Noname manuscript No.
(will be inserted by the editor)

Toward a Formal Analysis of Deceptive Signaling

Don Fallis · Peter J. Lewis

the date of receipt and acceptance should be inserted later

Abstract Deception has long been an important topic in philosophy (see Augustine 1952; Kant 1996; Chisholm & Feehan 1977; Mahon 2007; Carson 2010). However, the traditional analysis of the concept, which requires that a deceiver *intentionally* cause her victim to have a *false belief*, rules out the possibility of much deception in the animal kingdom. Cognitively unsophisticated species, such as fireflies and butterflies, have simply *evolved* to mislead potential predators and/or prey. To capture such cases of "functional deception," several researchers (e.g., Sober 1994; Hauser 1997; Searcy & Nowicki 2005; Skyrms 2010) have endorsed the broader view that deception only requires that a deceiver *benefit* from sending a *misleading* signal. Moreover, in order to facilitate game-theoretic study of deception in the context of Lewisian sender-receiver games, Brian Skyrms has proposed an influential formal analysis of this view. Such formal analyses have the potential to enhance our philosophical understanding of deception in humans as well as animals. However, as we argue in this paper, Skyrms's analysis, as well as two recently proposed alternative analyses (viz., Godfrey-Smith 2011; McWhirter 2016), are seriously flawed and can lead us to draw unwarranted conclusions about deception.

Keywords Deceptive Signals · Functional Deception · Measures of Inaccuracy · Misinformation · Sender-Receiver Games · Signaling Theory

1 Introduction

Signaling systems arise in order to convey information. One example of a signaling system is human languages. But animals as well as humans engage in signaling behavior. For instance, prairie dogs give alarm calls to warn neighbors about predators in the vicinity, hawks give aggressive displays to alert

Address(es) of author(s) should be given

competitors to their willingness to fight, and peacocks grow elaborate tails to signal their quality to potential mates.

David Lewis's (1969, pp. 122–59) framework of *sender-receiver games* has turned out to be an extremely powerful tool for studying signaling systems (see Skyrms 2010). For instance, it has illuminated inquiries into how the transmission of meaningful signals can develop even amongst very simple organisms, how languages can change over time and become more complex, how signaling systems can facilitate cooperation amongst large groups of agents, and of most relevance here, how communication between competing agents can become deceptive.

In the simplest version of these games, there are two players, a sender and a receiver. The *receiver* has a choice to make among several possible courses of action. The outcome of his choice depends on what the world is like. However, the receiver does not know exactly what the world is like. But before the receiver has to make his choice, the *sender*, who does get to observe the true state of the world, can send a signal to the receiver.

In his work on sender-receiver games, Lewis focused on *coordination games*, where the interests of the sender and the receiver are perfectly aligned (see Skyrms 2010, p. 7). But in his book, Brian Skyrms (2010) also considers games in which the interests of the sender and the receiver conflict. In such games, the sender may have a motivation to *deceive* the receiver about the state of the world.

Deception has long been an important topic in philosophy (see Augustine 1952; Kant 1996; Chisholm & Feehan 1977; Mahon 2007; Carson 2010). However, the focus has been almost exclusively on human deception. But it is fairly clear that deception occurs in the animal kingdom as well (see Sober 1994, pp. 71–92; Hauser 1997; Searcy & Nowicki 2005). For instance, an animal sometimes gives an alarm call, not because she sees a predator, but in order to frighten off potential competitors for food or mates (see Skyrms 2010, pp. 73-74). Unfortunately, the traditional analysis of the concept, which requires that a deceiver *intentionally* cause her victim to have a *false belief*, rules out the possibility of much animal deception. For instance, cognitively unsophisticated species, such as fireflies and butterflies, have simply *evolved* to mislead potential predators and/or prey.

To capture such cases of "functional deception," several researchers (e.g., Sober 1994; Hauser 1997; Searcy & Nowicki 2005; Skyrms 2010) have endorsed the broader view that deception only requires that a deceiver *benefit* from sending a *misleading* signal. Moreover, in order to facilitate game-theoretic study of deception in the context of Lewisian sender-receiver games, Skyrms (2010, p. 80) has proposed an influential formal analysis of this view (see section 3 below). Such formal analyses have the potential to enhance our philosophical understanding of deception in humans as well as animals.

Skyrms (2010, pp. 80-82) has used his analysis to draw substantive conclusions about the nature and scope of deception. Most notably, Skyrms (2010, p. 81) claims to have refuted Immanuel Kant's (1996, p. 196) famous claim that universal deception is impossible. Several other philosophers (e.g., Ruse 1986,

p. 262; Sober 1994, p. 80) have already pointed out that universal deception is "logically consistent in the sense of involving no contradiction" (Skyrms 2010, p. 82). Skyrms (2010, p. 82) attempts to go further and show that universal deception can be "evolutionarily consistent in the sense of being an equilibrium." In addition, Skyrms's analysis of deception has been utilized by several other researchers (e.g., Wagner 2012, pp. 570–71; Smead 2014, p. 858; Bruner 2015, pp. 659–60; Martínez 2015, pp. 217–18).

We think that Skyrms is correct that a *deceptive signal* is a misleading signal that the sender benefits from sending. We show that it is possible to identify such signals in the framework of sender-receiver games (see section 6 below). And we argue that two non-Skyrmsian analyses (viz., Godfrey-Smith 2011; McWhirter 2016) that have recently been proposed are incorrect (see sections 7 and 8 below). But we also think that Skyrms makes several critical mistakes when filling in the details of his own analysis. As a result, Skyrms's formal analysis incorrectly counts many signals that are not deceptive (including the very example that was supposed to refute Kant) as being deceptive signals (see sections 4 and 5 below). Thus, his analysis can easily lead us to draw unwarranted conclusions about deception.

2 The Skyrmsian View of Deceptive Signals

Skyrms's formal analysis can be viewed as a generalization of the traditional philosophical analysis of deception. On the traditional analysis, you deceive someone if and only if you *intentionally* cause her to have a *false belief* (see Mahon 2007, pp. 189–90; Carson 2010, pp. 47–49). However, the traditional analysis is too restrictive in the context of human as well as animal communication. It does not capture all instances of deceptive signaling.

First, not all signals that convey information result in outright belief. A signal may just shift the probability of some hypothesis. For instance, even if a weather report does not cause us to believe that it will rain today, it might lead us to assign a higher credence to rain. The same applies to deceptive signals (see Fallis 2009, p. 45). For instance, even if a big bet does not fully convince your opponent at the poker table that you have a strong hand, it may create enough doubt in his mind that he folds.

Second, not all signals that convey information are intended by the sender to alter the belief state of the receiver. Many signals, such as the peacock's tail, have simply evolved to convey certain information. In fact, some animals send signals even though they are not cognitively sophisticated enough to have intentions with respect to the beliefs of other animals. The same applies to deceptive signals in particular (see Fallis 2015a, p. 383). For instance, many species of insects have developed an appearance that misleads potential predators and/or prey.

Accordingly, Skyrms generalizes the traditional analysis of deception in two ways. First, instead of requiring that the receiver end up with a false

belief, he only requires that the receiver be *misled* by the signal.[1] Even if she does not acquire a false belief, someone can be misled if her credences end up further from the truth. Skyrms is not alone in adopting this broader notion of misleadingness. Several other philosophers (e.g., Chisholm & Feehan 1977, p. 145; Fallis 2009, p. 45; Staffel 2011) have claimed that you can deceive someone just by shifting her credences.

But Skyrms goes even further than this. Skyrms (2010, p. 7) wants a theory that "accommodates signaling where no plausible account of mental life is available." For instance, as Skyrms (2010, p. 29–31) points out, even bacteria can send and receive signals.[2] Thus, he does not talk about how signals affect the credences (or *subjective* probabilities) of the receiver. Instead, Skyrms (2010, p. 80) simply talks about how "receipt of a signal moves probabilities of states."[3] In this paper, we follow his lead in this regard.

Second, instead of requiring that the sender intend to mislead the receiver, Skyrms only requires that the *sender benefit* from misleading the receiver. An analysis of deception must provide a non-arbitrary criterion that distinguishes deceptive signals from signals that are *merely* misleading. In particular, it must rule out cases where it is a *mistake* or a *mere accident* that the receiver is misled (see Skyrms 2010, p. 76; Fallis 2015a, p. 383; McWhirter 2016, p. 759; Artiga & Paternotte forthcoming, section 2). The "intentionality" requirement is what allows the traditional analysis to do this. If a misleading signal is sent intentionally, then it is not an accident. But most animal deception researchers (e.g., Sober 1994, pp. 73–74; Hauser 1997, p. 116; Searcy & Nowicki 2005, p. 5; Skyrms 2010, p. 80) use a "sender benefit" requirement instead to draw this distinction. If the sender benefits from sending a misleading signal, there is a mechanism (such as selection pressure) that reinforces the sending of the misleading signal. Thus, it is still no accident that the receiver is misled. In other words, the sender benefiting provides an *explanation* for why the misleading signal is not just a random occurrence.[4]

Marc Artiga and Cédric Paternotte (forthcoming, section 3.4) have recently described some (hypothetical) cases of deception in the animal kingdom where the sender herself does not benefit. However, as they note, deceptive signaling occurs in these cases because it "increases chances of survival of relatives (for instance) so as to offset the individual's fitness loss." In other words, it is the sender's *genes* that benefit from misleading the receiver. Skyrms (2010, p. 25)

[1] Skyrms (2010, p. 80) equates the terms *misleading information* and *misinformation*. We just use the first term here as it fits better with the philosophical literature on deception.

[2] We might even want to talk about deceiving simple machines as well as deceiving living creatures (see Lynch 2001, pp. 13-14).

[3] Even though they may not be the receiver's credences, these are the probabilities *from the standpoint* of the receiver. The probabilities of the possible states of the world from an objective standpoint (or from the standpoint of the sender who has observed the true state of the world) are all either zero or one.

[4] Ideally, we are able to analyze what a phenomenon is without having to explain why that phenomenon occurs. But deceptive signals are distinguished from merely misleading signals precisely because there is some explanation for why they are sent.

clearly wants to adopt a broad enough notion of sender benefit to capture such instances of kin selection. In this paper, we follow his lead in this regard.[5]

Thus, according to Skyrms, S sends a *deceptive signal* to R if and only if

- S sends M to R,
- R is misled by M,
- and S benefits from misleading R.

Many other animal deception researchers (e.g., Sober 1994, pp. 73–74; Hauser 1997, p. 116; Searcy & Nowicki 2005, p. 5) have also endorsed this sort of view. We think that this "Skyrmsian view" of deception is correct. The devil is in the details, however. In particular, what exactly does it mean for a receiver to be misled? Also, what exactly does it mean for a sender to benefit? In this paper, we object to the way that Skyrms's formal analysis fills in these details.

3 Skyrms's Formal Analysis

Skyrms (2010, p. 80) presents his formal analysis in the following way:

> if receipt of a signal moves probabilities of states it contains information about the state. If it moves the probability of a state in the wrong direction—either by diminishing the probability of the state in which it is sent, or raising the probability of a state other than the one in which it is sent—then it is misleading information, or *misinformation*. If misinformation is sent systematically and benefits the sender at the expense of the receiver, we will not shrink from following the biological literature in calling it *deception*.

To illustrate Skyrms's analysis in the context of sender-receiver games, consider one of the games that Skyrms (2010, p. 81) discusses in his book. The payoffs to both the sender (left) and the receiver (right) are determined by what state the world is in and what act the receiver chooses to perform. (A stands for "act" and S stands for "state.")

	A1	A2	A3
S1	2, 10	0, 0	10, 8
S2	0, 0	2, 10	10, 8
S3	0, 0	10, 10	0, 0

Table 1 Game 1

[5] Artiga and Paternotte also point to instances of intentional *human* deception where the sender does not benefit in any respect (see also Fallis 2015b, pp. 412–13). Since it eschews talk of intentionality in favor of talk of costs and benefits, the framework of sender-receiver games may not be able to capture such instances of deception. But humans typically do intend to mislead others because they benefit from others being misled (see Smith 2005). Thus, the "sender benefit" requirement applies to the vast majority of instances of human deception.

Note that this is not a game matrix. This is just the decision matrix for the receiver. The matrix does not show the signaling options that are open to the sender. (Although the sender knows what the true state is, she does not get to choose it.) Except in one case where we explicitly assume otherwise (see section 6), we assume here that she can send as many distinct signals as she likes. Also, following Skyrms, we assume that signaling itself is cost-free.[6]

In order to make this example more concrete, we might imagine that the senders are females of a very simple animal species, and that the receivers are males of this species. The females and the males both do better by mating than by not mating.

There are three types of female of this species: Terrestrial (S1), Amphibious (S2), and Purely Aquatic (S3). Also, there are three possible locations for mating: on Dry Land (A1), in the Water (A2), or on the Beach (A3). Terrestrial females can successfully mate with the males on dry land or on the beach. Amphibious females can mate on the beach or in the water. But Purely Aquatic females can only mate in the water.

When the female is Purely Aquatic, it is in the interest of both the female and the male to mate in the water (since that is the only option for successful mating). However, beyond that, the interests of the females and the males diverge somewhat. Terrestrial females and Amphibious females do better by mating on the beach. However, males do better by mating with Terrestrial females on dry land and with Amphibious females in the water.

Whenever a female and a male encounter each other, the female sends the male a signal regarding her type.[7] After receiving this signal, the male decides on a location for the mating activity. But since the interests of the females and the males are not perfectly aligned, it may be best for a female to send a signal that does not completely reveal her type, and that possibly even misleads the male about her type.

In analyzing such games, we will assume that there is an initial probability distribution over the possible states of the world. For the sake of simplicity, in this paper, we assume that the possible states of the world are equiprobable.[8] So, at the outset of each play of Game 1, the probability distribution over S1, S2, and S3 is (1/3, 1/3, 1/3). Thus, before getting any signal from the sender, A2 has the highest expected payoff (6.7).[9]

In this paper, for each game, we will look at one or more possible signaling systems. A signaling system consists of a sender strategy and a receiver strategy.[10] The sender's strategy tells her what signals to send in each state.

[6] These constraints are not essential to the framework (see Martínez 2015, pp. 223–27; McWhirter 2016, p. 760). Indeed, signaling costs play an important role in many explanations of honesty in animal signaling (see Searcy & Nowicki 2005, pp. 9–10).

[7] We assume that the location where the female and the male encounter each other does nothing to give away her type.

[8] Again, this constraint is not essential to the framework.

[9] Throughout this paper, we round expected payoffs to the nearest tenth.

[10] Lewis restricts the term *signaling system* to pairs of strategies that are *equilibria* of a game (see Skyrms 2010, p. 7). But as Skyrms (2010, p. 78) points out, "a lot of life is lived

In Game 1, one possible signaling strategy is for the sender to always send one signal (M1, where M stands for "message") in both S1 and S2 and to always send a different signal (M2) in S3:

S1 → M1
S2 → M1
S3 → M2

In other words, it is as if the Terrestrial and Amphibious females both say, "I am either Terrestrial or Amphibious," while the Purely Aquatic female says, "I am Purely Aquatic" (see Fallis 2015a, p. 379).

The receiver's strategy tells him what action to take on receipt of each signal that the sender might send. In this paper, we always assume that the receiver's strategy is his best response to the sender's strategy.[11] Basically, it is as if the receiver knows what the sender's strategy is. But the fact that the receiver's strategy is his best response could just be the result of evolution (see Skyrms 2010, pp. 63–72). Or in the case of more sophisticated species, it could be the result of learning over time (see Skyrms 2010, p. 93–105). In any event, it is the response that the receiver *ought* to have to the sender's strategy, at least in the long run.

So, if the sender adopts the aforementioned strategy, when the receiver gets M1, his new probability distribution is (1/2, 1/2, 0) and A3 has the highest expected payoff (8). Also, when the receiver gets M2, his new probability distribution is (0, 0, 1) and A2 has the highest expected payoff (10). Thus, the receiver's best response to the aforementioned strategy is to adopt the following strategy:

M1 → A3
M2 → A2

In other words, when a female says, "I am either Terrestrial or Amphibious," the male takes her to the beach to mate. And when a female says, "I am Purely Aquatic," the male takes her to the water. Call this pair of strategies *Signaling System 1.1*.

In contrast to Signaling System 1.1 where a specific signal is sent in each state, the sender might randomly choose which signal to send in a given state with a certain probability. For instance, in Game 1, the sender might adopt the following signaling strategy (where the number in parentheses is the probability of a given signal being sent in a given state):

S1 → M1(7/9), M2(2/9)
S2 → M2(7/9), M1(2/9)
S3 → M3

out of equilibrium ... deception is one of the forces that drive[s] the system to equilibrium." Thus, for purposes of analyzing deception, we will not adopt this restriction.

[11] This constraint is also not essential to the framework. It simply insures that our examples of deception do not depend on the receiver being ill-informed or behaving irrationally. After all, deception should be possible (e.g., at the poker table) even if the receiver knows that the sender is engaged in deception. Also, if the sender can do no better given the receiver's strategy, this constraint insures that the signaling system is an equilibrium.

There are a few different ways in which this mixed strategy might play out in terms of the story above. It could be that the Terrestrial females say, "I am Terrestrial" 77.8% of the time, but say, "I am Amphibious" 22.2% of the time (and similarly for the Amphibious females). Alternatively, it could be that 77.8% of the Terrestrial females always say, "I am Terrestrial," but 22.2% of them say, "I am Amphibious." In other words, mixed strategies might be instantiated by different proportions of the population of senders playing pure strategies.

When the receiver gets M1, his new probability distribution is (7/9, 2/9, 0) and A3 has the maximum expected payoff (8). Also, when the receiver gets M2, his new probability distribution is (2/9, 7/9, 0) and A3 has the maximum expected payoff (8). Finally, when the receiver gets M3, his new probability distribution is (0, 0, 1) and A2 has the highest expected payoff (10). Thus the receiver's best response to the sender's strategy is to adopt the following strategy:

M1 $\rightarrow$ A3

M2 $\rightarrow$ A3

M3 $\rightarrow$ A2

Call this pair of strategies *Signaling System 1.2*.

Given a formal analysis of deception and a signaling system, we can ask whether sending a particular signal in a particular state counts as deception. For instance, sending M1 in S2 in Signaling System 1.1 counts as deception according to Skyrms's analysis. First, on Skyrms's analysis, sending M1 in S2 is misleading. This signal causes the probability that the receiver assigns to a false state (S1) to go up from 1/3 to 1/2.

Second, on Skyrms's analysis, the sender benefits from sending M1 in S2. Although he is not explicit about this in the quote above, Skyrms determines whether a player benefits by comparing how well the player does with how well she would have done had the receiver known the true state with probability 1 (see Fallis 2015a, p. 382; McWhirter 2016, p. 764). And in this case, the payoff to the sender is higher than her payoff if the receiver knew that the true state was S2 (10 rather than 2).

Finally, it is important to note that Skyrms requires that the sender benefit in a particular way. The sender must benefit *at the expense of the receiver*. For instance, on Skyrms's analysis, the receiver does suffer a cost from receiving M1 in S2 in Signaling System 1.1. Skyrms uses the same baseline to determine whether a player has suffered a cost as he uses to determine whether a player has benefited. And in this case, the payoff to the receiver is lower than his payoff if he knew for sure that the true state was S2 (8 rather than 10).

Along the same lines, it can be shown that sending M1 in S2 in Signaling System 1.2 is also deception on Skyrms's formal analysis.

4 Skyrms on Misleadingness

While we agree with Skyrms that a deceptive signal is a misleading signal that the sender benefits from sending, Skyrms's formal analysis of this view is too broad. One reason that it is too broad is that his analysis of misleadingness is too broad. That is, it counts as misleading signals that are not actually misleading.

Skyrms's analysis of misleadingness is intuitively plausible. But it has some unfortunate features. Most notably, whenever there are three or more possible states of the world, it will often be the case that evidence that causes a shift from probability distribution **p** to probability distribution **q** is misleading *and* that evidence that causes a shift from **q** to **p** is misleading. For instance, if the first state S1 is the true state, evidence e_1 that causes a shift from (1/3, 1/3, 1/3) to (2/5, 2/5, 1/5) is misleading (since the probability of the false state S2 has increased). But evidence e_2 that causes a shift from (2/5, 2/5, 1/5) right back to (1/3, 1/3, 1/3) is also misleading on Skyrms's analysis (since the probability of the false state S3 has increased).

Now, this is not an actual contradiction. But it is certainly a very strange implication. And things get even stranger. Evidence e_3 that causes a shift from (2/5, 2/5, 1/5) to (1/2, 1/4, 1/4) is also misleading on Skyrms's analysis (since the probability of the false state S3 has increased). But evidence e_1 followed by evidence e_3, which causes an overall shift from (1/3, 1/3, 1/3) to (1/2, 1/4, 1/4), is *not* misleading. Indeed, it brings about clear epistemic improvement as the probability of the true state has increased and the probability of each false state has decreased. Thus, on Skyrms's analysis of misleadingness, two epistemic wrongs can make an epistemic right.[12]

The underlying problem is that Skyrms's analysis of misleadingness counts as misleading signals that are not actually misleading. In particular, contra Skyrms, an increase in the probability of a false state is not sufficient for an agent being misled (see Godfrey-Smith 2011, pp. 1294-95; Fallis 2015a, pp. 384–87). To see this, consider the following case:

A street performer hides a ball under one of three opaque cups (or *shells*). He then gives you the opportunity to pick a cup at random to turn over. If you choose to turn over the third cup, there are two things that could happen. First, you might see the ball. In that case, your probability distribution over possible locations of the ball shifts from (1/3, 1/3, 1/3) to (0, 0, 1). And you have definitely learned something about the location of the ball. You are epistemically better off.

Second, you might not see the ball. In that case, your probability distribution shifts from (1/3, 1/3, 1/3) to (1/2, 1/2, 0). You do not know exactly where the ball is. But you have still learned something about the location of

[12] It is certainly possible for two pieces of evidence to be individually misleading, but jointly beneficial. The way this generally works is that the two pieces of evidence are each misleading on their own, but the second piece is not misleading to someone who is already in possession of the first piece. However, in this example, evidence e_3 is misleading on Skyrms's analysis *even after* we have gotten evidence e_1.

the ball. In particular, you know at least one place where it is not. And definitively eliminating a false possibility is clearly epistemic progress, regardless of where the ball actually is (see Earman 1992, p. 163-85). Thus, you have not been misled. Admittedly, the probability of one of the false states does go up, but it goes up in the same ratio as the probability of the true state (see Fallis & Lewis 2016, pp. 582–83).

Note that the shift in the probability distribution in the shell game is exactly the same as the shift in the probability distribution when M1 is sent in S2 in Signaling System 1.1. Thus, we have to conclude, contrary to Skyrms's analysis of misleadingness, that sending M1 in S2 is not misleading.

Admittedly, the shell game is just one counterexample to Skyrms's analysis of misleadingness. It is not unique, however. Whenever a false state is eliminated, the probabilities of the remaining states (including the false ones) necessarily go up. In such cases, and as long as there are three or more possible states of the world, Skyrms's analysis gives the result that the receiver has been misled. But how could simply eliminating a false state ever take us further from the truth?

Since M1 sent in S2 in Signaling System 1.1 is not misleading, and since misleadingness is necessary for deception, it is also not an example of deception. Skyrms's (2010, pp. 81–82) proposed counterexample to Kant is a modified version of Game 1 in which *all* of the signals cause false states to be eliminated thereby increasing the probability of other false states (see Fallis 2015a). But as with M1 sent in S2 in Signaling System 1.1, the probability distribution is actually more accurate in each instance. So, none of the signals are misleading. Thus, Skyrms's proposed counterexample to Kant is not an example of deception, much less an example of universal deception.[13]

While his formal analysis of misleadingness is incorrect, Skyrms *is* correct that a misleading signal is one that causes the probability distribution over possible states of the world to move further away from the truth. Skyrms just picked an inappropriate measure of *inaccuracy*. Formal epistemologists have not yet determined what the correct measure is, or even if there is a uniquely correct measure (see Joyce 2009; Fallis & Lewis 2016).[14] But all of the measures of inaccuracy that formal epistemologists have proposed agree

[13] Shea et al. (forthcoming, section 4.4) also contend that sending M1 in S2 in Signaling System 1.1 is "merely a case of *strategic withholding* of information by the sender, a phenomenon quite distinct from deception." For the same reason, Justin Bruner's (2015, p. 660) proposed example of deception is not a case of deception.

[14] Skyrms (2010, p. 36) uses the *Kullback-Leibler divergence* to measure the informational distance between probability distributions. When the first probability distribution is the one that assigns probability 1 to the true state of the world and probability 0 to all of the false states, this is equivalent to using the *logarithmic rule* to measure the inaccuracy of the second probability distribution (see Godfrey-Smith 2011, p. 1293). So, it would make sense for Skyrms to appeal to this rule in his analysis of misleadingness, which would amount to saying that a signal is misleading if and only if it diminishes the probability of the state in which it is sent. Several formal epistemologists (e.g., Levinstein 2012; Roche & Shogenji forthcoming) have defended the logarithmic rule as a measure of inaccuracy. But many others (e.g., Joyce 2009; Pettigrew 2016) have defended the *Brier rule*. For purposes of this paper, we remain agnostic on the issue as nothing here hangs on the outcome of this debate.

that, when S2 is the true state, (1/3, 1/3, 1/3) is less accurate than (1/2, 1/2, 0) (see Fallis & Lewis 2016, pp. 581–82).

If we say that a signal is misleading if and only if it makes the probabilities less accurate, we avoid the aforementioned unfortunate features with Skyrms's analysis. Note that it is not just "elimination experiments" that are not misleading even though the probability of a false state increases. For instance, all of the measures of inaccuracy that formal epistemologists have proposed agree that, if S1 is the true state, (2/5, 2/5, 1/5) is more accurate than (1/3, 1/3, 1/3) and (1/2, 1/4, 1/4) is more accurate than (2/5, 2/5, 1/5). Thus, neither evidence e_1 nor e_3 above is misleading.

Admittedly though, once we correct Skyrms's analysis of misleadingness, there is at least one bullet that has to be bitten. According to Skyrms (2010, p. 75), "a female firefly of the genus Photuris, when she observes a male of the genus Photinus, may mimic the female signals of the males species, lure him in, and eat him. I would say that this qualifies as deception, wouldn't you?" However, when this sender-receiver game is at equilibrium and the mating signal is sent by a female Photuris, the shift in the probability distribution is essentially what we see in the shell game and when M1 is sent in S2 in Signaling System 1.1. It simply eliminates the false state that there is neither a Photinus nor a Photuris in the area. Thus, the mating signal is not misleading and, since misleadingness is necessary for deception, it is not deceptive.

Now, when this mimicry first began, the mating signal sent by a female Photuris did make the probability distribution less accurate. It made the presence of a female Photinus much more probable than the presence of a female Photuris. Thus, the mating signal was originally misleading. However, at equilibrium, "the mating signal being sent raises the probability of both kinds of partner, but leaves the ratio unchanged" (Skyrms 2010, p. 77). Thus, it is no longer misleading.

There are a couple of strategies that would allow us to avoid the conclusion that the mating signal sent by a female Photuris is not deceptive. First, we might adopt an analysis of misleadingness that does not require that the probability distribution become less accurate. But this would have the counterintuitive implication that you *have been* misled when you learn that the ball is not under the third cup. Also, it would not be in the spirit of Skyrms's claim that misleading information *moves* probabilities in the *wrong* direction.

Second, we might adopt an analysis of deception that does not require misleadingness. Indeed, some deception researchers (e.g., Chisholm & Feehan 1977, pp. 143–46; Bell & Whaley 1991; Hauser 1997, pp. 114–15; Fallis 2015a, pp. 387–90) have taken the position that deception only requires preventing someone from ending up epistemically better off. Thus, they would count the mating signal sent by a female Photuris as deceptive. However, most philosophers, including Kant and *Skyrms*, endorse some version of the traditional philosophical analysis of deception and think that misleadingness is necessary for deception (see Fallis 2015a, pp. 386–87). If you merely prevent her from ending up epistemically better off, you are "keeping someone in the dark"

(Carson 2010, p. 53) or "keeping that person ignorant" (Mahon 2007, p. 187) rather than deceiving her.

Although we agree with Skyrms that it *seems* deceptive, we contend that the mating signal sent by a female Photuris is not *actually* misleading or deceptive. As Skyrms (2010, p. 77) himself notes, "if you want to think of it as saying "I am the kind who sends this signal," you can think of it as telling the truth."[15] Of course, this is not to say that a female Photuris is not being sneaky. But not all sneaky behavior (such as failing to reveal the *whole* truth) counts as deception.[16]

5 Skyrms on Sender Benefit

But even if we correct his analysis of misleadingness, Skyrms's analysis of deception is still too broad. The reason is that his analysis of sender benefit, just like his analysis of misleadingness, is too broad. Even if the sender benefits on Skyrms's analysis from sending a misleading signal, it may just be an accident that the receiver is misled. Thus, Skyrms fails to distinguish between deceptive signals and misleading signals that are merely a random occurrence. In this section, we provide an example of this.

Even though sending M1 in S2 in Signaling System 1.1 is not misleading, there are signaling systems that definitely involve the sending of misleading signals. For instance, consider the following game:

	A1	A2	A3
S1	12, 10	0, 0	12, 6
S2	4, 0	0, 10	12, 6

Table 2 Game 2

Note that the story used to illustrate Game 1 above might be modified to capture this game as well. For instance, we would need to imagine that there are just Terrestrial females and Amphibious females. Also, in this case, the interests of the males and the Terrestrial females are fairly well aligned, but the interests of the males and the Amphibious females are not.

Once again, we assume that all of the states are equiprobable. Thus, at the outset of each play of the game, the receiver's initial probability distribution is (1/2, 1/2).

Now, suppose that the sender adopts the following signaling strategy:

S1 → M1(3/4), M3(1/4)

S2 → M1(3/8), M2(3/8), M3(1/4)

[15] Skyrms (2010, p. 77) does go on to say, "But it is only a half-truth." However, as Peter Godfrey-Smith (2011, p. 1295) points out, "to tell half the truth is not to tell a half-truth."

[16] Although it is possible to mislead someone by withholding information, it is not *always* misleading (see Carson 2010, pp. 56–57). Nevertheless, withholding information can be sneaky even when it merely prevents someone from ending up epistemically better off.

Given this signaling strategy, when the receiver gets M1, his new probability distribution is (2/3, 1/3) and A1 has the highest expected payoff (6.7). When the receiver gets M2, his new probability distribution is (0, 1) and A2 has the highest expected payoff (10). When the receiver gets M3, his new probability distribution is (1/2, 1/2) and A3 has the highest expected payoff (6). Thus, the receiver's best response to the sender's strategy is to adopt the following strategy:

M1 → A1
M2 → A2
M3 → A3

Call this pair of strategies *Signaling System 2.1*.

In Signaling System 2.1, M1 is sent twice as often in S1 than in S2. So, when she sends M1 in S2 (i.e., when an Amphibious female says, "I am Terrestrial"), the sender is sort of pretending that the true state is S1 when it is really S2. In any event, this signal causes the probability of the true state (S2) to go down (from 1/2 to 1/3). And it causes the probability of the false state (S1) to go up (from 1/2 to 2/3). This kind of shift in the probabilities is clearly sufficient for the receiver being misled. However we measure inaccuracy, when the probability of the true state goes down and the probability of the only false state goes up, the probability distribution is less accurate (see Fallis & Lewis 2016, p. 581).

In addition to actually misleading the receiver by sending M1 in S2, the sender benefits on Skyrms's analysis. The payoff to the sender is higher than her payoff if the receiver knew that the true state was S2 (4 rather than 0). Also, the payoff to the receiver is lower than his payoff if he knew that the true state was S2 (0 rather than 10). Nevertheless, we do not think that sending M1 in S2 in Signaling System 2.1 is necessarily an instance of deception. We think that it could be a mere accident that a misleading signal is sent in this case.

On the Skyrmsian view of deception, the sender benefiting is what is supposed to insure that it is no accident that a misleading signal is sent and, thus, that the receiver is misled. In the remainder of this section, we argue that, despite satisfying the conditions of Skyrms's formal analysis, the sender in Signaling System 2.1 need not benefit in a way that reinforces the sending of M1 in S2. In particular, we show that the sending of this misleading signal is not part of an equilibrium strategy, that it is not reinforced at the level of an individual signal, and that it is not necessarily reinforced as part of a full strategy. So, basically, we put the burden on Skyrms to identify some other respect in which the sending of M1 in S2 is guaranteed not to be mistake or a mere accident.

5.1 Part of an Equilibrium Strategy?

As mentioned in footnote 10, we do not want to analyze deception in such a way that it can only occur at equilibrium. But the fact that a certain sort

of behavior is part of an equilibrium is a very good explanation for why that behavior persists. Unfortunately, Signaling System 2.1 is not an equilibrium. The sender could unilaterally do better for herself by sending M3 more often in S2.

5.2 Reinforced at the Level of Individual Signals?

In general, revealing the whole truth will not be one of the sender's immediate options. For instance, in Signaling System 2.1, none of the available signals (M1, M2, or M3) will change the probability of S2 to 1.[17] But while sending M2 in S2 does not completely disclose the true state, it does change the probabilities so that the receiver acts in the same way that he would if he knew the true state for sure. Thus, Skyrms's analysis of sender benefit does explain why M1 is sent rather than M2 in S2.

However, Skyrms's analysis of sender benefit does not explain why M1 is sent at all in S2. There is another signal, M3, that could be sent and that would give the sender an even higher payoff (12 rather than 4). Moreover, since M3 is sent just as often in S1 as in S2, it does not change the probabilities. Thus, it does not mislead the receiver. So, it seems like an accident that M1 is sent in S2 and that the receiver is misled.

Now, even if sending M1 in S2 does not yield the highest possible payoff for the sender, there might still be selection pressure toward sending M1 in S2. Skyrms (2010, p. 54) standardly assumes that signaling systems evolve according to the *replicator dynamics*. That is, as the sender-receiver game is played multiple times, the probability of a particular signal being sent in a particular state increases (decreases) just in case the payoff for that signal in that state is greater than (less than) the average payoff.[18] And it could be the case that the probability of M1 being sent in S2 increases over time even though it does not yield the highest possible payoff.

However, as it happens, the probability of M1 being sent in S2 decreases over time under the replicator dynamics. In Signaling System 2.1 (and anywhere nearby in the space of possible signaling systems), the payoff for M1 in S2 is less than the average payoff (4 rather than 4.5). So, it still seems like an accident that the M1 is sent in S2 and that the receiver is misled.

[17] The sender might be able to cook up some completely new signal. But such a signal would not change the probabilities since signals have no meaning outside of the context of an overall signaling strategy (see below).

[18] Basically, we can think of the states as urns containing different colored balls representing the different possible signals (see Skyrms 2010, pp. 13–14). On each play of the game, nature chooses an urn and a ball is chosen at random from that urn in order to determine which signal is sent. Before the next play of the game, balls of the chosen color are added to (removed from) the urn, where the number of balls added (removed) depends on the payoffs.

5.3 Reinforced at the Level of Full Strategies?

So far, we have sought an explanation for why M1 is sent in S2 by focusing on the possible payoffs in S2. However, there is another way to go. Namely, we might appeal to the *expected* payoffs of adopting *full strategies* that involve sending misleading signals rather than the immediate payoffs for sending individual misleading signals (see Skyrms 2010, pp. 9–12).[19]

It actually makes a lot of sense to focus on full strategies. As Skyrms (2010, p. 8) emphasizes, the meaning of signals is *purely conventional*. As a result, a signal in isolation has no particular meaning. The information that a signal conveys to the receiver about the possible states of the world depends entirely on the full strategy of which it is a part. In particular, a signal is only misleading in the context of the full strategy. For instance, M1 sent in S2 is misleading because it is part of Signaling System 2.1. M1 sent in S2 is *not* a misleading signal if M1 is always sent in all states. In that case, it is simply uninformative. However, Skyrms's analysis of sender benefit does not guarantee that adopting a strategy that involves sending misleading signals is not just a mistake.

Skyrms's analysis of sender benefit *can* explain why the sender adopts the signaling strategy in Signaling System 2.1 *rather than* always revealing the whole truth. Consider the strategy of sending a distinct signal in each state in Game 2, say, M1 in S1 and M2 in S2. The receiver's best response to this strategy is to play A1 when he gets M1 and to play A2 when he gets M2. Call this pair of strategies *Signaling System 2.2*. The signaling strategy from Signaling System 2.1 has a higher expected payoff than the signaling strategy from Signaling System 2.2 (8.3 rather then 6). So, if the sender had been revealing the whole truth in her interactions with the receiver up until now, there would have been reason to adopt Signaling System 2.1 instead. Thus, sending M1 in S2 could be an instance of deception.[20]

However, the history might have been different. For instance, up until now, the sender might not have been conveying any information to the receiver. And

[19] In order to compute the expected payoff of the sender's strategy, an assumption must be made about how the receiver will respond. As noted in section 3, we assume that the receiver's strategy is his best response to the sender's strategy. If Skyrms's analysis does not guarantee that it is no accident that the signal is sent when the receiver's strategy is his best response, Skyrms's analysis does not guarantee that it is no accident that the signal is sent.

[20] Strangely enough, there are examples that satisfy Skyrms's formal analysis of deception, but where always revealing the whole truth has a *higher* expected payoff than the strategy that involves sending the misleading signal. For instance, suppose that the sender in Game 2 adopts the strategy:
S1 → M1(4/5), M2(1/5)
S2 → M2(1/2), M1(1/2)
And suppose that the receiver adopts the strategy:
M1 → A1
M2 → A2
In that case, sending M1 in S2 is deception on Skyrms's analysis, but the expected payoff to the sender of adopting this strategy is less than the expected payoff for always revealing the whole truth (5.8 rather than 6).

Skyrms's analysis of sender benefit *cannot* explain why the sender adopts the signaling strategy in Signaling System 2.1 *rather than* simply continuing to "keep her mouth shut." Consider the strategy of sending the same signal in all states in Game 2. The receivers best response to this strategy is to always play A3. Call this pair of strategies *Signaling System 2.3*. The signaling strategy from Signaling System 2.3 has a higher expected payoff than the signaling strategy from Signaling System 2.1 (12 rather then 8.3). Moreover, unlike Signaling System 2.1, Signaling System 2.3 is an equilibrium for the game and the sender always gets the maximum payoff (12). So, even if there were selection pressure toward sending M1 in S2, it would be washed out in the long run. Thus, even at the level of full strategies, Skyrms's analysis of sender benefit does not guarantee that it is not just an accident that the sender adopts a strategy that involves sending misleading signals.

So, even though sending M1 in S2 in Signaling System 2.1 is misleading and the sender benefits on Skyrms's analysis, there is no guarantee that it is not just a mistake that M1 is sent in S2 and that the receiver is misled. And this holds whether we focus on reinforcement at the level of individual signals or at the level of full strategies. So, it is not necessarily an instance of deception. And Skyrms's way of drawing the distinction between deceptive signals and signals that are merely misleading seems arbitrary.

Admittedly though, sending M1 in S2 in Signaling System 2.1 is just one counterexample to Skyrms's analysis of sender benefit. It could be that, in the majority of cases, when Skyrms's analysis says that a misleading signal is deceptive, it is. However, this would need to be shown. Also, if Skyrms were really getting at what deception is, there should not be *any* exceptions.

6 Two Examples of Deceptive Signaling

Skyrms's formal analysis of deception is too broad. It counts as deceptive signals that are not misleading. It also counts as deceptive signals that are accidentally misleading. Admittedly, Skyrms's analysis does have the advantage that it is fairly easy to determine whether sending a particular signal in a particular state counts as deception. But we do not want to be like the proverbial drunk who looks under a lamppost for his keys, even though he lost them elsewhere, simply because the light is better under the lamppost.[21]

Given that Skyrms's formal analysis is too broad, we are not justified in using it to make *existential* claims about deception. Most notably, Skyrms's own claim that universal deception is possible at equilibrium is unfounded. In the example that he offers in support of this claim, the receiver is not actually misled (see section 4).

But even though Skyrms's formal analysis is incorrect, that does not imply that the Skyrmsian view of deception from section 2 is incorrect. The view that a deceptive signal is a misleading signal that the sender benefits from sending is

[21] The story of the lost keys apparently derives from a story about the Sufi master, Mulla Nasrudin (see Shah 1966, p. 24).

still a plausible extension of the traditional analysis of deception and one that many animal deception researchers have endorsed. We should not abandon it just because Skyrms made some mistakes filling in the details. Furthermore, as we illustrate below, even without a fully worked-out formal analysis of deception, it is possible to identify clear examples of deceptive signaling in the framework of sender-receiver games using the Skyrmsian view.

6.1 The Signal is Misleading

While sending M1 in S2 in Signaling System 1.1 is not deceptive, we contend that sending M1 in S2 in Signaling System 1.2 *is*.[22] In other words, the signal is misleading and the sender benefits in a way that guarantees that it is no accident that the signal is sent. In order to see this, we first have to show that sending the signal is misleading.

In two-state cases, it is easy to tell that a signal is misleading even without an agreed upon measure of inaccuracy. If the probability of the true state goes down, the probability of the false state must go up. In that case, the new probability distribution is clearly less accurate. However, when there are three or more states, things are not as straightforward. It is not even clear that a signal that decreases the probability of the true state is necessarily misleading (see Fallis 2007, p. 234).

If the probability of the true state goes down, the probability of some false states must go up. But the probability of some other false states might go down, which is epistemically good. For instance, in Signaling System 1.2, when M1 is sent in S2, the probability of the true state goes down from 1/3 to 2/9. Also, the probability of the false state S1 goes up from 1/3 to 7/9. But in addition, the probability assigned to the false state S3 goes down from 1/3 to 0. And it is not immediately clear whether the epistemic costs outweigh this epistemic benefit (i.e., whether the new probability distribution as a whole is less accurate).

Even so, there is good reason to think that $\mathbf{r} = (7/9, 2/9, 0)$ is less accurate than $\mathbf{s} = (1/3, 1/3, 1/3)$ when S2 is the true state. First, note that $\mathbf{r}$ is clearly less accurate than $\mathbf{q} = (2/3, 1/3, 0)$ when S2 is the true state. $\mathbf{q}$ assigns a higher probability to the true state and $\mathbf{q}$ assigns at least as low a probability to all of the false states (see Fallis & Lewis 2016, p. 581).

Second, all of the measures of inaccuracy that formal epistemologists have proposed agree that $\mathbf{s}$ is at least as accurate as $\mathbf{q}$. $\mathbf{s}$ and $\mathbf{q}$ assign the same probability to the true state. They just differ with respect to how the overall probability assigned to false states is distributed over the false states. Now, in formal epistemology, there is a debate about whether such "falsity distributions" affect the accuracy of the overall probability distribution (see Fallis & Lewis 2016, p. 580). If falsity distributions do not affect overall accuracy, then $\mathbf{s}$ is exactly as accurate as $\mathbf{q}$. And if they do affect overall accuracy then

[22] In this case, Skyrms's formal analysis gets the right result (albeit for the wrong reasons).

s is *more* accurate than **q** since it is clearly better not to have more of the probability assigned to false states piled onto a particular false state.[23] Thus, either way, **s** is at least as accurate as **q**.

Finally, since **s** is at least as accurate as **q**, and **r** is less accurate than **q**, it follows from the transitivity of inaccuracy that **r** is less accurate than **s**. Thus, sending M1 in S2 in Signaling System 1.2 is misleading.

6.2 The Sender Benefits from Sending the Signal

Next, we have to show that the sender benefits from sending M1 in S2 in Signaling System 1.2 in a way that guarantees that it is no accident that the signal is sent. In this regard, note that Signaling System 1.2 is actually an equilibrium of Game 1.[24] No player can unilaterally do any better for herself. And as noted above, the fact that a certain sort of behavior is part of an equilibrium is a very good explanation for why that behavior persists.

Moreover, by adopting the signaling strategy from Signaling System 1.2, the sender gets the maximum possible payoff (10) *in every state*. So, the expected payoff of this strategy is at least as high as the expected payoff of any other strategy that the sender might adopt.[25] And it is higher than the expected payoff of many strategies, such as the strategy of always revealing the whole truth (10 rather than 4.7). Thus, the sender does at least as well for herself by adopting this strategy no matter what strategy she previously adopted.

Admittedly, the probability of M1 being sent in S2 does not increase over time under the replicator dynamics. But it does not decrease either. In Signaling System 2.1 (and anywhere nearby in the space of possible signaling systems), the payoff for M1 in S2 is equal to the average payoff (10).

6.3 Possible Objections

There are many equilibria of Game 1. So, one might worry that it is an accident that the system reaches the specific equilibrium of Signaling System 1.2. Moreover, some of the equilibria of this game do not involve sending misleading signals. For instance, Signaling System 1.1 is also an equilibrium of this game and (as discussed in section 4) all of its signals are informative. So, it might be suggested that it *is* an accident that misleading signals are sent.

[23] According to the standard measures of inaccuracy that do take into account falsity distributions, such as the Brier rule and the spherical rule, symmetric distributions are (*ceteris paribus*) more accurate.

[24] Not every signal sent at this equilibrium is deceptive. For instance, M2 sent in S2 is not even misleading. So, it is still not the example of *universal* deception that Skyrms (2010, p. 81) was looking for.

[25] Thus, there is a reason for the sender to adopt this strategy regardless of whether the receiver's strategy is his best response.

However, a large number of the possible equilibria of this game *do* involve sending misleading signals. For instance, consider the family of signaling systems where $1/5 < x, y < 4/5$ and $x/y > 2$:

S1 $\rightarrow$ M1(x), M2($1 - x$)
S2 $\rightarrow$ M2($1 - y$), M1(y)
S3 $\rightarrow$ M3
M1 $\rightarrow$ A3
M2 $\rightarrow$ A3
M3 $\rightarrow$ A2

All of these signaling systems are equilibria and sending M1 in S2 is misleading.

One might also worry that nothing has yet been said about whether the sender benefits *at the expense of the receiver* by sending M1 in S2 in Signaling System 1.2. In fact, she does on Skyrms's analysis. The payoff to the receiver is lower than his payoff if he knew that the true state was S2 (8 rather than 10).[26] However, we do not think that the receiver suffering a cost is actually required for deception.

Although some researchers in animal signaling (e.g., Hauser 1997, p. 116, Skyrms 2010, p. 80) include an "expense to the receiver" requirement in their analyses of deception, others (e.g., Searcy & Nowicki 2005, p. 5) do not.[27] And there are actually good reasons not to. First, an "expense to the receiver" requirement rules out the possibility of *altruistic* deception (see Fallis 2015a, p. 391).[28] We often deceive someone else for that person's own good. For instance, we tell "white lies" in order to save someone from embarrassment or from upsetting news.

Second, and most importantly, an "expense to the receiver" requirement does nothing to help explain why a misleading signal is sent (see Fallis 2015a, p. 391). Without an "intentionality" requirement in the analysis of deception, we need a "sender benefit" requirement in order to insure that it is no accident that a misleading signal is sent. But the receiver suffering a cost does not contribute to this explanation in any way.

6.4 Deception Outside of Equilibrium

We have now seen an example of deception at equilibrium. But deception can also occur outside of equilibrium. For instance, consider the following game:

[26] Similarly, in our second example of deception below, the payoff to the receiver is lower than his payoff if he knew the true state (0 rather than 10).

[27] Skyrms (2010, p. 76) himself is actually ambivalent on the issue. He writes that "one could argue over whether the clause about the detriment of the receiver should be included ... I do not think that much hangs on the choice."

[28] It might seem like the "sender benefit" requirement itself rules out the possibility of altruistic deception. But even though altruism requires benefiting someone else *at a cost to oneself*, the standard explanation for such behavior is that it benefits one's genes (see Skyrms 2010, p. 25). And as noted in section 2, we and Skyrms include benefits to the sender's genes as part of sender benefit.

	A1	A2	A3
S1	10, 10	0, 0	2, 2
S2	5, 0	10, 10	2, 2
S3	0, 0	0, 0	10, 10

Table 3 Game 3

As usual, we assume that all of the states are equiprobable. So, at the outset of each play of the game, the receiver's initial probability distribution is (1/3, 1/3, 1/3).

Note that, once again, we can interpret this game in terms of Terrestrial (S1), Amphibian (S2), and Purely Aquatic (S3) females. But in this case, given the payoffs, the females would clearly all prefer to reveal their identities to the males. That would yield the maximum possible payoff to the sender (10) in every state.

However, suppose that, at the moment, the females only have two signals available.[29] And suppose that the sender adopts the following signaling strategy:

S1 → M1
S2 → M2
S3 → M2

In other words, it is as if the Terrestrial female says, "I am Terrestrial," while the Amphibious and Purely Aquatic females both say, "I am either Amphibious or Purely Aquatic." The receiver's best response to this strategy is to adopt the following strategy:

M1 → A1
M2 → A3

Call this pair of strategies *Signaling System 3.1*.

In this case, while the Terrestrial and Purely Aquatic females still prefer to be truthful about their identities, there is actually pressure for the Amphibious female to lie sometimes and claim to be Terrestrial. The payoff for M1 in S2 is greater than the average payoff (5 rather than 2). Thus, under the replicator dynamics, the probability of M1 being sent in S2 will slowly increase over time. For instance, at some point, the sender's strategy will be:

S1 → M1
S2 → M2(5/7), M1(2/7)
S3 → M2

The receiver's best response to this strategy will still be to play A1 when he gets M1 and to play A3 when he gets M2.[30] Call this pair of strategies *Signaling System 3.2*.

Of course, Signaling System 3.2 is not an equilibrium of the game either. Given the receiver's strategy, the sender could do better by sending M1 even

[29] The sender may not have had time yet to develop the capacity to send more than two different signals.

[30] The male may not initially realize that the Amphibious female lies sometimes. But even once he does catch on, it will not lead him to alter his strategy.

more often in S2. Under the replicator dynamics, the sender will ultimately end up *always* playing M1 in S2:

S1 → M1
S2 → M1
S3 → M2

It is now as if the Terrestrial and Amphibious females both say, "I am either Terrestrial or Amphibious," while the Purely Aquatic female says, "I am Purely Aquatic." And all the way to this point, the receiver's best response remains the same (play A1 in response to M1 and play A3 in response to M2). Call this pair of strategies *Signaling System 3.3*.

Much like Signaling System 1.1, Signaling Systems 3.1 and 3.3 do not involve any misleading signals. But when the receiver gets M1 in Signaling System 3.2, his new probability distribution is (7/9, 2/9, 0). And as we have already seen, when S2 is the true state, (7/9, 2/9, 0) is less accurate than (1/3, 1/3, 1/3). So, M1 sent in S2 is misleading.

Moreover, it is no accident that M1 is sent in S2 in Signaling System 3.2 and that the receiver is misled. There is no other signal that could be sent and that would give the sender a higher payoff.[31] And more importantly, the probability of M1 being sent in S2 increases over time under the replicator dynamics. In Signaling System 3.2 (and anywhere nearby in the space of possible signaling systems), the payoff for M1 in S2 is greater than the average payoff (5 rather than 2.9).[32] Thus, despite not being part of an equilibrium, sending M1 in S2 in Signaling System 3.2 is deceptive.[33]

In conclusion, even if a deceptive signal is a misleading signal that the sender benefits from sending, it may be quite difficult to operationalize the concept. First, as discussed in section 4, formal epistemologists are not yet in agreement about how to measure inaccuracy. So, it is not clear how to give necessary and sufficient conditions for when a signal is misleading. Second, there are several different (and not necessarily overlapping) ways in which a sender might benefit from sending a signal. So, it is not yet clear how to give necessary and sufficient conditions for when it is no accident that a misleading signal is sent.

Nevertheless, as the two examples discussed in this section illustrate, it is often possible to identify instances of deceptive signaling using the Skyrmsian

[31] The sender might be able to cook up some completely new signal. But since such a signal would not have any meaning for the receiver, the probabilities would not change. In that case, the receiver would play A3 and the sender would still get a lower payoff (2 rather than 5).

[32] Moreover, sending M1 in S2 in Signaling System 3.2 is reinforced at the level of full strategies if the sender had previously adopted the signaling strategy from Signaling System 3.1. The signaling strategy from Signaling System 3.2 has a higher expected payoff than the signaling strategy from Signaling System 3.1 (7.6 rather then 7.3).

[33] Even though sending M1 in S2 in Signaling System 3.2 is misleading and it is no accident that M1 is sent in S2, this example does not count as deception on Skyrms's analysis. The payoff to the sender is *lower* than her payoff if the receiver knew that the true state was S2 (5 rather than 10). So, this example indicates that Skyrms's analysis is too narrow (as well as being too broad) and, thus, that it is not safe to use Skyrms's analysis to make *universal* claims about deception.

view. After all, in most cases, the measures of inaccuracy that formal epistemologists have proposed will agree that a shift in probabilities is misleading. Also, in many cases, it will be clear that the sender benefits in a way that makes it no accident that a signal is sent. For instance, the signal might be part of an equilibrium strategy and, even if it is not, the evolutionary dynamics might tend to increase the probability that the signal is sent.

7 Godfrey-Smith on Non-Maintainingness

A couple of other philosophers have also recently criticized Skyrms's formal analysis of deception. But they have gone even further and rejected the Skyrmsian view as well.[34] For instance, Peter Godfrey-Smith (2011, p. 1295) analyzes deception in terms of a distinction between "the maintaining and the non-maintaining uses of the signal." According to Godfrey-Smith, "some uses contribute to stabilization of the sender-receiver configuration and some, if more common, would undermine it." He claims that the former (i.e., the maintaining uses) are not deceptive whereas the latter (i.e., the non-maintaining uses) are deceptive.

Here is a fairly obvious way to cash out Godfrey-Smith's notion of non-maintainingness in the context of sender-receiver games:[35] Consider a signal M, a state S, and a signaling system Z. And let A be the receiver's response to M in Z.[36] We then imagine modifying the sender's strategy by increasing the probability that M is sent in S. If we can find such a modification where the receiver has a better response to M than A, then sending M in S in Z is a *non-maintaining use*.

[34] Shea et al. (forthcoming, section 4.5) offer yet another non-Skyrmsian analysis of deception. But the only example of "*bone fide* deception" that they provide also counts as deception on the Skyrmsian view. The receiver is misled (as sending M1 in S2 shifts the probabilities from (1/4, 3/4) to (1/2, 1/2)) and the sender benefits (as sending M1 in S2 is part of an equilibrium strategy).

[35] Manolo Martínez (2015, p. 219–21) suggests a different way of cashing out the notion of non-maintainingness. He claims that sending signal M in state S is a non-maintaining use if and only if the receiver would shift from a separating strategy to a pooling strategy if M were always sent in S. That is, the receiver starts out responding differently to some signals and, as the probability of M being sent in S increases, ends up responding the same way to all signals. However, this analysis seems to give the wrong result in at least some cases. For instance, Martínez describes a signaling system in which he claims that sending M1 in S2 is a non-maintaining use. But on his own analysis, it is *not* a non-maintaining use. It is true that, when the probability that M1 is sent in S2 is equal to 1, the receiver always performs the same act. But that does not mean that the receiver has adopted a pooling strategy. In the original signaling system, the sender always sends M1 in the other states. Thus, when the probability that M1 is sent in S2 increases to 1, the only signal that the receiver ever gets is M1. But the receiver *would* do something different if he ever did receive some other signal. It is not the case that the receiver "stops listening" to what the sender says. In contrast, the analysis that we suggest in the text gives the correct result that sending M1 in S2 in this signaling system is a non-maintaining use.

[36] Just as the sender might randomly choose which signal to send with a certain probability in each state, the receiver might randomly choose which act to perform with a certain probability in response to each signal. Thus, A might be a mixture rather than a pure act.

Godfrey-Smith's proposal makes a lot of sense. Non-maintainingness frequently goes hand-in-hand with deception. As J. L. Mackie (1977, p. 183) points out, if you engage in deception too often, it is likely to cease to be an effective strategy at some point (see also Smith 2005, p. 20).

However, we think that non-maintainingness is neither necessary nor sufficient for deception. First, as noted above, sending M1 in S2 in Signaling System 1.2 and in Signaling System 3.2 are pretty clearly instances of deception. The receiver is misled and the sender benefits in a way that makes it no accident that the signal is sent. And misleadingness together with such benefit to the sender seems to be sufficient for deception. But sending M1 in S2 in Signaling System 1.2 is not a non-maintaining use. Even if the probability of M1 being sent in S2 increases all the way to 1, A3 continues to have the highest expected payoff (8). Thus, the receiver continues to perform A3 if he gets M1. Similarly, sending M1 in S2 in Signaling System 3.2 is clearly not a non-maintaining use. Thus, non-maintainingness is not necessary for deception.[37]

We can also use Signaling System 1.2 to see that non-maintainingness is not sufficient for deception. Sending M2 in S2 is clearly not an instance of deception. The receiver is not even misled. When he gets M2, the probabilities shift from (1/3, 1/3, 1/3) to (2/9, 7/9, 0). Since the probability of the true state (S2) goes up and the probability of both false states goes down, the receiver has not been misled. Thus, as misleadingness is necessary for deception, this is not an instance of deception. But sending M2 in S2 *is* a non-maintaining use. In Signaling System 1.2, the receiver always performs A3 if he gets M2. However, if the probability of M2 being sent in S2 increases (beyond 8/9), he reverts to performing A2 (which is what he would have done if he got no information from the sender in this game). In other words, the receiver "stops listening" to what the sender says.

It might be suggested that our analysis of non-maintainingness does not quite capture the idea of the receiver ignoring what the sender says. Instead of just requiring that the receiver do something different as the probability that the signal is sent increases, maybe we should require specifically that the receiver goes back to doing what he would have done if he had gotten no signal from the sender. As noted above, however, sending M2 in S2 in Signaling System 1.2 satisfies this more stringent condition. So, it is still an example of a non-maintaining use that is not an instance of deception.

8 McWhirter on Misuse

The analyses of deception that we have considered so far apply even if there is only a single sender, or if there is a homogenous population of senders all

[37] Martínez (2015, p. 221–23) also claims to have found a signaling system that involves "deception without non-maintaining uses of signals." However, we think that sending M3 in S1 in this signaling system *is* a non-maintaining use. When the probability that M3 is sent in S1 increases to 1, the receiver does not adopt a pooling strategy. But he does abandon his previous response to M3 and always performs A2 when he gets M3.

using the same strategy. However, according to Gregory McWhirter (2016, p. 767), deception arises when an individual sender benefits from using a signal in a way that diverges from the way that the population of senders as a whole uses it.

Suppose that the receiver does not know what strategy a particular sender has adopted.[38] He only knows what the *average sender's* strategy is. As a result, the receiver's strategy is his best response to the average sender's strategy. In that case, a *rogue sender* can potentially fool the receiver by "misusing" a signal (i.e., by diverging from the average sender's strategy). In particular, she can get him to underestimate the chances of the true state and to overestimate the chances of one or more of the false states.

According to McWhirter (2016, p. 768), sender G sending signal M in state S in signaling system Z counts as misuse if the following two conditions hold: First, the probability of the true state is higher when the signal comes from G than when it comes from an average member of the population. In other words, Pr(S | M from G) > Pr(S | M from the population). Second, the probability of some false state is higher when the signal comes from an average member of the population than when it comes from G. In other words, there is a state S' not equal to S such that Pr(S' | M from the population) > Pr(S' | M from G).

As with Godfrey-Smith's analysis of deception, McWhirter's analysis makes a lot of sense. Misuse is often closely associated with deception. For instance, if most people say, "I am innocent," or "I love you," only when these things are true, a rogue who says these things (to his own advantage) when they are not true seems to be engaged in deception.

However, we think that McWhirter's analysis is neither necessary nor sufficient for deception. First, suppose that a population of senders *all* adopt the signaling strategy from Signaling System 1.2. Thus, there can be no misuse. But as noted above, when a sender sends M1 in S2 in Signaling System 1.2, it is pretty clearly an instance of deception. The receiver is misled and the sender benefits in a way that makes it no accident that the signal is sent. Thus, misuse is not necessary for deception.[39]

Moreover, McWhirter's analysis is not sufficient for deception. Although McWhirter gets rid of the misleadingness requirement of the Skyrmsian view, he keeps the sender benefit requirement. Indeed, McWhirter takes his analysis of sender benefit directly from Skyrms's analysis. So, McWhirter's analysis of deception is going to be too broad for essentially the same reason that Skyrms's analysis is too broad. That is, a sender can benefit from misusing a signal on his analysis when it is really an accident that signal is sent.

But even if we were to correct his analysis of sender benefit, McWhirter's analysis of deception would still be too broad. A rogue sender can misuse a

[38] In order for deception to occur on McWhirter's analysis, there must be two or more senders. Similarly, there might be two or more receivers. Just for the sake of simplicity, we assume here that there is a single receiver.

[39] Similarly, M1 sent in S2 in Signaling System 3.2 shows that misuse is not necessary for deception.

signal without misleading the receiver. For instance, suppose that, when a population of senders sends a signal M1, it causes the probabilities to shift from (1/3, 1/3, 1/3) to (2/5, 2/5, 1/5). Also, suppose that a rogue sender adopts the following signaling strategy:

S1 → M1
S2 → M1
S3 → M2

When the rogue sender sends M1 in S1, it is misuse. The probability of the true state is higher when the signal comes from the rogue sender than when it comes from an average member of the population:

Pr(S1 | M1 from rogue) = 1/2
Pr(S1 | M1 from population) = 2/5

Also, the probability of some false state is lower when the signal comes from the rogue sender than when it comes from an average member of the population:

Pr(S3 | M1 from rogue) = 0
Pr(S3 | M1 from population) = 1/5

However, as discussed in section 4, when S1 is the true state, (2/5, 2/5, 1/5) is clearly more accurate than (1/3, 1/3, 1/3). Thus, M1 sent in S1 is not misleading. In fact, it is informative. So, it would be a stretch to call it deception even if the rogue sender were to benefit from sending this signal.

9 Conclusion

In broad strokes, Skyrms is right about deception (and Godfrey-Smith and McWhirter are wrong). A deceptive signal *is* a misleading signal that the sender benefits from sending. But Skyrms gets the details wrong. In developing a formal analysis of deception, we need to utilize an appropriate measure of inaccuracy and an appropriate measure of sender benefit. Otherwise, we are liable to draw erroneous conclusions about deceptive signaling in human and animal communication.[40]

References

1. Artiga, M., & Paternotte, C. (forthcoming). Deception: A functional account. *Philosophical Studies*.
2. Augustine. (1952). *Treatises on various subjects*. New York: Fathers of the Church.
3. Bell, J. B., & Whaley, B. (1991). *Cheating and deception*. New Brunswick: Transaction Publishers.
4. Bruner, J. P. (2015). Disclosure and information transfer in signaling games. *Philosophy of Science* 82, 649-66.
5. Carson, T. L. (2010). *Lying and deception*. New York: Oxford University Press.

[40] For extremely helpful feedback on earlier versions of this material, we would like to thank Jeff Barrett, Justin Bruner, Terry Horgan, Kay Mathiesen, Brian Skyrms, Rory Smead, Eyal Tal, Dan Zelinski, two anonymous referees, and audiences at the Freedom Center, University of Arizona, the School of Information, University of Arizona, and the Department of Logic and Philosophy of Science, University of California, Irvine.

6. Chisholm, R. M., & Feehan, T. D. (1977). The intent to deceive. *Journal of Philosophy* 74, 143-59.
7. Earman, J. (1992). *Bayes or bust?* Cambridge: MIT Press.
8. Fallis, D. (2007). Attitudes toward epistemic risk and the value of experiments. *Studia Logica* 86, 215-46.
9. Fallis, D. (2009). What is lying? *Journal of Philosophy* 106, 29-56.
10. Fallis, D. (2015a). Skyrms on the possibility of universal deception. *Philosophical Studies* 172, 375-97.
11. Fallis, D. (2015b). What is disinformation? *Library Trends* 63, 401-26.
12. Fallis, D., & Lewis, P. J. (2016). The Brier rule is not a good measure of epistemic utility (and other useful facts about epistemic betterness). *Australasian Journal of Philosophy* 94, 576-90.
13. Godfrey-Smith, P. (2011). Review of *Signals: Evolution, learning, and information* by Brian Skyrms. *Mind* 120, 1288-97.
14. Hauser, M. D. (1997). Minding the behaviour of deception. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II* (pp. 112–43). Cambridge: Cambridge University Press.
15. Joyce, J. M. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber and C. Schmidt-Petri (Eds.), *Degrees of belief* (pp. 263–97). Dordrecht: Springer.
16. Kant, I. (1996). *Practical philosophy*. Cambridge: Cambridge University Press.
17. Levinstein, B. A. (2012). Leitgeb and Pettigrew on accuracy and updating. *Philosophy of Science* 79, 413-24.
18. Lewis, D. (1969). *Convention*. Cambridge: Harvard University Press.
19. Lynch, C. A. (2001). When documents deceive: Trust and provenance as new factors for information retrieval in a tangled web. *Journal of the American Society for Information Science and Technology* 52, 12-17.
20. Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin Books.
21. Mahon, J. E. (2007). A definition of deceiving. *International Journal of Applied Philosophy* 21, 181-94.
22. Martínez, M. (2015). Deception in sender-receiver games. *Erkenntnis* 80, 215-27.
23. McWhirter, G. (2016). Behavioural deception and formal models of communication. *British Journal for the Philosophy of Science* 67, 757-80.
24. Pettigrew, R. (2016). *Accuracy and the laws of credence*. Oxford University Press.
25. Roche, W., & Shogenji, T. (forthcoming). Information and inaccuracy. *British Journal for the Philosophy of Science*.
26. Ruse, M. (1986). *Taking Darwin seriously*. New York: Basil Blackwell.
27. Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication*. Princeton: Princeton University Press.
28. Shah, I. (1966). *The exploits of the incomparable Mulla Nasrudin*. New York: Simon and Schuster.
29. Shea, N., Godfrey-Smith, P., & Cao, R. (forthcoming). Content in simple signalling systems. *British Journal for the Philosophy of Science*.
30. Skyrms, B. (2010). *Signals*. Oxford: Oxford University Press.
31. Smead, R. (2014). Deception and the evolution of plasticity. *Philosophy of Science* 81, 852-65.
32. Smith, D. L. (2005). Natural-born liars. *Scientific American Mind* 16, 16-23.
33. Sober, E. (1994). *From a biological point of view*. Cambridge: Cambridge University Press.
34. Staffel, J. (2011). Reply to Sorensen, 'knowledge-lies'. *Analysis* 71, 300-2.
35. Wagner, E. O. (2012). Deterministic chaos and the evolution of meaning. *British Journal for the Philosophy of Science* 63, 547-575.