that it lies at the intersection of various areas of research, such as the evolution of cooperation, animal communication, ethics or epistemology. This essay focuses on the biological approach to deception and argues that standard definitions put forward by most biologists and philosophers are inadequate. We provide a functional account of deception which solves the problems of extant accounts in virtue of two characteristics: deceptive states have the function of causing a misinformative states and they do not necessarily provide direct benefits to the deceivers and losses to the targets.

1. Introduction

Deception is ubiquitous in nature. Subordinate tufted capuchin monkeys (*Cebus apella nigrurus*) send antipredator alarm calls in the absence of predators to distract dominant individuals and access more resources (Wheeler 2009). Fireflies of the species *Photuris* perform lightning patterns which mimic the mating signals of female fireflies of the species *Photinus* to lure males and devour them (Lloyd 1975). Male scorpion flies (*Hylobittacus apicalis*) mimic females in order to receive insects (necessary for courtship to succeed) from other males, only to later offer them to real females (Thorhill, 1979). The Western hog-nosed snake behaves like a venomous snake (which it is not) to deter predators; if this fails, it simulates death and emits a decaying odour (Platt 1969). Fiddler crabs that lost their original claws by predation or aggression sometimes use their regenerated claws (which look similar but are weaker than the original ones) to frighten rivals (Wilson and Angilletta 2015). Similar cases of deception occur in a wide variety of organisms across the living world (Stegmann 2013). Needless to say, deceptive behaviour is also pervasive in the human species. It has even been suggested as one of the driving forces in the evolution of human intelligence (Trivers 2011; Hippel and Trivers 2011).

Nonetheless, deception had historically received scant attention from philosophers and scientists until recently. In biology at least, recent interest in the evolution of cooperation (Sterelny et al. 2013) and animal communication theory (Stegmann 2013), among others, have revealed deception as a central phenomenon that requires a careful analysis. Cognitive science, ethics and epistemology have witnessed a parallel revival, with special interest in philosophical and psychological aspects of self-deception, as well as philosophy of language with a prominent emphasis on lying (Mahon 2015). This recent interest on deception has generated a growing body of literature, but also a whole range of different and usually incompatible definitions of deception, which hamper a common approach and a more comprehensive understanding of this phenomenon.

In this paper, although we focus on a biological approach of deception, we aim to make it consistent with non-biological accounts. This is by no means an original choice, as the existing accounts of biological deception typically draw from intuitions regarding human, intentional deception, as we will see. However, such continuity is rarely acknowledged and explicitly embraced. We will argue that the case of human intentional deception helps locate

insufficiencies in existing definitions of biological deception and motivates an original account. As an added benefit, it may provide a starting point for a general theory of deception across the biological world, humans included.

These aims set the structure of the paper. First, section 2 puts forward a set of conceptual and empirical conditions with regard to which an account of deception may be deemed satisfying. Section 3 introduces a typical definition of deception in evolutionary biology, before discussing why it is lacking – namely because, as many other accounts, it incorporates the condition that deceptive traits should benefit their holder and harm their targets. Section 4 develops and discusses our own functional account of deception, which both escapes pitfalls of existing accounts and shares common points with some accounts of intentional deception (e.g. Smith's recent approach to self-deception). In particular, we argue that deception is consistent with any possible deceiver/deceived payoff profile and fits the other conditions identified in section 2. Section 5 defends our account in the light of two recent ahistorical takes on deception and meaning, due to McWhirter and Birch, drawing from Godfrey-Smith's 'modern history' approach to functions. Section 6 concludes.

2. Conditions on the definition of deception

In order to properly assess current accounts of deception, we first make explicit and discuss a set of conditions which any satisfactory account may meet. While not necessarily exhaustive, this list will, however, be sufficient to delineate an often implicit common background for the discussion to come and to eliminate alternative candidate accounts. Still, our three conditions do not have the same normative status. The first one is an uncontroversial necessary condition, which can be seen as a constraint for any acceptable account. The second condition constitutes an aim for a satisfying account – a virtue that, while not being necessary, is nonetheless desirable whenever attainable. Finally, we discuss a third condition that, while being widely accepted as necessary, should actually be dropped, or so this paper will argue.

The first one (the Error Condition – ErrC), which is a deeply entrenched assumption, is that deception should be distinguished from error (Searcy & Nowicki 2005:4, Skyrms 2010:76, McWhirter 2015:3). To be deceived is not just to acquire a false belief, or to misperceive. While closely connected on most views, deception and such mistakes are distinct phenomena that should not be confused.

The second one, which we call the 'Extensional Condition (ExtC)', concerns the scope of deception. This condition, which we take to be tacitly accepted to a degree by many theorists, is rarely explicitly stated. It states that we should prefer an account that attributes deception in those cases in which science and common sense do. We do not – nor do we need to – claim that it is a necessary condition for a definition of deception; only that others things being equal, an account that accommodates it should be preferred to one that does not. One reason to defend the extensional condition stems from the principle of charity. Failing to satisfy this desideratum would imply that most researchers and philosophers are deeply confused about their subject matter, which is very implausible. When an account excludes agreed examples of deception, this divergence should be motivated and offset by additional theoretical benefits. Also note that the condition does not entail that an account cannot – as ours will – identify new, previously neglected examples of deception.

An important corollary of this condition, especially relevant for the goal of the paper, is that we should prefer a theory that can accommodate cases of deception across different scientific disciplines. In particular, an approach that encompasses intentional and non-intentional deception will provide a unified and more comprehensive perspective.

This last point is worth expanding. Why should we hope for intentional deception (including but not limited to human deception) and non-intentional deception to be covered by a common account? This is for three reasons. First, there is a long tradition of analogies between intentional behaviours and adaptations - traits evolved by natural selection. Adaptations seem to be designed and have a purpose in the same way intentional behaviours do. For instance, just like human altruism and cooperation (say) have evolutionary counterparts and are understood and characterised by similar models (e.g. the Prisoner's Dilemma, the Stag Hunt, etc.), deceptive traits may look as if their bearer intended to deceive its target and be displayed in similar situations. Thus, we think there is an intuitive connection between cases of deception in different domains that motivates an encompassing account.

Second, intuitions about biological deception drawn from its intentional counterpart already pervade the literature. Searcy and Nowicki see their definition of biological deception as "the behaviour [that] has the effects of deception without necessarily having the cognitive underpinnings that we would require of deception in humans" (2005:5); this entails that intentional and biological deception can be identified in the same manner. Skyrms sees intentional and biological deception as similar: "Deception is widespread enough in human affairs, but it is not confined to our own species." (2010:73); so that we can use our intuitions about intentional deception to identify cases of biological deception. Likewise, to defend his claim that in a uniform population, no behaviour can be deceptive, McWhirter (2015:9) uses an analogy with the intentional case. Fallis (2015) explicitly aims for an analysis of deception that "does not require intentionality on the part of the sender and, thus, it can be applied to animal signaling as well as human signaling" (395). In other words, the idea that both humans and non-human animals can deceive is already present in the shared background of the literature on biological deception. As a consequence, an account that treats biological and non-biological deception separately or deals exclusively with one of them would not allow one to understand this formal and conceptual capillarity.

The third reason that motivates the extensional condition: it may be difficult to determine whether a case of deception is intentional or not, for instance when organisms such as primates are involved. As a result, deceptive behaviour among monkeys may be excluded by any account that sharply distinguishes between intentional and non-intentional deception – but not by accounts that encompass both cases. For these reasons, we think a theory that accommodates cases of intentional and non-intentional deception within a single framework should be preferred.

Finally, we would like to mention a widely accepted condition that we reject. There is a third claim that many endorse, namely that deception *has to* benefit the deceiver and harm the deceived. That is, many hold that when A deceives B, B must incur a cost and A must profit from this interaction. Let us call this condition the 'Harm-benefit Condition' (HbC). We reject this condition and will argue at length that a definition of deception should refrain from appealing to costs and benefits. Of course, it is hard to deny that there is a general tendency for deceivers to benefit from this interaction and for deceived agents to pay a cost (which might be regarded as a weak version of the Harm-benefit condition), but we reject the idea that deception necessarily requires a benefit for the sender and a cost for the receiver.

Whereas our proposal will *explain* why deception is usually associated with certain costs and benefits, one of the main theses of the paper that will be extensively argued for is that this tendency should not be mistaken for conceptual necessity.¹

In short, we think any satisfactory account of deception should accommodate the Error and the Extensional conditions, but should not meet the Harm-benefit condition (a claim that we will defend below). The next section shows that no current account of deception fits this portrait.

3. Theories of Deception

This section concentrates on an analysis of deception in biology, which mostly focuses on non-human animals. Nonetheless, it is useful to start with the human case, which is the paradigmatic model scientists use in order to understand deception in other organisms.

a. Deception in Humans

To a first approximation, human deception involves the production of a false belief (Fallis, 2010:5). For a subject A to deceive another subject B, it is necessary that B acquires a false belief (or fails to acquire a true belief). Nonetheless, this feature fails to provide a sufficient condition; for one thing, it does not distinguish between deception and error. A standard addendum is to require that, among other things, the deceiver must have some kind of complex intentional state (whence the label ‘intentional deception’, Hauser 1996). Thus, one of the central requirements in the standard approach to deception in humans is that a subject A must *intend* B to acquire a false belief (Carson 2009). A significant amount of discussion concerns the kind of propositional attitude (intention, belief...) that is required for intentional deception to occur (Mahon 2015). Unfortunately, definitions including complex intentional states are obviously inapplicable to many animals and for this reason fail to satisfy ExtC. When female fireflies of the species *Photuris* imitate the lightning pattern of the female of another species, they are clearly deceiving males while surely lacking complex cognitive states such as intentions or beliefs (Lloyd, 1975). Thus, current accounts of intentional deception face a difficult dilemma: if they do not appeal to intentions, they fail to meet ErrC; however, if they do, the resulting account is incompatible with ExtC. Most people interested in human deception choose the latter horn of the dilemma, and so give up the project of providing a common theory for human and non-human instances of deception.

b. Deception in non-human animals

The perspective adopted by biological approaches to deception differs from the previous view in two crucial respects. First, biologists replace intentional states and propositional attitudes by appeals to the organism's interests. Instead of intentionally trying to produce a false belief, it is said that deception requires an exchange that should be beneficial for the sender and detrimental for the receiver (usually measured in fitness). The second significant modification is that instead of producing a false belief, the sense in which deception involves false or misleading information is captured by the requirement that the signal itself must be inaccurate

¹ In the remainder of the paper, we will be explicit when we refer to this weak version; by default, talk of ‘Harm-Benefit Condition’ refers to the necessary version.

(Mitchell, 1986). Thus, deception² involves three different elements: a signal carrying false information, a sender benefiting from the interaction and the receiver performing a sub-optimal action. Proposals along these lines have been suggested by Hauser (1997), Maynard-Smith and Harper (2003), Searcy and Nowicki (2005), Semple and McComb (1996), Wilson and Angilletta (2015) and have become mainstream.³

Recent formal accounts of signalling have provided a more precise way of stating these ideas (Martinez 2015; McWhirter 2015). These approaches to deception use game-theoretical models, with a focus on signalling games. In the simplest versions of these models, two players, usually called 'sender' and 'receiver', face a set of situations ('states'). Each player gets a certain payoff depending on the response given to these states, which only the sender can observe. Often, the two players have common interests, represented by what Lewis (1969) called a 'coordination problem'. More formally, let $Q = \{q_1, q_2, \dots, q_n\}$ be a set of states of the world, $M = \{m_1, m_2, \dots, m_q\}$ a set of signals and $A = \{a_1, a_2, \dots, a_s\}$ a set of actions available to the receiver. Nature probabilistically chooses a state of world, observed by the sender. She then chooses a signal, observed by the receiver, who in turn chooses an action. A pure strategy for the sender is a function from states of nature onto signals, $S : Q \rightarrow M$ and a pure strategy for the receiver is a function from signals onto actions, $R : M \rightarrow A$. Sender and receiver payoffs are respectively given by the following functions: $\pi_s, \pi_r : Q \times A \rightarrow \mathbb{R}$.⁴ The set $\{Q, M, A, \pi_s, \pi_r\}$ defines a signalling game.

From this framework, which specifies the key elements of the simplest signalling system, several models can be built in order to define deception. Here we focus on Skyrms' (2010) influential account, based on information theory. First, Skyrms suggests the following measure for the amount of information I provided by a message m on a state of the world q:⁵

$$I(m, q) = \log_2 (P(q|m)/P(q))$$

In plain English, the amount of information of a signal on a state depends on how much the occurrence of the signal changes the probability of the state. The logarithm ensures that the amount of information is 0 if the message and the state of nature are independent (i.e. $P(q|M) = P(q)$, then $I(m, q) = 0$). Signals carry information about states of the world to the extent that they change their prior probabilities.

Second, Skyrms defines a concept of misinformation. A state carries misinformation just in case it moves the probabilities in the wrong direction, that is, if it either decreases the probability of an actual state of nature or increases the probability of a non-actual state. More

2 This set of cases is sometimes labelled 'functional deception'. However, given that we will defend a functional theory of deception (according to which all cases of deception are functional), we will sometimes refer to these cases of deception as non-intentional.

3 Some of these accounts (Searcy and Nowicki 2005, Fallis 2015) do not require the signal to be detrimental for the receiver, but still contain a built-in condition about a deceiver benefit, which makes them problematic, or so we will argue.

4 If messages are costly, these functions should also take signals into account, becoming $\pi_s, \pi_r : Q \times M \times A \rightarrow \mathbb{R}$.

5 This measure derives from Kullback and Leibler (1951).

formally, given an actual state q and a non-actual state q' , a signal carries misinformation if and only if either $I(m,q) < 0$ or $I(m,q') > 0$.⁶

With these conceptual tools, Skyrms' approach to deception can be reconstructed as follows (McWhirter 2015, 8):

(D) The use of signal M is deceptive if and only if:

1. The use of M carries misinformation;
2. Signals of type M are systematically sent to the benefit of the sender (or, more precisely, if $BR_s(x)$ refers to the highest payoff a receiver could get from state q and $R(M)$ is the receiver's strategy when M is received, $\pi_s(q,R(M)) > \pi_s(q,BR_r(q))$).
3. Signals of type M are systematically sent to the detriment of the receiver. (or $\pi_r(q,R(M)) < \pi_r(q,BR_r(q))$).

This spells out precisely the central ideas suggested by standard approaches to deception in the biological literature. Indeed, although different measures of, or variants on information have been provided (e.g. Lachmann and Bergstrom 2004, McWhirter 2015), the intuitions underlying Skyrms' model are widely shared (de Waal 1992, Wiley 1994, Martinez 2015). In short, deception appears to involve a misinformative signal, which benefits the sender and harms the receiver. Flashes emitted by females of the *Photuris* fireflies, for instance, are deceptive because they carry misinformation (increase the probability of a non-actual state, namely the presence of a female of another species), benefit the sender and harm the receiver.

How does D fare with the desiderata set up at the beginning? At first glance, D classifies as deceptive many intuitive cases of deception, such as the *Photuris* fireflies signals just mentioned. The fact that it seems to satisfy ExtC partly explains why it is the standard framework for studying deception. Furthermore, by appealing to the sender's benefit and the receiver's cost, it distinguishes between deception and error and thus meets ErrC.

However, we argue in the next sections that this definition, as similar ones, has two important drawbacks: first, its focus on signals causes it to fail ExtC; second, and more importantly, its explicit mention of sender benefits and receiver losses, which allows it to meet ErrC, also causes it to fail ExtC. We now consider these two points in turn.

c. Misinformative signals

The first problem with the previous definition is that, contrary to a widespread assumption (Hasson 1994), deception may not involve any signal. Mimics are a case in point. Flowers of the genus *Passiflora*, which are consumed by the herbivory larvae of Heliconius butterflies, mimic nearly hatched eggs, avoided by Heliconius (Williams and Gilbert 1981). Likewise, the Livingston cichlid of the Lake Malawi feigns to be dead in order to attract little scavengers, which he attacks when they come sufficiently close. Another example is the sea dragon (*Phyllopteryx eques*), an Australian sea-horse that resembles a sea weed (Ruxton et al.

6 Note that any signal that satisfies the first inequality will also satisfy the second one; since the total probability should be equal to 1, a decrease of the probability of the actual state q implies an increase of the probability of a non-actual state.

2004:23). Even humans employ many deceptive strategies that do not involve signals; just like piping plovers, we can deceive by limping. As these clear cases of deception fail to involve signals, this cannot be considered a requirement for deception to occur.

Indeed, the idea that such states are not signals can be motivated in various ways. For instance, take Maynard-Smith and Harper's (2003) widely embraced definition of signals as “any act or structure that alters the behavior of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved”. The previous examples fail to satisfy the last condition: the receiver has not evolved in order to give a precise response, a mimic being successful precisely when it fails to trigger an evolved response. Similarly, they do not seem to fall in any of the categories listed by Searcy & Nowicki (2005): alarm calls, food calls, begging, badges of status, mating signals or weapon displays.⁷ A further reason for denying their status as signals is that many theories assume that signaling usually requires a minimal level of common interest between sender and receiver, which is missing in these examples (Godfrey-Smith and Martinez, 2013; Artiga, 2014; Stegmann, 2009). Finally, interpreting these states as signals threatens to collapse the widespread distinction between signals and cues (Stegmann, 2013; Scott-Phillips and Kirby, 2013). Cues are states from which an organism can acquire information, but which have not evolved for that purpose. For instance, since the lengthening of daylight hours correlates with the beginning of spring, some plants use it as a cue for flowering, but this is not a signal. Likewise, a bird that spots a caterpillar whose ending mimics the face of a snake (such as the astonishing *Dynastor darius darius*) might wrongly infer that its prey is actually a potential predator, but it would be extremely counterintuitive to classify this protuberance as a signal. Not all deceptive states are signals, even if many certainly are.⁸ As a consequence, definitions of deception that presuppose a signalling framework fail to meet ExtC – they rule out cases that common sense as well as scientific intuition acknowledge as deceptive.

d. Costs and benefits

The second problem with D concerns its mention of costs and benefits. According to conditions 2 and 3 of (D), a deceptive signal should systematically benefit the sender and harm the receiver (which we called the 'Harm-benefit condition'). This choice of benefit and harm as systematic (rather than plausible, commonplace or widespread) is ubiquitous in the literature on deception. In the same vein, Mc Whirter appeals to misuse rather than misinformation, but otherwise retains Skyrms' conditions regarding systematic sender benefit and receiver harm (2015: 13). Semple & McComb define deception as a situation in which “the receiver benefits while the signaller pays a cost” (1996: 434). For Searcy & Nowicki, in addition to other conditions deception happens when “the receiver responds [to a signal] in a

7 Cases of organisms that withhold information may be interpreted as additional examples of deception without signalling. For instance, many avian and mammalian species perform a call when discovering food, but in certain occasions (e.g. when the food is unlikely to be discovered by others) an individual might not send the call (Hauser, 1997). However, such cases may be accommodated by D, by interpreting silence as a further signal. (Skyrms 2010).

8 Note that states that are not signals might still carry information (and misinformation) about the world. This follows, for instance, from Skyrms' proposal: a state carries information about another state iff it changes its prior probability, and this condition can be fulfilled by non-signals. However, note that this notion of information-carrying is extremely liberal. In this sense, for instance, any state carries information about its actual cause and carries misinformation about all other possible (non-actual) causes.

way that benefits the signaller” (2005: 5): although they drop the receiver harm condition, systematic benefit is still necessary. Similarly, Wilson & Angilletta hold that “dishonesty occurs when a receiver registers X from a signaller and responds in a way that not only benefits the signaller but would also benefit the receiver if X means Y, however, Y is false.” (2015: 206-7). In all these accounts, receiver cost and/or sender benefit have to be systematic for deception to occur; even those who drop one of the two conditions still assume fitness payoffs conditions to be conceptually necessary for the definition of deception.

Now, two caveats are important. First, note that the cost-benefit requirement is not simply an optional add-on that might be embraced or rejected without changing the core elements of the theory. Since biological approaches cannot define deception by appealing to the intentional states of agents (because most organisms lack sophisticated beliefs or intentions) they need to resort to costs and benefits in order to distinguish deception from mere error. For example, among the three conditions included in Skyrms' approach, only 1 does not appeal to costs and benefits, and it simply claims that the signal must carry misinformation. Thus, without 2 and 3, which basically state to harm-benefit requirement, deception would be defined as a misinformative signal, and in that case the theory would not satisfy the Error constraint. Consequently, the strong reading of the harm-benefit condition is not just a claim that most theories in fact endorse, but a core commitment that they have to accept if they want to distinguish deception from error.

Second, there is of course a difference between those who think that every case of deception has to harm the deceived and benefit the receiver and those who relax this demand and only require that it must fall within a pattern of situations in which these costs and benefits *systematically* occur. However, we think there are strong reasons against any form of harm-benefit condition, as long as it is a requirement for deception to occur. We now argue that explicit conditions about benefit and cost are not necessary for a trait to be deceptive.⁹

Some authors have already claimed that there may be deception even if the *target* incurs no cost. As seen above, Searcy & Nowicki (2005:5) only include a clause about the deceiver's benefit in their definition of deception; Fallis (2015) also insists that receiver costs are irrelevant. This is consistent with our intuitions regarding mimicking cases. The existence of a twig-like species of insects would only be costly for would-be predators if catching each prey involved some great effort. However, when other prey species are abundant, predators should not suffer from the failure to catch one of them (provided it is not nutritiously superior to other preys). Moreover, several empirical studies identify clear cases of deception without assessing the target cost. De Waal (1982) mentions the case of Yeroen, a chimpanzee (*Pan troglodytes*) in the colony of Arnhem Zoo, who had hurt his hand during an aggressive encounter with another male, Nikkie. After this incident, Yeroen limped, but only when Nikkie was around. Let us suppose that Nikkie saw this behaviour, which caused him to refrain from aggressing him. That would constitute a clear case of deception. Nonetheless, as Semple and McComb (1996, p. 436) acknowledge, we do not know whether Nikkie is actually paying any cost for not pressing his advantage further; for instance, by aggressing him, he could significantly increase his dominance with little effort. Independently of this

9 Several authors have recently criticised Skyrms' account as too liberal, because of its focus on misinformation; they identify intuitive cases of non-deceptive signals that Skyrms' definition would rule as deceptive (Martinez 2015, McWhirter 2015). As a result, they would probably see Skyrms' account as providing necessary (but not sufficient) conditions for deception. By contrast, if our arguments are correct the conditions put forward by Skyrms are not even necessary for deception.

particular point, the example shows that behaviours can be labelled as deceptive regardless of whether the target incurs any cost.

Deception may not involve any target cost. We intend to make a much stronger claim though, namely that deception can not only benefit the target, but could even harm the deceiver in principle. Again, the initial intuition comes from the case of intentional deception. Humans frequently tell white lies, i.e. falsehoods that benefit the receiver, which we learn to tell between 3 and 7 (Talwar and Lee 2002). In a signalling system framework, Fallis (2015) uses such cases to argue against the relevance of any receiver-related cost/benefit consideration for deception.

But intentional deception goes even further, as it can also harm the deceiver. In a study of white lies, Erat & Gneezy (2011) identify four kinds of lies. White lies benefit the target: ‘Pareto white lies’ are beneficial to the deceiver and ‘altruistic white lies’ detrimental. Black lies harm the target: ‘selfish black lies’ are beneficial to the deceiver and ‘spiteful black lies’ detrimental (2011:2). It seems that, in humans at least, one can find cases that fall under all these categories. Selfish black lies correspond to typical cases of deception, covered by Skyrms’ and others’ accounts. Pareto white lies fit the cases just discussed: falsehoods that benefit everyone involved. An example of a spiteful black lie could consist in overtly spreading a damaging but false rumour about someone else, risking to be perceived as a gossip or a schemer. Altruistic white lies may seem unlikely, but Erat & Gneezy’s study show that “a non-negligible fraction of participants are willing to tell an altruistic white lie that hurts them a bit but significantly helps others” (2011:1).¹⁰ Thus, human deception is compatible with any cost/benefit configuration.

Particularly interesting to us is the existence of altruistic and spiteful white lies, both of which harm the deceiver, because they constitute intuitive cases of deception that have been neglected in the biological literature. For biological cases, we will label deception as selfish, mutualistic, spiteful or altruistic (where mutualistic deception is analogous to Pareto white lies) in order to remain consistent with the vocabulary of social evolution (West et al. 2007). The usually discussed kind of biological deception is selfish deception. What about other kinds? We have not been able to find any real case of biological mutualistic, spiteful or altruistic deception in nature.¹¹ However, we can come up with possible cases, which is enough for our purpose.

First take a variant on *Dicrocoelium dendriticum*, the parasitic fluke famously mentioned by Dawkins, which during its life cycle forces its ant host to move to the top of a blade of grass in order for it to be ingested by a cow. Imagine a group of organisms that find and move dead ants so as to be similarly detected and eaten by cows; once in the cow, they multiply while contributing to the cow’s digestive system (ultimately leaving the cow’s body). Cows would profit from ingesting this on-nutritious dead ant. This would be a case of imperfect symbiosis (in which symbionts cannot be transmitted vertically), triggered by an initial episode of deception – mutualistic deception.

10 An extreme example of an altruistic white lie can be found in the movie “Love Me No More”, in which the terminably-ill main character, rather than to reveal his condition, chooses to start behaving obnoxiously with his family and friends, so that they are less affected when he dies.

11 We discuss possible reasons for this in section 5. At that point it suffices to say that this fact connects with the weak version of the Harm-Benefit Condition: since deceptive states tend to benefit the deceiver and harm the deceived, cases where this does not happen should be harder to spot.

Now consider members of a species (e.g. monkeys or birds) that perform alarm calls for certain predators. Imagine that a mutation enables some members to detect an elusive predator much more efficiently; they come to warn others by using an already existing alarm call. This call may allow the group to flee in a non-ideal way (adapted to another danger), although still preferable to any other possible reaction.¹² Moreover, as is often the case, the alarm call also makes the sender more likely to be spotted and targeted by the predator. Such behaviour would qualify as altruistic deception: it falsely indicates a danger, harms the sender and benefits the receivers. It could be adaptive as long as the fleeing mode, although imperfect, still increases chances of survival of relatives (for instance) so as to offset the individual's fitness loss.

That spiteful behaviours are rare to begin with makes spiteful deception all the more unlikely to exist in nature. However, it is possible. Red fire ants (*Solenopsis invicta*) are known for sometimes killing queens that lack a specific gene of theirs (Foster et al. 2001). This behaviour is spiteful because of the costs involved in the killing. Now imagine ants that would cause other workers to kill the queen, for instance by laying on her pheromones that normally signal an intruder (this pheromone-producing trait would be individually harmful, e.g. because the pheromone could "leak"). This does qualify as spiteful deception: deceptive (the queen is not an intruder), costly to the deceiver as well as to the receivers (the workers that kill their related queen). Just like spiteful (and altruistic) traits, it could and would have to evolve by kin selection (Foster et al. 2001).

Imaginary cases suffice to make our point, as they all feature traits that we want to call deceptive (just as we feel that white and black lies are lies). Encompassing them as cases of deception fulfils ExtC, because of the analogy with black and white lies in intentional deception. However, most existing definitions firmly exclude them, as they build in sender benefit and receiver harm. In other words, the traditional approach to deception assumed by biologists is inadequate. Although it satisfies ErrC, it fails to fulfil ExtC.¹³ In the next section, we provide an account that avoids these difficulties.

4. A Functional Theory of Deception

a. Deceptive states

The compatibility of deception with all combinations of deceiver and target interests dismisses the Harm-Benefit Condition, according to which it is *necessary* for deception to be

12 It may be objected that the example involves meaning change rather than deception: the meaning of the alarm call would now be a disjunction ('one of the two predators is close'). However, the fact that the fleeing behaviour is not adapted to the new predator, but merely preferable to standing still, mitigates this intuition. In Skyrms' (2010) vocabulary, this signalling system would feature a bottleneck: a signal used for two different states of the world, for which different actions would be preferable. And bottlenecks are misinformative – at least according to Skyrms' account.

13 Could these scenarios hold at equilibrium? This is a difficult question we cannot address here (see, for instance, Wagner 2014). Nonetheless, note that deception can obviously take place out of equilibrium (see below), so these cases would constitute counterexamples to standard definitions of deception, even if they were wiped out at equilibrium.

harmful to the target and beneficial to the deceiver. As a consequence, deception should be definable without any reference to the involved agents' interests.

A sensible worry, however, is whether an account of deception that does not appeal to costs and benefits can be satisfactory. McWhirter argues that “a sensible definition of deception should involve the interests of both the receiver and the sender.” (2015:6), because it allows one to meet the Error Condition – to differentiate deception from mere error. McWhirter discusses two cases that fail to meet ErrC because of their neglect of interests. First, some theorists have defined deception as “[occurring] when not all states of the world are uniquely identifiable from a signal” (Ibid.:5), which is compatible with error.¹⁴ Second, for Lachmann and Bergstrom (2004), “a message is deceptive if it has a negative value of information”, where the value of information actually depends on the receiver's interests. McWhirter objects that cases of error that are detrimental to the sender would then still count as deception, thus claiming that the sender's interests should then be considered in any definition of deception. The upshot is that definitions that fail to define deception in terms of costs and benefits may also fail to meet the Error Condition.

There is, however, an alternative way to meet the Error Condition. In the evolutionary literature, a classical way to distinguish the erroneous effects of traits from their ‘normal’ ones is to treat traits as functional. For to say that the function of a trait is to produce a certain effect is compatible with the trait sometimes failing to produce it. Accordingly, the remainder of this section develops a functional account of deception, which meets the Error¹⁵ and Extensional Conditions while rejecting the Harm-Benefit one (or so we will argue).

Our account assumes an etiological approach to function, according to which functions are reasons for existence. For instance, what explains the existence of many artefacts is the explicit intention of a designer. A screwdriver has the function of screwing because what explains its existence is the fact that someone intended it to be used in that way. This is true even if it has never been actually used. In the case of biological traits, intentions are replaced by natural selection (Dawkins' Blind Watchmaker). Here, the function of a trait is the effect that explains why evolutionary units bearing traits of this type were selected for. For instance, hearts have many effects like pumping blood or making thump-thump noises, but only the former accounts for its current existence. This is why the heart's function is to pump blood and not, for instance, to make thump-thump noises.

What should the function of deceptive traits be? This is where the concept of misinformation still plays a role. Even though we reject the view that deception necessarily involves misinformative *signals* (partly because some deceptive states are not even signals), the central idea of the functional account of deception is that deceptive states have the function of producing misinformation. In other words, misinformation is something that deceptive states have the function to produce, rather than something they carry.

14 McWhirter adds that this definition usually works for signalling models because sender benefit and receiver harm are built in.

15 Note that the Error Condition is distinct from what is called the problem of error in the context of defining propositional content. The latter is the problem of defining propositional content in such a way that a signal can have *false* propositional content (which lies beyond the scope of this essay). The former concerns the conceptual distinction between deceptive and erroneous states.

We can now provide our definition:¹⁶

(FD): M is a deceptive state iff

1/ M has the (etiological) function of causing a misinformative state (or failing to acquire a particular piece of information)

2/ M leads to a misinformative state.¹⁷

Several comments are in order. First, this definition appeals to misinformative states. As it is well known, providing an adequate account of misinformation is an extremely difficult task, which might even require solving the problem of meaning (Godfrey-Smith and Sterelny, 2016). Thus, addressing this question in detail would require a paper on its own. Nonetheless, we would like to stress two points. On the one hand, it is worth emphasizing that in principle any approach to information is compatible with our account. For instance, one could adopt Skyrms' (2010, p. 74-75) view, according to which a misinformative state is one that decreases the probability of the actual world state or increases the probability of a non-actual state. Alternatively, one could assume a teleological account and define misinformation in functional terms – e.g. that a state S is misinformative at t_1 iff (1) it has been produced by a mechanism whose function is to produce S when another state R obtains and (2) R does not obtain at t_1 . FD is fully compatible with these and other accounts of misinformation, and we take that to be a virtue of our approach. On the other hand, note that other theories also use the notion of misinformation and (crucially) its definition is taken to be independent of the other conditions included in the analysis of deception.¹⁸ In other words, in no account does a single condition take care of everything. Extant theories assume that the choice of the right concept of information is logically independent from the core conditions in the analysis of deception (harm-benefit, evolution, etc.). We agree with and follow this consensus.¹⁹

16 Note that FD is similar to Smith's (2014) recent proposal in the context of self-deception, with two significant differences. First, Smith's notion of function does not cover intentional cases (see footnote 11). Consequently, his approach does not satisfy the Extensional Condition. Second, his account is committed to teleosemantics, which, although a promising and increasingly popular theory of representation (Millikan, 1984, Shea 2007, Martinez 2013, Neander 2013), renders his account more specific than ours.

17 Mahon (2007, p. 185) argues that the false belief must also be caused in a normal way (e.g. inserting something into your brain might cause you to have a false representation, but it is not a case of deception). This is not just a philosophical quibble. Many parasites affect the nervous system of a host species in order to increase its susceptibility of predation and, in this way, reach their final host (Lafferty 1999). For instance, some members of the *Microphallus* species induce profound behavioural changes in their amphipod hosts, which makes them swim at the surface, rather than the bottom, of the water. A full theory of deception might need to take that into account.

18 For instance, in (D) (McWhirter's reconstruction of Skyrms' account, given in section 3.b), condition 1 concerns misinformation, while conditions 2 and 3 concern harm and benefit aspects. This isolation is also a feature of alternative definitions – for instance Searcy & Nowicki (2005: 5) and McWhirter's (2015) own.

19 It may be argued that the account risks being circular because misinformation itself may be defined in functional terms (for instance following Millikan's functional account of misrepresentation). However, our definition is compatible even with such a view. That a deceptive signal has the function of producing a misinformative state is consistent with a receiver's misinformation being in turn defined functionally. IN any cases, Authors such as Skyrms, McWhirter and Martinez for instance all adopt non-functional concepts of misinformation.

Second, condition 1 includes the failure to acquire a particular piece of information (for a discussion, see Fallis 2015), which captures cases in which animals deceive by withholding information (Kirkpatrick 2007, Hauser 1997). Third, condition 2 requires that the target is actually deceived (misinformed), according to the intuition that to deceive is a factive verb. Just as someone cannot be convinced without having acquired the intended belief, a genuine case of deception requires that the state's function be performed.²⁰

b. Conditions for deception

We now argue that FD fits the portrait established in section 2: it meets the Error and the Extensional Conditions, but rejects the Harm-Benefit Condition, which as we have seen is misguided.

b1. Error Condition

As demanded by the Error Condition, FD distinguishes deception from error, due to its appeal to etiological functions. When a tufted capuchin monkey performs an alarm call because he falsely believes there is a predator around, this call causes misinformation without having the function to do so. What explains the existence of these signals in the population is that often enough they carried information about predators; so their function is not to produce a misinformative state. (Furthermore, assuming that the monkey does not intend to mislead, a function to deceive intentionally cannot be ascribed either.) They have not been designed (either intentionally or by evolution) so as to produce misinformation.

b2. Extensional Condition

FD also meets the Extensional Condition. Fiddler crabs' display of the regenerated claws to their rivals has the function of misinforming the latter about the former's strength. The lightning pattern of *Photuris* fireflies has the function of causing *Photinus* males to represent the presence of a female of their own species. The shape and colour of passiform flowers derive from a historical process of selection, in which flowers possessing egg-like structures were favoured over other variants lacking them because they produced misinformation. On the etiological approach assumed here, they have this function because there is an evolutionary story that explains why these behaviours exist nowadays in the population.

FD also labels cases of camouflaging as deceptive. Several existing analyses of deception are sympathetic in principle to the inclusion of camouflaging (e.g. Mitchell 1986, Hasson 1994, Güzeldere et al. 2002). However, their requirement that deception necessarily involve signalling creates a tension. Debatable ways out include considering camouflage as a limiting case of signaling (e.g. Güzeldere et al. 2002), adopting an implausibly broad definition of signaling (e.g. Hasson 1994) or distinguishing between different kinds of deception (e.g. Mitchell, 1986).²¹ By contrast, FD accommodates camouflaging without ad hoc claims.

20 As an additional virtue, this approach easily accounts for cases in which there is a (non-successful) attempt to deceive; here condition 1 is satisfied, but not condition 2.

21 Furthermore, FD can account for different cases of camouflaging depending on the kind of misinformative state they involve. In crypsis, when the organism is not supposed to be detected, misinformation involves the presence of a prey. In masquerading, where an organism is detected but pretends to be something else, misinformation involves a miscategorization of the object. In disruptive coloration, misinformation involves the organism's form or shape, and in cases of motion dazzle the predator wrongly estimates speed and

Finally, FD also accounts for deception in humans and non-human animals. This is both because FD involves deceptive states (rather than only traits or behaviours) and because the notion of etiological function covers both intentional and non-intentional deception: both have the function of producing certain effects (Scott-Phillips and Kirby, 2013, p.431).²² For instance, Clinton's utterance of 'I did not have sexual relations with that woman' had the function to producing a false belief in the audience. He deceived those who believed his words. Here, the function does not derive from natural selection but from Clinton's intention. Thus, the etiological theory of function accommodates cases of deception across different domains.²³ Note that this characteristic is especially important for studying deception in organisms such as primates. Monkeys provide difficult cases for any analysis: on the one hand, intentional theories of deception, which rely on attributions of complex intentional states, probably fail because many primates lack them. On the other, standard approaches to deception are also inadequate, since they require systematic benefits and costs (which, among others things, preclude one-shot cases of deception). However, by appealing to the function of certain behaviours, all kinds of primates may be said to deceive, in the same sense in which the rest of natural world does.

In short, the FD account also meets the general Extensional Condition, because it includes many intuitive cases of deception ruled out by alternative accounts thanks. The two key features that help accommodate more cases are its appeal to deceptive *states* (instead of signals or traits), and to their *functions*.

b3. Harm-Benefit Condition

Regardless of what ones thinks of the benefits of a unified account of deception, the main interest of FD lies in its neglect of considerations about benefit and harm, which allows it to cover not only selfish but also mutualistic, altruistic and spiteful deception. More precisely, FD does not rule mutualistic, altruistic or spiteful deception as impossible by definition. Deceptive states may benefit or harm the deceiver or the target.

In the intentional case, it is enough that an agent intend to mislead or misinform another for this effect to be the function of her communicated message, which is consequently deceptive – whether it is a black or white lie, and more generally regardless of who it benefits or harms. In the non-intentional case, to say that a trait is deceptive implies that it has been selected for its misinformative effect, regardless of more specific fitness values. Depending on

trajectory (Stevens and Mirailta 2011: 5). Thus, misinformation may result from the misidentification of an object or from the misattribution of properties.

22 Smith's (2014) account also relies on the etiological approach to function. According to him (and following Millikan's 1984), for a trait to have a function it has to be a reproduction of a past item. Consequently, one-shot intentions cannot be said to warrant functions. Smith concludes that his approach supports non-intentionalist accounts of self-deception. By contrast, as said above, our more liberal etiological approach is compatible with both kinds of deception, which further fulfils the Extensional Condition.

23 Similarly, Searcy and Nowicki remark that "deception defined in this way has sometimes been termed 'functional deception' (Hauser 1997), meaning that the behavior has the effects of deception without necessarily having the cognitive underpinnings that we would require of deception in humans." (2005:5). This further reveals the appeal of a functional approach in the light of the Extensional Condition. However, their definition falls short of meeting the condition, as it still includes deceiver benefit.

background factors such as kin and population structure, natural selection may cause the evolution of traits that harm their bearer even if they benefit other agents; a fortiori, it may cause the evolution of traits in more favourable fitness configurations.

If all configurations of benefit and harm are possible for the deceiver and the deceived, then such harm/benefit configurations should not feature in conditions of the definition of deception, which is precisely what FD's appeal to functions allows. A functional definition is one way to avoid any explicit reference to harms and benefits.

Nonetheless, even if deception is not defined in terms of costs and benefits, FD can capture the weak version of the Harm-Benefit Condition, which claims that deception *tends* to benefit the deceiver and harm the deceived. According to FD, this would be the case if at least two additional assumptions are made: first, that false informational states tend to lead to suboptimal behaviour (or that, so to speak, true beliefs tend to make one fitter); second, that a trait typically acquires a function to produce certain effects when these tend to benefit the trait's bearer. When these assumptions are correct (which is, of course, an empirical matter) FD will entail that deception tends to benefit the sender and harm the receiver.

We suspect that the temptation to make deceiver benefit and target harm explicit in the definition of deception, has several possible motivations. The first motivation, namely the intuition that this is a necessary feature of deception, is unwarranted, as we have seen.

A second possible motivation is that stipulating that a given trait must be beneficial to its bearer and harmful to some other agents makes it easy to understand its evolution. If the bearer and these other agents are conspecific, then the trait automatically tends to make its bearer fitter than them. Even if they are not conspecific, that the bearer benefits from the trait would tend to make it fitter than organisms that occupy the same niche (i.e. are potential targets for the same predators). In other words, the harm/benefit configuration plays a double role for deception, as it participates both in its definition and its (evolutionary) explanation. This need not be a problem. However, this redundancy is not necessary. Our functional definition of deception dissociates definitional and explanatory aspects, by stipulating that deception must have been selected for (in non-intentional cases), regardless of how straightforward its evolutionary history is.²⁴

A third motivation to include deceiver benefit and target harm in the definition of deception is that in field studies, ethologists typically identify deceptive traits by relying on such conditions – they look for situations in which a false signal is sent, in such a way that the sender benefits and the receiver pays a cost. However, typical characteristics need be neither necessary nor sufficient. The mere fact that we had to rely on (biologically inspired) thought examples for mutualistic, altruistic and spiteful deception reveals that these cases should be rare in the biological world, if they exist at all. If the vast majority of cases of deception must involve deceiver benefit and target harm, these can, and even should, be used as proxies to for identifying deceptive traits. Still, to repeat, that the harm/benefit perspective is of heuristic value for identifying actual cases of deception explain its appeal without granting it any

24 This argument echoes Godfrey-Smith's (1994) point that it would be "vacuous to say that [a] trait persisted because some specific effect was its *function*." (p.354; original emphasis). This is because mentioning an etiological function presupposes that its effect has persisted. Similarly, it would be not vacuous, but redundant to add conditions on fitness that justify the persistence of a trait while calling this trait functional.

definitional worth. Indeed, we think the very same reasoning can justify the use of formal models of signalling (such as depicted in section 4) to study deception.

5. Objections to the functional approach

While it is clear that harm/benefit considerations should not feature a functional definition of deception (or so we have argued), at least two kinds of objections can be levelled against the appeal to functions in the non-intentional case. For a trait to be a function (in the non-intentional case), it needs to have been selected for during enough evolutionary time. A first kind of objection holds that evolutionary time is not necessary for non-intentional deception to exist. On the contrary, a second kind of objection claims that more time is necessary for functions than deceptive traits could ever enjoy.

The first kind of objection can be drawn from two accounts of the recent literature. Having reconstructed Skyrms' account of deception as described in section 3.a, McWhirter suggests a modification, targeting not the harm/benefit conditions but the misinformation one. McWhirter remarks that there could be misinformation, but not deception, in a population of uniform signallers. This would be the case when a signal is a half-truth – if it is sent when any of two states of the world is the case. Whenever emitted, it would raise the probability of the state that is not the case, thus being misinformative about it. Still, if this signal was used by the whole population, we would not want to call it deceptive, because “no one is using a message in a misleading manner in a uniform population. Receivers have learned to react to how the senders in the population use each message.” (McWhirter 2015:9). McWhirter follows up by redefining a signalling strategy as deceptive when a message is *misused* (plus sender benefit and receiver harm) – used differently from how the average population member does. The core idea is that a message can only be considered as deceptive if used in a different way than it normally is.

We will not comment on the replacement of misinformation by misuse – FD leaves the concept of misinformation underspecified so as to leave room for such amendments. However, note that on McWhirter's approach, deception can be defined by taking a snapshot of the population's signalling strategy at a given time, because it only depends on the typical use of a message. Nonetheless, the *causes* of misuse seem to be equally important. These could be intentions, or stem from natural selection. Deception builds on the existing meaning of a message (regardless of how meaning is defined), either intentionally or not. In particular, in the non-intentional case, deceptive traits must have built on pre-existing signalling strategies.

Imagine a sudden mutation simultaneously affecting a fraction of a population so that they start misusing an existing message. This would count as deception for McWhirter. But this should not, as the spread of the mutation does not stem from its effect. Rather, it is akin to a collective error. Note that in the above quote, McWhirter acknowledges such a historical dimension – deception would cease to exist if receivers learnt to detect it. Similarly, deception *starts* to exist when senders “learn” to be misinformative. As noted in the previous sections, a neglect of the historical dimension loses sight of the Error Condition. One could object that the sender benefit and receiver harm conditions make deceptive signalling strategies fitter and so ensure their future spread; but they do so by being unduly restrictive, at least if one deems

mutualistic, altruistic and spiteful deception possible. McWhirter's account is forward-looking at best; but non-intentional deception depends on the past.²⁵

A second kind of objection goes as follows. Functional traits are adaptations, which take time to spread and reach fixation in an entire population. But deceptive traits cannot be stable for long enough; deception and detection traits take part in an arm's race. Under constant selective pressure to surpass one another, they incur an endless sequence of modifications. In other words, deceptive traits are never stable enough to qualify as functional, or so the objection goes.

This is a tempting argument. First, some deceptive traits certainly never reach fixation in the whole population (for instance in cases of intraspecific deception, such as when males mimic females). If one follows McWhirter's intuition, maybe they never can.²⁶ Second, deceptive traits may never be stable. Godfrey-Smith (2011), for instance, argues that deception must be non-maintaining, that is, that if it reached a certain frequency in a population it would start being selected against.

However, we think that traits can be called functional (or adapted) even if they do not reach fixation or are negatively frequency-dependent. It is clear that deception can evolve when it is relatively non-frequent, so even if it is non-maintaining, it may always be present in a population. Consider the analogy with lies: maybe a society in which everyone lies would collapse; but a certain amount of lying seems to have been present in all human societies. Intraspecific deceptive traits, such as in male scorpion flies that mimic females, have not reached fixation; there is, however, no doubt that they are adaptations.

The reconciliation of adaptations with short evolutionary periods is not problematic in itself. Indeed, it is well known that many traits are adaptations in virtue of changes they underwent in a relatively recent past. This is the case with many so-called exaptations – traits that gain a new function after having had a first, different one (e.g. feathers used for flight after first being used to regulate body temperature). We think it is reasonable to adopt Godfrey-Smith's (1994) 'modern history theory of functions', according to which "functions are dispositions and powers which explain the recent maintenance of a trait in a selective context" (356). Even if there is no principled way of determining how recent the evolutionary history for which a function is relevant must be, this perspective shows that temporary or unstable traits can qualify as adaptations. Indeed, most existing traits are probably unstable when envisaged in a long enough evolutionary period. Moreover, specific contexts can lead traits to a certain

25 Birch (2014) suggest that we define the meaning of a signal by what it would have at the closest separating equilibrium of a signalling system ("an equilibrium at which there is a one-to-one mapping from states of the world to signals" (503). If 'closest' is interpreted as one to which the population will converge, this view, once applied to deceptive signals, would amount to another forward-looking account of deception (which is not Birch's focus). However, we could reply that deception also depends on what happened before the current state, and that one should add the condition that it has appeared after enough evolutionary time has passed. Just as it would be strange to grant that signals become meaningful immediately after their first appearance, so too traits need time to be considered deceptive.

26 Cases of interspecific deception may appear as counterexamples (in mimicking, all members of one species typically mimic something that helps them escape some predators). However, McWhirter only tackles interspecific deception (in which members of the same population exchange signals and anyone can be sender or receiver).

frequency within a population in a relatively short span of time. The functional definition of deception requires nothing more.

6. Conclusion

The main approaches to deception fail to satisfy key desiderata for any theory of deception – the Error and Extensional Conditions – while they mistakenly meet a third one – the Harm-Benefit Condition. We have presented a functional theory of deception that satisfies the former while rejecting the latter. It offers a unified approach to both intentional and non-intentional forms of deception. Crucially, it does not presuppose that deception must benefit the deceiver and harm its target, without abstracting away from considerations regarding benefit and loss: any benefit/loss combination that is compatible with successful misinformation is allowed. Biological deception can be mutualistic, altruistic and even spiteful, whether intentional or not.

Acknowledgements

We thank the members of the Munich Center for Mathematical Philosophy 2015 reading group on biological information and the audience of the 3rd Philosophy of Biology in the U.K. Conference (Bristol, 2016), the 21st Valencian Philosophy Conference (Castelló de la Plana, 2016) and the 4th Catalan Philosophy Conference (Vilafranca del Penedés, 2015). This research was partly supported by the Alexander von Humboldt Foundation (at the Munich Center for Mathematical Philosophy), the postdoctoral grant FPDI-2013-16764 and the project “La Complejidad de la Percepción: Un Enfoque Multidimensional” (FFI2014-51811-P).

References

- Artiga, M. (2014) Signaling Without Cooperation. *Biology and Philosophy* 29(3): 357-378
- Birch, J. (2014) Propositional Content in Signalling Systems. *Philosophical Studies* 171 (3): 493-512
- Carson, Th. (2010) *Lying and Deception*, Oxford University Press
- Erat, S. and Gneezy, U. (2011) White Lies, *Management Science, Articles in Advance*, pp. 1–11
- Fallis, D. (2010) Lying and Deception. *Philosophers' Imprint* 10 (11)
- Fallis, D. (2015) Skyrms on the Possibility of Universal Deception. *Philosophical Studies* 172 (2): 375-397
- Foster, K.R., Wenseleers, T. and Ratnieks, F.L.W. (2001) Spite: Hamilton's unproven theory. *Annales Zoologici Fennici* 38:229-238.
- Godfrey-Smith, P. (1994) A Modern History Theory of Functions. *Noûs* 28 (3):344-362.

- Godfrey-Smith, P. (2011). Signals: Evolution, learning & information, by Brian Skyrms. *Mind*, 120(480):1288–1297.
- Godfrey-Smith, P. and Martínez, M. (2013) Communication and Common Interest. *PLOS Computational Biology* 9 (11)
- Godfrey-Smith, P. (2014) Sender-Receiver Systems Within and Between Organisms. *Philosophy of Science* 81 (5):866-878
- Godfrey-Smith, P and K. Sterelny (2016) Biological Information. *Stanford Encyclopedia of Philosophy*.
- Güzeldere, G.; Nahmias, E, and Deaner, R. (2002) Darwin's continuum and the building blocks of deception. In M. Bekoff; C. Allen; G. Burghardt, (ed.), *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*. MIT Press
- Hasson, O. (1994) Cheating signals. *Journal of Theoretical Biology*, 167(3):223-238
- Hauser, M. (1997) Minding the behaviour of deception. In A. Wythen; R. Byrne (ed.) *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge University Press
- Hinton, H. E. (1976) Possible significance of the red patches of the female crab-spider. *misumenia vatia*. *Journal of Zoology*, 180:35-39,
- Von Hippel, W. and Trivers, R. (2011) The Evolution and Psychology of Self-Deception. *Behavioral and Brain Sciences* 34 (1):1
- Kirkpatrick, C. (2007) Tactical deception and the great apes: Insight into the question of theory of mind. *Totem*, 1:31-37
- Kight et al. (2013) Communication as information use: insights from statistical decision theory. *Animal Communication Theory*, p. 89-111
- Kullback, S.; Leibler, R.A. (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22 (1): 79–86
- Lachmann M, Bergstrom CT. (2004) The disadvantage of combinatorial communication. *Proceedings. Biological Sciences / the Royal Society*. 271: 2337-43.
- Lafferty, K.D. (1999) The Evolution of Trophic Transmission. *Parasitology Today* 15(3):111-115.
- Lewis, D. (1969) *Convention: A Philosophical Study*. John Wiley and Sons.
- Lloyd, J. (1975) Aggressive mimicry in photuris fireflies: signal repertoires by femmes fatales. *Science* 187(4175): 452–453
- Mahon, J. A. (2007) A Definition of Deceiving. *International Journal of Applied Philosophy*, 21(2):181-194

- Mahon, J. A. (2015) The Definition of Lying and deception. *Stanford Encyclopedia of Philosophy*
- Martinez, M. (2013) Teleosemantics and indeterminacy. *Dialectica*, 67(4):427-453
- Martínez, M. (2015) Deception in Sender–Receiver Games. *Erkenntnis* 80 (1):215-227.
- Maynard-Smith, J. (1991) Honest signalling: the philip sidney games. *Animal Behavior*, 42:1034-1035
- Maynard-Smith, J. and Harper, D. (2003) *Animal Signals*. Oxford Series in Ecology and Evolution
- McWhirter, G. (2015), Behavioural Deception and Formal Models of Communication, *British Journal for the Philosophy of Science*.
- Mitchell, R. (1986) A Framework for Discussing Deception. In R. Mitchell; N Thomson (ed.), *New Essays on Singular Thought*. SUNY Press
- Millikan, R. (1984). *Language, Thought and Other Biological Categories*. MIT Press.
- Neander, K. (2013) Toward an Informational Teleosemantics. In D. Ryder; J.Kingsbury; K. Williford, editor, *Millikan and her critics*. Wiley-Blackwell
- Platt, D. R. (1969) Natural history of the hognose snakes, *Heterodon platyrhinos* and *Heterodon nasicus*. University of Kansas Publications, Museum of Natural History 18:253–420.
- Ruxton, G., Sherratt, Th., Speed, M. (2004) *Avoiding Attack: The Evolutionary Ecology of Crypsis, Warning Signals and Mimicry*, Oxford University Press
- Scott-Phillips, T. and Kirby, S. (2013) Information, influence and inference in language evolution. In U. Stegmann (ed.) *Animal Communication Theory: Information and Influence*, pages 421-442. Cambridge University Press
- Searcy W. and Nowicki, S. (2005) *The Evolution of Animal Communication*. Princeton University Press
- Semple S. and McComb K (1996) Behavioural deception *Trends Ecol Evol*. 11(10):434-7.
- Shea, N. (2007) Consumers Need Information: Supplementing Teleosemantics with an Input Condition. *Philosophy and Phenomenological Research*, 75(2):404-435
- Skyrms, B. (2010) *Signals: Evolution, Learning, and Information*. Oxford University Press, Oxford
- Smith, D. (2014) Self-deception: A teleofunctional approach. *Philosophia*, 42(1):181-199,

Stegmann, U. (2009) A Consumer-Based Teleosemantics for Animal Signals. *Philosophy of Science*, 76 (5) 864-875

Stegmann, U. (2013) *Animal Communication Theory: Information and Influence*, Cambridge University Press (ed.)

Stevens, M. and Mirailta, S. (2011) *Animal Camouflage*, Cambridge University Press

Sterelny, K.; Joyce, R., Calcott, B. and Fraser B. (eds.) (2013). *Cooperation and its Evolution*, MIT Press.

Talwar V. and Lee, K. (2002) Emergence of White-Lie Telling in Children Between 3 and 7 Years of Age, *Merrill-Palmer Quarterly*, 48 (2)

Thornhill, R. (1979) Adaptive female-mimicking behavior in a scorpionfly. *Science*. 27; 205 (4404):412-4.

Trivers, R. (2011) *Deceit and Self-deception: Fooling Ourselves the Better to Fool Others*. London: Penguin

de Waal, F. (1982) *Chimpanzee Politics: Power and Sex among Apes*. The John Hopkins University Press

Wagner, E. (2012). Deterministic Chaos and the Evolution of Meaning. *British Journal for the Philosophy of Science* 63 (3):547-575.

Wagner, E. (2015) Conventional Semantic Meaning in Signalling Games with Conflicting Interests. *British Journal for the Philosophy of Science* 66 (4):751-773.

West, S.A., Griffin, A.S. and Gardner, A. (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology* 20(2):415-32.

Wheeler, B. C. (2009) Monkeys Crying Wolf? Tufted capuchin monkeys use anti-predator calls to usurp resources from conspecifics. *Proceedings of the Royal Society B* 276 (1669) 3013-3018

Williams, K. and Gilbert, L. (1981) Insects as Selective Agents on Plant Vegetative Morphology: Egg Mimicry Reduces Egg Laying by Butterflies. *Science*. 212(4493):467-469

Wilson R. and Angiletta, M. (2015) Dishonest Signaling during Aggressive Interactions: Theory and Empirical Evidence. In Duncan J. Irschick, M. Briffa, J. *Animal Signaling and Function: An Integrative Approach*, Wiley-Blackwell, pp. 205-227.