Deception in Sender-Receiver Games

Manolo Martínez

mail@manolomartinez.net; manolo.martinez@uab.cat

Logos – Logic, Language and Cognition Research Group

Universitat Autònoma de Barcelona

Barcelona (Spain)

1. Sender-Receiver Games

In a *sender-receiver game* (also, interchangeably, *signalling game*; Lewis 1969), the *sender* observes in which one of a set *S* of jointly exhaustive and mutually exclusive *states* the world is and then, following its *sender rule*, sends one of a set *M* of *messages* (also, interchangeably, *signals*) to the *receiver*. Upon receipt of the message, the receiver follows the *receiver rule* to do one of a set *A* of *acts*.

There is a *payoff* associated to each triple of state, message and act. Such payoffs may differ partially, completely, or not at all between sender and receiver. Payoffs can be summarised in a *payoff matrix* such as the one shown in Table 1.

	Act 1	**Act 2**	**Act 3**
State 1	2,10	0,0	10,8
State 2	0,0	2,10	10,8
State 3	0,0	10,10	0,0

Table 1: A payoff matrix. At any cell, c_{ij}, the first number is the payoff for the sender of the combination of state S_i and act A_j . The second number is the payoff for the receiver. In this game, payoffs only depend on the combination of state and act – there's no cost associated with the production of signals. From Skyrms (2010), p. 81.

A sender rule is characterised by a function from *S* to the set of probability distributions over *M*. An example of sender rule might be:

- $S_1 \rightarrow M_2(1/2), M_3(1/2)$—This is an abbreviation of “If the world is in S_1 send M_2 with probability 0.5 and M_3 with probability 0.5”.
- $S_2 \rightarrow M_1(1)$
- $S_3 \rightarrow M_1(1/5), M_3(4/5)$

A receiver rule is characterised by a similar function, from *M* to the set of probability distributions over *A*.

Given a particular combination of sender and receiver rules and the unconditional probabilities of states in *S*, we can calculate an average payoff for sender and receiver. A *Nash equilibrium* (also, simply *equilibrium* henceforth) in a sender-receiver game is a combination of sender and receiver rules such that neither sender nor receiver can increase their average payoff by changing their rule unilaterally.

Sender-receiver games have been receiving increasing attention among philosophers of mind since Skyrms's seminal 1996 book. One reason for this is that the wholly non-intentional goings-on in such games appear to provide insight into the nature of intentional properties such as *having meaning*: at equilibrium it is often the case that signals carry information about the state the world is in, and also that such information is used by the receiver in guiding its actions. This combination of properties of the signals exchanged in a sender-receiver game – the fact that they carry information about the world, and their role in the production of behaviour – makes them plausible precursors of ulterior fully intentional vehicles of communication.

Communication, in sender-receiver games as in real life, is often jeopardised by the possibility of *deception*. Suppose that the state of the world S_j is one in which the sender has a certain desirable quality; for example, it is a suitable mating partner for the receiver (Johnstone 1997) or a good candidate for a job offered by the receiver (Spence 1973). If the world is, indeed, in S_j, it is in the interest of both parties that the sender let the receiver know that this is the case and, for example, use signal M_i to send information about S_j, so that the receiver then might use M_i as a cue to produce S_j-adequate behaviour – such as mating with, or hiring, the sender. On the other hand, once sender and receiver have established a link between M_i and S_j – once, that is, M_i "means" that the world is in S_j – a sender lacking the desirable quality will be tempted to fake its way into mating or the job by sending M_i regardless of the state the world is in. This seems to lead to a situation in which the sender sends M_i no matter what, and the informational content of this signal (and, therefore, the incentive that the receiver has in letting its actions be guided by the signal) disappears: communication breaks down.

The possibility of deception has figured prominently in the study of communication in sender-receiver games. One influential approach to this problem has been to suggest that the temptation to deceive can be kept in check if there is a cost associated with the production of certain signals: perhaps it is only profitable to send M_i for senders in S_j (e.g., only fit male peacocks can afford the exuberant display characteristic of this animal – see Zahavi 1975); perhaps faking is costly because, say, the sender needs to spend resources in making it seem that it is high-quality when it is not – "lemon market" games, such as the one presented in section 6 of this paper, are like this. Such *costly signalling* (a relatively recent survey is Searcy and Nowicki 2005) can prevent liars from destabilising communication. In the absence of signalling costs, communication can be preserved if the interests of sender and receiver (i.e., which acts they prefer in which states) are sufficiently aligned[1] (Crawford and Sobel 1982, Skyrms 2010).

The foregoing informal presentation of the problem of deception is typical of the literature on

[1] As it turns out, a remarkably imperfect alignment of interests can be sufficient for communication. See Godfrey-Smith et al. (2013), Wagner (2012), Zollman et al. (2013)

this topic in that I have assumed that liars are *subversive*, in the following sense: they exploit a communicative agreement for their own interest and, if they take this behaviour too far, the agreement can break down. While this subversive aspect of deception is undoubtedly of great importance – indeed arguably the aspect that gives deception such a central place in the discussion of signalling games – one might wonder whether subversiveness is, moreover, *necessary* for some behaviour to count as deceptive.

In a recent critical piece, Peter Godfrey-Smith has argued that Skyrms's (2010) purely informational characterisation of deception is unsatisfactory: many signals which Skyrms would count as deceptive should, in fact, be read as communicating (albeit partial) truths. Godfrey-Smith suggests that deceptive signals always have uses which "if more common, would undermine [the sender-receiver configuration]" (Godfrey-Smith, 2011 p. 1295) – *non-maintaining* uses. In this paper I defend Skyrms's liberal notion of deception from Godfrey-Smith's claim that it is essentially subversive.

After quickly reviewing Skyrms's treatment of deception (section 2) and Godfrey-Smith's objection to this treatment (section 3), in section 4 I present a game that, *contra* Godfrey-Smith, has signals that count as deceptive according to Skyrms, but such that no partial truth is communicated by them. The discussion in this section motivates a certain game-theoretic regimentation of the notion of non-maintaining signal, which is then used, in section 5, to show that it is possible to have signalling games in which certain uses of signals are deceptive but not non-maintaining. Section 5 discusses, first, a concrete example of one such game, and then reports on the outcome of a more systematic exploration of the space of games with three states, three signals and three acts, and the prevalence of not non-maintaining, deceptive signals in that space.

The discussion up to and including section 5 follows Skyrms and Godfrey-Smith in focusing on "cheap talk" cases: sender and receiver payoffs depend only on the state the world is in and the act the receiver does, not on the signal sent. It might be wondered whether the possibility of not non-maintaining deceptive signals extends to cases of costly signalling (in which the sender payoff does depend on the type of signal) – given, in particular, as I have explained above, that costly signalling is generally taken to be less accommodating to deception than cheap talk. Section 6 adjudicates this issue by providing an example of a game with costly signals, with a relatively natural biological interpretation, in which some of the signals sent are deceptive but not non-maintaining.

The conclusion will be that, while the importance of subversive cases of deception is not in question, the claim that deception is essentially subversive apparently cannot be sustained. Skyrms's purely informational treatment of this notion is, thus, vindicated.

2. Information and Deception

A message M_j in a sender-receiver game *carries information* about a certain state S_i whenever the probability of the state, conditional on the presence of the message, is different from its unconditional probability: $P(S_i|M_j) \neq P(S_i)$ (Skyrms 2010, p. 35f). For example, the following combination of sender and receiver rules is a Nash equilibrium in the game presented in Table 1 (assuming that all three world states are equiprobable):

Sender Rule:

- $S_1 \rightarrow M_1$
- $S_2 \rightarrow M_1$
- $S_3 \rightarrow M_2$

Receiver Rule:

- $M_1 \rightarrow A_3$
- $M_2 \rightarrow A_2$
- $M_3 \rightarrow A_2$ – This one will never be used.

What is going on here? On the one hand, the sender is not distinguishing S_1 and S_2 with its signalling behaviour. It is, on the other hand, letting the receiver know whenever they are in S_3. This is explained by the fact that, in S_3, they agree on which act is best (A_2), while they disagree about this on S_1 and S_2. The receiver, as a consequence, can deliver A_2 in S_3, but is forced in S_1 and S_2 to chose an act, A_3, that is the best for the sender, but only the second best for itself. Intuitively, then, the sender is manipulating the receiver to the former's own advantage. It is easy to see that this is a consequence of its *mis*informing the receiver: according to Skyrms (2010, p. 74), a signal carries misinformation about a state if it decreases the probability of the state it is sent in, or increases the probability of a state it is not sent in. In the example, $P(S_1)=1/3$, but $P(S_1|M_1)=1/2$. When sent in S_2, M_1 carries misinformation about S_1 – the message increases the probability of a nonactual state. When sent in S_1, it carries misinformation about S_2. That is: as we have just seen, the sender is misleading the receiver as to which one of S_1 or S_2 the world is in.

Skyrms suggests that if the results of sending such misinformative signals is consistently to the interest of the sender, as it is in the equilibrium I have been discussing, it is a case of *deception*. If the signal carries both correct and incorrect information, as with M_1, it is a (deceptive) *half-truth*.

3. Half the Truth and Half-Truths

Godfrey-Smith (2011) argues that M_1 is not really deceptive:

> I do not think [M_1] is a ‘half-truth’, in the sense usually associated with that term. What the sender is doing is refusing to tell the whole truth, but what is said is simply true. The sender is saying something logically weaker than what they know. Expressed propositionally, [M_1] says ‘State 1 or 2 is actual’. To say something logically weaker than what you might say is not to deceive. To tell half the truth is not to tell a half-truth. (Godfrey-Smith 2011, p. 1294f)

Later in the same piece Godfrey-Smith suggests that only *non-maintaining* uses of a signal should count as deceptive. Informally, certain uses of a signal are non-maintaining if they somehow subvert a mutually beneficial interaction between sender and receiver to the exclusive advantage of one of the parties. As a result, non-maintaining uses cannot be arbitrarily frequent: they destabilise the sender-receiver configuration, and too many of them would undermine it (Godfrey-Smith 2011, p. 1295).

According to Godfrey-Smith, M_1 is not non-maintaining, and therefore not deceptive. We should take the message to be telling the truth, and there is certainly a way to do so: after all,

$P(S_3|M_1)=0$, and we are entitled to read M_1 as communicating to the receiver the (true) information that the world is either in S_1 or S_2.

	Act 1	**Act 2**	**Act 3**
State 1	2,10	0,0	10,8
State 2	0,0	2,10	10,8

Table 2: Deception at equilibrium (with a non-maintaining message)

4. Deception Without True Information

While Godfrey-Smith's idea that only non-maintaining uses of signals can be deceptive is appealing, it does not seem to be true in general. To see this, we need a more rigorous understanding of what it is for a certain use of a signal to be non-maintaining.

Consider now the sender-receiver game in Table 2, and the following sender and receiver rules:

Sender Rule:

- $S_1 \rightarrow M_1$
- $S_2 \rightarrow M_1(1/2), M_2(1/2)$

Receiver Rule:

- $M_1 \rightarrow A_1(4/5), A_3(1/5)$
- $M_2 \rightarrow A_2$
- $M_3 \rightarrow A_1(4/5), A_3(1/5)$

Assuming that $P(S_1)=2/3$ and $P(S_2)=1/3$, this is a Nash equilibrium.[2] Now, when produced in S_2, M_1 carries misinformation about S_2: $P(S_2)=1/3$, but $P(S_2|M_1)=1/5$. This misinformation is beneficial for the sender: it's preventing the receiver from producing A_2 in S_2 in those cases – which would be great for the receiver, but disastrous for the sender. According to Skyrms, then, M_1 is deceptive. But here we cannot read M_1 as telling any partial truth: both $P(S_1|M_1)$ and $P(S_2|M_1)$ are greater than zero. The only "truth" the sender is communicating, expressed propositionally, is the empty claim that the world is either in S_1 or S_2. It would be at least awkward to insist that M_1 is not deceptive, and to claim that the truth it is expressing is that the world in in some state or other.[3] This is the example of Skyrms-deceptive signals without the communication of partial truths that I announced in the introduction.

[2] The calculation of equilibria for the games in this paper, from Table 2 onwards, has been carried out using the implementation of Lemke's (1965) algorithm provided by the program *Gambit* (McKelvey, McLennan, Turocy 2010).

[3] There is another option (suggested to me by an anonymous reviewer): denying that M_1 – which, as we have seen, has a tautologous propositional content – has meaning at all. It would be, perhaps, rather a *non-message*. This is, I think, a good illustration of the way in which Skyrms' informational contents are more explanatory than propositional contents, at least in the context of these games: if M_1 was just a non-message one would expect the receiver to respond to it by producing A_3 alone; after all, non-messages, presumably, are communicatively inert, and A_3 is the best response to the mix of two thirds of S_1 and one third of S_2 that the unconditional probabilities of these states produce. In fact, the receiver responds by mixing four fifths of A_1 and one fifth of A_3 – that is, by correctly adapting its response to the information that M_1 carries about world states. This provides evidence that M_1 is meaningful.

On the other hand, Godfrey-Smith is not forced to endorse the verdict that M_1 is not deceptive: he advocates for tying deceptiveness to non-maintenance, and the tokens of M_1 that are sent in S_2 *do* seem non-maintaining. These tokens undermine the sender-receiver configuration in the following clear sense:

Currently the receiver "listens" to what the sender has to say; that is, it does not produce the same response to every message—it does not *pool*. But, if the frequency with which the sender issues M_1 in S_2 grows too much, this will stop being so: if, in the limit, M_1 is the only message sent in S_2 the receiver is free to stop listening (is free to pool) without loss. This is because the sender would be producing only tokens of M_1 in every state, and there is a best response to this sender rule in which the receiver pools every message to act A_3: communication has been lost.

The idea, then, is to think of non-maintaining uses of signals as comparatively rare, abnormal happenings in an otherwise mutually beneficial signalling interaction. If a sufficient proportion of the signals sent in a certain state were of the non-maintaining kind, the receiver would lose interest in what the sender has to say, and would start pooling.

A game-theoretic way of capturing what non-maintaining messages are, therefore, is the following:

Non-Maintaining Message:

> Consider a sender rule that includes a statement of the form $S_i \rightarrow M_1(P_1), \ldots, M_j(P_j), \ldots, M_n(P_n)$, such that the best response by the receiver to this sender rule is not pooling. M_j, when sent in S_i, is non-maintaining iff, if we substitute the former statement with $S_i \rightarrow M_j(1)$, the receiver has a pooling best response.

This definition makes the correct predictions, and provisionally vindicates Godfrey-Smith's point: in Skyrms original game (Table 1), there are no non-maintaining uses of signals but, according to Godfrey-Smith, nor is there deception. In the game presented in Table 2, M_1, when sent in S_2, is deceptive, but also non-maintaining.

5. Deception Without Non-Maintaining Uses of Signals

But there *can* be deception in equilibrium, in the absence of non-maintaining messages. Consider now the game in Table 3.

	Act 1	**Act 2**	**Act 3**
State 1	10, 5	5, 10	0, 0
State 2	10, 10	5, 5	3, 3
State 3	10, 0	0, 10	5, 15

Table 3: Deception without non-maintaining uses of signals

Here the receiver would rather do A_2 in S_1, and A_3 in S_3. The sender, on the other hand, prefers A_1 in both these states. They agree on what's best in S_2. When all three states are equiprobable, the following combination of sender and receiver rules is an equilibrium in this game:

Sender Rule:

$S_1 \rightarrow M_1(24/25), M_3(1/25)$

$S_2 \rightarrow M_3$

$S_3 \rightarrow M_2(13/25), M_3(12/25)$

Receiver Rule:

$M_1 \rightarrow A_2$

$M_2 \rightarrow A_3$

$M_3 \rightarrow A_1(1/3), A_2(1/3), A_3(1/3)$

When sent in S_2, M_3 carries correct information about this state. But, when sent in S_1, M_3 carries misinformation about this state: $P(S_1|M_3)=1/38$, much lower than $P(S_1)=1/3$.

Receiving M_3 increases the probability of being in S_2; the receiver, thus, mixes A_1 as part of its response to M_3, in the hope of reaping the benefits of an S_2/A_1 pair. In consequence, from time to time it finds itself doing A_1 in S_1, contrary to its interests, and in favour of the sender's interests: M_3 is deceptive (according to Skyrms's characterisation) in S_1.

Again here, reading M_3 as communicating "half the truth" is not particularly attractive: all of $P(S_1|M_3)$, $P(S_2|M_3)$, $P(S_3|M_3)$ are greater than zero, and the putative partial truth cannot be anything stronger than "The world is in some state".

Finally, M_3 is not non-maintaining. If the sender changes its rule to

$S_1 \rightarrow M_3$

$S_2 \rightarrow M_3$

$S_3 \rightarrow M_2(13/25), M_3(12/25)$

then a best response by the receiver is

$M_1 \rightarrow A_2$

$M_2 \rightarrow A_3$

$M_3 \rightarrow A_2$

This response has a expected payoff of 9.2 – higher than the payoff of pooling all messages to A_1 (which is 5.0), A_2 (8.3) or A_3 (6.0). The receiver, that is, cannot afford to stop listening to the sender, even if the sender overdoes the proportion of deceptive messages in S_1.[4]

The reader who find this example compelling might still wonder whether this kind of not non-

[4] Both here and in the game discussed in section 6 the receiver resorts to partial pooling. This could be taken to imply that the receiver's responsiveness to the sender is decreasing: in the receiver's final response, A_2 carries less information about the message sent by the sender than it did in the original Nash equilibrium. On the other hand, A_3 carries *more* information about the message sent by the sender than it did in the Nash equilibrium: the receiver, one might say, is increasing its responsiveness to the faithful messages that the sender is still sending.

maintaining, deceptive messages is just a theoretical curiosity. If only contrived, “laboratory” games have equilibria with such messages, it might be theoretically sensible simply to disregard them, and insist that Godfrey-Smith's non-maintenance condition is still necessary for a message to be deceptive “in the wild”.

In fact, an exploration of the space of games with 3 states, 3 messages and 3 acts, with equiprobable states and payoffs depending only on the combination of states and acts, shows that not non-maintaining deceptiveness is, while certainly infrequent, not terribly so: random sampling yields that about 0.5% of all such games have at least one equilibrium with at least one non-maintaining, deceptive message.

More precisely, in a random sample of 50,000 such games (i. e., games in which every payoff for sender and receiver is an independently generated pseudo-random number between 0 and 1), 251 games had at least one Nash equilibrium in which at least one pair of message and state, $\langle M_i, S_j \rangle$, had the following combination of features:

- The mutual information between states and acts is nonzero – i.e., the sender is providing information with its messages, and the receiver is using this information to guide its actions.
- $P(S_j|M_i) < P(S_j)$ – i.e., M_i misinforms about S_j when sent in S_j.
- $\forall k [P(S_k|M_i) > 0]$ – i.e., the message cannot be interpreted as communicating a partial truth.
- If we substitute the S_j statement in the sender rule with $S_j \rightarrow M_i(1)$, the receiver has a best response with which the mutual information between states and acts is still nonzero, and which has a strictly better expected payoff than any pooling strategy – i.e., the deceptive message is not non-maintaining, and making it the only message sent in the state in which it's misinformative does not drive the receiver to stop listening.

Deceptive, not non-maintaining messages, then, appear to be far from a mere theoretical curiosity, and should probably be considered a legitimate constraint in any elucidation of deceptiveness.[5]

[5] This conclusion, as I have said, is based on data gathered randomly from the space of all cheap-talk 3x3x3 games. It remains an open question whether restricting the data-gathering to biologically salient regions of this space (i.e., regions in which games can be used to model actual sender-receiver interactions in nature) would show deceptive, not non-maintaining messages to be less (or more, or equally) prevalent.

There are at least two other respects in which this numerical exploration should be supplemented by further work: first, although casual inspection does not reveal salient common features among the 251 games in the sample that have deceptive, not non-maintaining signals, it is possible (perhaps likely) that more systematic exploration might uncover such similarities in the payoff structure of these games. Second, in this paper I am using Nash equilibria as the target equilibrium concept. This should be complemented with a study of how accessible to evolution, and how dynamically stable, those equilibria are.

6. Not Non-Maintaining Deception and Costly Signalling

The discussion so far has focused on “cheap talk” games: those in which the sender can choose which message to produce, at no cost for them. In the literature on sender-receiver games it is often suggested that information exchange between sender and receiver, in the absence of perfect alignment of interests, can be stable if there is cost associated to the production of some signals. The following example – a variation on the kind of asymmetric-information games in "lemon markets" first discussed by Akerlof (1970); see also Zollman et al. (2013) for similar examples of differential-cost signalling games – shows that costly signalling is compatible with deception that is not non-maintaining:

	Act 1		Act 2		
	M_1	M_2	M_1	M_2	M_3
State 1	4, 4	4, 4	0, 0	0, 0	0, 0
State 2	2, 2	3, 2	-2, 0	-1, 0	0, 0
State 3	1, -8	2, -8	-3, 0	-2, 0	0,0

Table 4: Deception without non-maintaining uses of (costly) signals

	Act 1		Act 2		
	M_1	M_2	M_1	M_2	M_3
State 1	$p, H - p$	$p, H - p$	$0, 0$	$0, 0$	0, 0
State 2	$p - c, M - p$	$p - e, M - p$	$-c, 0$	$-e, 0$	0, 0
State 3	$p - d, L - p$	$p - f, L - p$	$-d, 0$	$-f, 0$	0, 0

Table 5: Deception without non-maintaining uses of costly signals – deriving the payoff matrix

We can think of the sender as a seller trying to make the most out of her stock, which includes both good items and others of medium and low quality. On the other hand, the receiver is a buyer trying to get the best return on investment. While I do not wish to make any strong claims regarding the biological relevance of this game, it appears to admit of relatively natural biological interpretations: for example, the sender could be thought of as a tree which bears fruits of different nutritional value – from very nutritious to slightly poisonous –, and the receiver a pollinator bird who would rather not waste its time on one of the less attractive fruits.

This game is an example of differential-cost signalling: the sender is able to present the quality of its stock as better than it is, but doing so has an associated cost – which is higher the bigger the lie is.

For our current purposes, what this game shows is that lying, even when costly, need not be non-maintaining: even if the seller overdoes it, and flags *all* of its low-quality goods as high quality, it might be (in the game to be described, it is) still providing reliable information about the goods of middle quality, so that the buyer still finds incentive in listening to the seller's messages.

A Description of the Game

States: In S_1 the sender has high-quality goods on offer (the tree bears highly nutritious fruit, say). In S_2, medium-quality goods. In S_3, low-quality goods.

Messages: They consist in the sender displaying its goods prominently (M_1), less prominently (M_2), or not at all (M_3). Costly signalling comes at this point: A sender with goods of medium or low quality needs to spend resources in making them look good, and this is more expensive for low quality than for medium quality goods, and more expensive for those more prominently displayed than for those less so. Perhaps the tree has to spend resources in making the medium and bad quality fruit mimic the appearance of the better variety, and the quality of the mimicry is more important if the fruit is prominently displayed than if it is hidden among the leaves.

Acts: The receiver has the option of buying the goods (act A_1) or refraining from doing so (act A_2). In the case of the pollinator bird, say, it has the options of incurring in the resource expenditure associated with actually flying all the way to the fruit and eating it, or simply leaving it be.

Payoffs: The actual values of the payoff matrix in Table 3 are calculated in the following way: A receiver has to spend 4 resource units (p) if she wants to obtain goods of any kind. The real value of the high quality goods is 8 units (H); the medium quality goods, 6 units (M); the low quality, -4 units (L) – they are actually deleterious for the receiver. This means that a receiver gets 8 - 4 = 4 resource units if it chooses to buy (do act A_1) in state S_1, it gets 6 – 4 = 2 resource units if it buys in S_2, and -4 – 4 = -8 if it buys in S_3. The receiver loses nothing and gains nothing if it chooses not to buy.

As regards the sender, making medium-quality goods in prominent display look good costs 2 units (this is quantity *c*); making medium-quality goods in less prominent display look good costs 1 unit (*e*). The costs for bad-quality goods in prominent and less prominent display are 3 (*d*) and 2 (*f*) units, respectively. So, for example, if the sender bears bad-quality goods (i.e., if the world is in state S_3) and chooses to display them prominently (i.e., if it sends message M_1), then it will incur in a cost of 3 resource units. If it manages to fool the receiver into buying, its net payoff will be 4 – 3 = 1 – this is the entry in the payoff matrix under state S_3, message M_1, act A_1; if it doesn't fool the receiver, its net payoff will be 0 – 3 = -3 – and this is under S_3 / M_1/ A_2 in the payoff matrix.

Table 5 summarises how the payoffs in Table 4 are calculated – the variables correspond to the letters between brackets in the foregoing two paragraphs.

If all three states are equiprobable, the following combination of sender and receiver rules constitutes a Nash equilibrium in this game:

Sender Rule:

- $S_1 \rightarrow M_1$
- $S_2 \rightarrow M_1(1/4), M_2(3/4)$
- $S_3 \rightarrow M_1(9/16), M_2(3/16), M_3(1/4)$

Receiver Rule:

- $M_1 \rightarrow A_1(3/4), A_2(1/4)$
- $M_2 \rightarrow A_1(1/2), A_2(1/2)$
- $M_3 \rightarrow A_2$

That is, when the seller has high-quality goods, they only send message M_1. When they have medium-quality goods, they send a combination of M_1 and M_2. When they have low-quality goods, they mix all three messages.

The first piece of behaviour increases the probability of good-quality goods conditionally on M_1. The other two pieces of behaviour are, partially, about taking profit of this established informational link. In this equilibrium, for example, the receiver resorts to buying three out of four times in which they receive M_1. This strategy gives a good chance of buying goods of high quality, but it also implies that the receiver will, sometimes and against their interest, buy low-quality goods. Analogously, the receiver buys half of the times it receives M_2; this will get it many medium-quality goods, but also some low-quality ones. The sender is also, part of the time, honest about the lower end of its stock: M_3 is only sent when it has low-quality goods on offer. And, sure enough, the receiver never buys when it receives M_3.

More formally, given these sender and receiver rules, M_1 carries misinformation about S_3 when sent in S_3: $P(S_3)=1/3$, but $P(S_3|M_1)=9/29$. This misinformation is beneficial to the sender: it is allowing it to trick the receiver into buying goods of low quality some of the time. Moreover, again here, all of $P(S_1|M_1)$, $P(S_2|M_1)$ and $P(S_3|M_1)$ are greater than zero: we cannot help ourselves to the maneuver of reading M_1 as telling half the truth. Indeed, M_1 (displaying the goods prominently) does appear pretheoretically to be deceptive when sent in S_3 (that is, when the goods are of bad quality): the sender is bluffing about the quality of its stock, by placing bad-quality goods in the situation in which, normally, high-quality goods are displayed.

Finally, is M_1, when sent in S_3, non-maintaining? It is not. First of all, intuitively, these uses of M_1 are part of what makes the sender-receiver configuration worth the while for the sender. If it was prevented from bluffing from time to time about the lower end of its stock, it would not be able to afford providing correct information about its quality the rest of the time. And they are not non-maintaining in the formal regimentation of this notion introduced in section 4: if the sender changes the third statement in its rule to

- $S_3 \rightarrow M_1(1)$

the receiver cannot afford to stop listening. That is to say, its best response is not pooling. To see this, consider the sender rule with the new S_3:

Sender Rule:

$S_1 \rightarrow M_1$

$S_2 \rightarrow M_1(1/4), M_2(3/4)$

$S_3 \rightarrow M_1$

And the following candidates for a receiver rule:

a)	*b)*	*c)*
$M_1 \rightarrow A_2$	$M_1 \rightarrow A_2$	$M_1 \rightarrow A_1$
$M_2 \rightarrow A_1$	$M_2 \rightarrow A_2$	$M_2 \rightarrow A_1$
$M_3 \rightarrow A_2$	$M_3 \rightarrow A_2$	$M_3 \rightarrow A_2$

In b) the receiver is pooling, refraining from buying in every case. In c) it is buying wherever it can[6]. In a) it buys when it receives M_2, abstains otherwise. As it turns out, a) is a better response than b) or c): in a) the receiver nets $1/3 \cdot 3/4 \cdot 2 = 1/2$ units; in b), 0 units; in c), -2/3 units.

This is because, even if the most prominent display makes now low-quality goods as likely as high-quality ones, the sender is still providing correct information with M_2, the less prominent display, about the presence of medium quality goods. The receiver is better off using this information than ignoring it. Therefore, no pooling response is a best response, and M_1 in S_3 is not non-maintaining. This is, then, a case of deception without non-maintaining uses of signals, without the communication of any correct information by the deceptive signal.

Godfrey-Smith proposes making non-maintenance a constitutive feature of deception in sender-receiver games. I have shown that one can describe cases (and quantitative exploration has shown that their proportion is far from negligible) in which deceptive signals can be arbitrarily frequent, without this undermining the sender-receiver configuration: whenever the sender sends several types of signal, increasing the frequency of a certain lie, even all the way to 1, might still leave room for fruitful communication. This, at least, casts doubt on the proposed constitutivity of non-maintenance to deception.

Funding

Research for this paper was supported by the Spanish government via research grants MCINN FFI2011-26853 and CSD2009-0056 (CONSOLIDER INGENIO)

Acknowledgements

I would like to thank Peter Godfrey-Smith and two anonymous reviewers for their helpful comments to earlier drafts.

References

Akerlof, G., 1970, “The Market for ‘Lemons:’ Quality Uncertainty and the Market Mechanism,” *The Quarterly Journal of Economics*, 84, pp. 488–500.

[6] Although this is not pooling in the strict sense, I will grant that this response counts as pooling for the purposes of the definition of non-maintaining message: the game does not allow for a pooling, *always-buy* strategy, and this is as close as the receiver can get to such a strategy.

Crawford, V. P. and Sobel, J., 1982, "Strategic Information Transmission". *Econometrica,* 50 (6), pp. 1431–1451

Godfrey-Smith, P., 2011, “Signals: Evolution, Learning & Information, by Brian Skyrms,” *Mind,* 120 (480), pp. 1288 – 1297.

Godfrey-Smith, P. & Martínez, M., 2013, “Communication and Common Interest,” *PLOS Computational Biology,* 9(11)

Johnstone, R. A., 1997, “The Evolution of Animal Signals” in (Krebs, J. R., Davies, N. B., eds.) *Behavioural Ecology: An Evolutionary Approach,* Oxford: Blackwell, pp. 155 – 178

Lemke, C.E., 1965, “Bimatrix equilibrium points and mathematical programming,” *Management Science*, 11, pp. 681-689

Lewis, D., 1969, *Convention*, Cambridge: Harvard University Press.

McKelvey, R.D., McLennan, A.M., and Turocy, T.L., 2010, Gambit: Software tools for game theory, Version 0.2010.09.01. http://www.gambit-project.org.

Searcy, W. and Nowicki, S., 2005, *The Evolution of Animal Communication*, Princeton University Press

Skyrms, B., 1996, *Evolution of the Social Contract*, Cambridge: Cambridge University Press.

Skyrms, B., 2010, *Signals: Evolution, Learning & Information*, New York: Oxford University Press.

Spence, M., 1973, "Job Market Signaling," *Quarterly Journal of Economics*, 87 (3), pp. 355 – 374.

Wagner, E., 2012, “Deterministic Chaos and the Evolution of Meaning,” *The British Journal for the Philosophy of Science,* 63, pp. 547–575

Zahavi, A., 1975, “Mate Selection – A Selection for a Handicap”, *Journal of Theoretical Biology*, 53, pp. 205–214

Zollman, K. J., Bergstrom, C. T. & Huttegger, S. M., 2013, “Between cheap and costly signals: the evolution of partially honest communication,” *Proceedings of the Royal Society B* 20121878. http://dx.doi.org/10.1098/rspb.2012.187