ARTICLE

Assertions, Handicaps, and Social Norms

Peter J. Graham*

University of California, Riverside, CA, USA *Corresponding author. Email: peterg@ucr.edu

(Received 6 June 2019; accepted 18 November 2019)

Abstract

How should we undertand the role of norms – especially epistemic norms – governing assertive speech acts? Mitchell Green (2009) has argued that these norms play the role of handicaps in the technical sense from the animal signals literature. As handicaps, they then play a large role in explaining the reliability – and so the stability (the continued prevalence) – of assertive speech acts. But though norms of assertion conceived of as social norms do indeed play this stabilizing role, these norms are best understood as deterrents and not as handicaps. This paper explains the stability problem for the maintenance of animal signals, and so human communication; the mechanics of the handicap principle; the role of deterrents and punishments as an alternative mechanism; and the role of social norms governing assertion for the case of human communication.

Keywords: epistemology; norms; animal cognition; signaling; social epistemology

For the most part, the philosophical literature on the norms of assertion relies on intuitive reactions to cases and reflection on conversational data to uncover *the* epistemic norm governing assertion: of the myriad norms by which we might evaluate an assertion, which one is *the* norm of assertion?¹ This literature seems to show relatively little interest in related literatures on animal signaling, the evolution of language, and the roles social norms play in stabilizing reliable informative communication. Mitchell Green's "Speech Acts, the Handicap Principle and the Expression of Psychological States" is one of a handful of notable exceptions (Green 2009).²

Following a well-established tradition in speech act theory, Green argues that to assert is to express a belief, and to express a belief is to give your audience strong evidence that you have the belief (e.g. Searle 1969; Bach and Harnish 1979; Davis 2002). This is the speaker's half of the "Gricean handshake" underwriting distinctively human ostensive-inferential communication. Green's innovation is to ask how it is we provide strong evidence for our beliefs – how do we "express" our beliefs by asserting? He answers that assertions give strong evidence for our beliefs because assertions are "handicaps" in the technical sense of the word from signaling theory. Handicaps in this sense are costs that the signaler pays designed to ensure the honesty (the accuracy)

¹The norms of assertion literature continues to grow. Williamson (1996) has been especially influential. Goldberg (2015) is a handy guide to the literature.

²See also Knight (1998), Scott-Phillips (2010), Graham (2015) and Turri (2016).

[©] The Author(s), 2020. Published by Cambridge University Press.

of the signal. That's why assertions are strong evidence for the speaker's belief. And why are they handicaps? They are handicaps, Green argues, because of the role of norms for sincere and informed asserting. Norms as handicaps – norms that impose costs that honest signalers pay – thereby explain the very possibility of assertive speech acts.

I admire Green's turn to the animal signals literature for reflection on human communication – we are animals, after all. I agree that social norms help explain the reliability of assertion, especially in cases where speaker and receiver interests do not necessarily coincide. But I disagree that norms are handicaps in the technical sense. The norms governing assertion are *deterrents*, not handicaps. That's how norms – *social* norms – help explain the reliability of assertive speech acts. Or so I will argue.

1. The stability problem

To understand handicaps in the technical sense, you need to understand the stability problem for animal signaling systems, for handicaps purport to solve the problem. In this section I explain the problem. In the next I explain handicaps as the solution.

Animals signal in a variety of ways for a variety of purposes. Birds sing to attract mates. Dogs dole out urine to mark territory. Chimps hoot when they find food. Ground squirrels make high-pitched whistles to warn offspring of coyotes. Cranes dance to announce their pair bond. A male spider vibrates the female's web to encourage her to mate with him instead of dining on him. In general, animal signals are evolved signal-response *pairs* – what we might also call evolved signaling *systems* – for the sender's signal and the receiver's response evolved in response to one another (Maynard Smith and Harper 2003). The sender's signal evolved to benefit the sender (to assist survival and reproduction). That's why the sender sends the signal. The receiver's response evolved to benefit the receiver. That's why the receiver responds to the sender's signal. How do they benefit? The sender benefits by influencing the receiver. The receiver benefits by receiving valuable information. The sender effectively trades information for influence.

Take the roaring of male red deer when competing for mates. If they fight, both get hurt. Since the larger deer nearly always wins, they will both be better off if the smaller deer simply retreats without a fight. But how do they know which one is bigger, which one is smaller? By signaling. They bellow to signal their size. They then learn relative size, and the smaller deer retreats. Roar to signal your size to get the other to retreat. Retreat when you learn you are smaller than your rival. Both sides benefit. Animal signals are then evolved sender-response pairs – evolved behavioral strategies – where both sides, on average, benefit (Johnstone and Grafen 1993).

When the signal-response pair is "at" or "in" equilibrium – when neither side benefits by changing their behavioral repertoire – we say that the signal system (the signalresponse pair) is *stable*. An equilibrium just is a stable system: neither side (neither the sender nor the receiver) will do better by changing strategies; an equilibrium just is a strategy from which it does not pay to deviate (to "defect"). When a signal system is stable, a mutant – a defector – would then be worse off. Mutants might occur, but they won't last for long, and so they won't evolve. The "stability" problem is then explaining why a signal system remains stable; it then just is explaining why neither side would be better off changing their strategies, why mutants aren't expected to survive and reproduce.³

³For the sake of ease of exposition, I will often write without hedges or qualifications when presenting the evolutionary logic of animal signals. Even so we should be aware that we are dealing with probabilities,

Why should this be a problem? Why should there be something to explain? Because when it comes to animal communication, there is evolutionary pressure for the sender to take advantage of opportunities to send misleading information. If the sender can fool the receiver into providing a response that benefits the sender, but without sending useful information to the receiver, then there will be pressure for the sender to do just that. And so imagine a mutation (a "defection") that leads a smaller red deer to signal that he is much bigger than he really is. He then fools larger red deer when competing for mates; his opponent retreats. He wins a competition he would have otherwise lost. The mutant smaller deer will then pass on his mutation to his offspring, thereby increasing the rate of false signals in the population. Dishonesty spreads.

So far so good. But soon enough a mutation to ignore such signals (and fight anyway) will appear in the population. Then deer with the counter-mutation to ignore signals of size – especially larger deer – will benefit by ignoring those signals. They will not retreat but fight, for now there is a good chance they will win. Once they win often enough, the larger mutants will have more mutant offspring and pass on the mutation to ignore such signals. Ignoring signals of size will take over in the population. Though dishonesty paid in the short-term, dishonesty no longer works.

We started with a signal-response pair that benefitted both senders and receivers. But once evolution pressured senders to defect and to send dishonest signals, evolution in turn pressured receivers to defect as well and ignore signals of size. The signal system – the sender-response pair – collapsed.

Sending the signal might disappear from the population as well. Why waste energy? Aesop's fable of the boy who cried 'Wolf!' vividly illustrates the idea. Send too many false signals, and no one will believe you anymore. You might as well stop talking.

The problem for studying animal signals is then to explain why a signaling system stays stable, why signals tend to be honest on average, especially in cases of non-coincident interests. Given pressures on senders to "cheat" the system by sending dishonest information for short-term biological gain, defection that in turns prompts a response that undermines the system, why do signal systems remain stable, when they do? Once honest signaling evolves, what keeps mutant defectors from invading for short-term gain? What keeps dishonest signals (for the most part) out of the stable signaling business (Johnstone 1995; Maynard Smith and Harper 2003; Scott-Phillips 2010)? That's the stability problem.⁴

The stability problem is not just a problem for non-human animals but for human animals too. Just like deer, we compete in a variety of ways on various occasions for a variety of goods. So why is human communication – like any other stable animal signaling system – as honest as it is? Human communication – as Grice (1975) made abundantly clear – presupposes high levels of cooperation; a great deal of the time

contingencies, and a number of other influencing factors, that would complexify the complete story for the evolution of any signaling system. I will also help myself to the common practice of using the intentional stance to explain the evolutionary forces at work. It is a helpful device, not to be confused with the underlying ontology of changes in frequencies in genes over generations within populations that help build behavioral traits.

⁴When a system is stable, does that mean that the signals are invariably "honest"? No. How frequently signals need to be honest for the system to remain stable varies case-by-case, depending on the overall costs and benefits of relying on signals, among other issues. For example, if believing a falsehood doesn't cost much, then falsehoods may spread without destabilizing a system. A good deal of "dishonesty" might co-exist with honest signals in a stable system. Stable systems seem to require honesty "on average" or "for the most part" – or at least stable systems on average and for the most part do. For some discussion of the range of possible deception in a stable system, see Mokkonen and Lindstedt (2016).

we say what we believe and what we have evidence for. Why in our world – a Darwinian world – should this be so?

One answer should already have come to mind: unlike many animals, humans, like ants and bees, frequently have coincident interests, and so we frequently have every reason to share accurate information with our interlocutors. Only when our interests conflict is there a reason to "defect" and send dishonest signals. Fortunately for us our interests are so frequently aligned.

I agree with a good deal of this sentiment. In particular, I think that a large part of the explanation for the origins of language in the first place involves kin selection – sharing useful information with your offspring and close kin (Fitch 2010; Laland 2017). I also think the need to forage collaboratively to survive played a large role in the spread of language amongst non-kin throughout human evolution (Tomasello 2008). These are both cases of coincident interests – whether genetic in the case of kin selection or prudential in the case of collaborative projects. Helping kin, and helping ourselves by helping our collaborators, both go a long way towards explaining why we honestly provide relevant information.

On the other hand, I do not think coincident interests tells the whole story when it comes to human communication and its origins. I agree instead with Michael Lachmann, Szablocs Számádo, and Carl Bergstrom when they claim that

[though] language can plausibly arise and be maintained by natural selection when individuals have coincident interests ... [even so] human language almost certainly did not evolve in an Eden of coincident interests ... [C]onflicting interests would have been frequent during the origin of language (as they are now), and that the problem of honesty would have exerted a continuing influence on the development of language. (Lachman *et al.* 2001: 13189)

Hugo Mercier and Daniel Sperber echo this sentiment in their recent book *The Enigma of Reason*:

Human communication is definitely not limited to common interest where trustfulness and trust are mutually advantageous to the interlocutors. Linguistic signals can be produced at will to inform or to mislead. Human communication takes place not only among close kin or cooperators but also with competitors and strangers. Lying and deception are in everyone's repertoire. (Mercier and Sperber 2017: 189; cf. Sperber 2001; Sperber *et al.* 2010)

Unlike ants, bees and termites and closely related ground squirrels, human interests do not always coincide.

And even when it comes to close kin and others with whom we generally share interests, there will always be cases where, on the occasion, our interests do not align. How, then, outside of sharing information with close kin (and sometimes even when sharing information with close kin) do we solve the stability problem for human communication? When there are considerations in favor oflying or misleading or speaking without sufficient evidence, but we toe the line anyway, what counter-considerations help explain our honesty? Why is telling the truth – at least often enough to keep listeners listening – the stable strategy?

2. The handicap solution

For some time zoologists clearly recognized the stability problem but lacked a clear solution. Then in 1975, Amotz Zahavi proposed a solution – the so-called "Handicap

Principle" – in a justly famous article entitled "Mate Selection – A Selection for a Handicap" (Zahavi 1975, 1977, 2007, 2008).

Zahavi's article addressed an issue that puzzled Darwin (1871). Why should ostentatious traits – especially traits like the peacock's tail – exist, given that they often come at high energy costs to produce and maintain, and often put their bearer at greater risk of predation? Darwin's answer was *sexual* selection. They exist, he argued, because members of the opposite sex choose their mates on the basis of such traits. Those traits are then "sexually" selected. If you've got them, you reproduce. If you don't, you go out of the reproduction business. The benefits to reproduction thereby outweigh the costs to survivability. Applied to the peacock's tail, the idea is that peahens choose mates because of their tails. Though the tail may reduce survivability, it enhances reproducibility.

But why do peahens choose mates with such outlandish tails? The logic of natural selection suggests peahens choose based on reproductive fitness. Somehow their choice will lead to differentially more viable offspring over time. So how does choosing mates with large tails increase reproductive fitness?

Zahavi's idea was that peahens choose mates for the overall quality of their genes. But since peahens cannot simply peer inside peacocks and detect the quality of their genes, peacocks would have evolved to provide a signal of the quality of their genes. That's what, Zahavi claims, the ostentatious tail does. The peacock's tail (actually long feathers) provides the valuable information the peahen needs. A male then grows a large tail to influence mate choice, by conveying information about the bearer's fitness. "Look at my long tail," the male peafowl seems to say to the female peafowl, "it proves how genetically fit I am, and so you should choose me." The peafowl, detecting the tail, then selects the male with the largest, most ostentatious display, for that best improves her reproductive potential.

But don't forget the stability problem. Less fit males, males with lower genetic quality, will produce dishonest signals that also say "I've got great genes" if they can, for they too would clearly benefit from producing such signals, for they would fool peahens into mating with them. If lower genetic quality males can produce a dishonest signal that's just as effective at persuading females as an honest signal from high-quality males, even if it only works in the evolutionary short-term, that's what we should expect to evolve. The system is then on its way to collapsing.

Zahavi's "handicap" solution purports to explain why this isn't likely to occur, why defectors are not likely to evolve, why the signaling system is apt to remain stable. Here's his idea, stepwise. If all the steps hold, then we have a sufficient condition for stability; dishonesty won't invade and destabilize the system.

First, we assume an honest signal has evolved, so that signalers trade honest information to influence receivers. We start with a signaling system and then explain its stability.

Second, we note that the signal – the tail – comes with a cost to survivability and so to fitness: the peacock's tail weighs a good deal, can be difficult to grow and maintain, and peacocks have to drag it around everywhere they go; it slows them down when fleeing from danger; it is just as visible to friend as foe.

Third, we note the obvious point that an *honest* signaler will have high genetic fitness and a dishonest signaler will not, for the signal means "I have great genes."

Fourth, we note that the overall benefits to the honest signaler are positive, despite the costs in fitness. Though expensive to produce, the honest peacock who is high in fitness has enough fitness to survive and so to successfully mate and reproduce.

Fifth, we note that the overall net benefits to the dishonest signaler are negative. Because of the fitness costs to produce such a tail, the dishonest signaler, low in fitness, cannot afford to produce such a tail. A mutation in a less fit peacock might produce a large tail, but it won't be strong enough to drag it around while foraging for food, or fast enough to evade predators.

Sixth, we then draw out the consequences of this logic: though mutants might emerge – low quality peacocks might grow long tails – they will not evolve, for the fitness costs for dishonest peacocks are too high; the peacock with the dishonest tail isn't likely to survive long enough to reproduce or have viable offspring. Like buying a car you cannot afford and still pay your bills, the peacock's tail comes with a cost that the dishonest signaler can't really afford. Defection doesn't pay. If it costs a lot to produce a signal, where only the honest signaler can afford to pay that price, then dishonest signals won't evolve and undermine the signaling system. That's Zahavi's idea.

Though we've focused on one example involving sexual selection,⁵ Zahavi's idea is more general. For example, if you are looking to date someone who is really rich, look for someone who drives a Ferrari and owns a house with a number of empty bedrooms (like Mr. Darcy), for someone who isn't rich probably couldn't afford such expensive goods and still pay the bills. Someone who can waste all that money is sure to have plenty more where that came from, whereas someone who isn't rich who finances such expensive goods is sure to go broke trying. Given the cost of producing such a signal, dishonesty doesn't pay. Zahavi's idea is actually an old idea (cf. Veblen 1899; Spence 1973).

A "handicap" is then a signal that costs something to produce, with a cost that only the honest signaler can afford. That's Zahavi's idea.

The key idea is *differential* cost. If receivers only respond to certain signals with a cost, a cost that honest signalers can afford but dishonest signalers cannot, then dishonest signaling will not invade and destabilize the system. The idea is not simply that signals are costly. The idea is not simply that honest signalers invest in a quality to advertise their quality. The idea rather is that dishonest signalers cannot equally afford the investment to engage in false advertising. If they could also afford effective but false advertising, the signaling system would eventually collapse.⁶

⁵Zahavi's explanation of the peacock's tail is not the only explanation of the tail in the literature on sexual selection. Sir Ronald Fisher, the famous statistician and geneticist, offered another explaination. He held the view that once females (in the animal kingdom it is typically, though not always, females that chose with whom to mate, for sperm is cheap and eggs are expensive) start to choose males on the basis of having a particular trait - females "see" that trait as "sexy" - then where their choosiness is inherited, those traits (even exaggerated traits that harm the survivability and so handicap males) are bound to evolve in the population. The idea is that once other females choose males for those traits, then any female who doesn't is in big trouble. For if she mates with a male without such a trait, her sons won't have that trait either. Then her sons won't be chosen by those other females who choose for the "sexy" traits that her sons lack. She'll then risk selecting herself out of the reproduction business, for she's much less likely to have any grandchildren. Hence once other females are choosing for sexy traits, all females will do so to ensure that they have sexy sons, and so grandsons and granddaughters. Fisher's side is then called the "sexy-son" or "good-taste" side of the debate. On Fisher's view, the peacock's tail isn't a signal. It is simply a trait that females prefer. Zahavi's side disagrees. Zahavi's side holds that females choose instead for mates with good genes, so their male and female offspring are more likely to survive and reproduce. The tail signals genetic quality. Zahavi's side is then called the "healthy-offspring" or "good-sense" side of the debate. (For reviews of the debate, see Cronin 1991; Zuk and Simmons 2018). Regardless of the particular case - regardless of whether the tail is a signal or simply a "sexy" trait - Zahavi identified a mechanism that can stabilize signals, even if the peacock's tail isn't a signal.

⁶Alan Grafen draws an important and useful distinction between *efficacy* costs and *strategic* costs. The efficacy costs of a signal are the costs required to get your message across to your audience, whether the message is honest or dishonest. If someone is standing right next to you, you can whisper. If they are standing a football field's length away, you'll need to shout as loud as you can. Depending on the circumstances –

A key element of Zahavi's idea is that the signal means that the sender has the *quality* that the signal costs to produce. The peacock's tail means "high genetic fitness" for only the male with good genes can produce such a signal. The Ferrari means "I am rich" for only the rich can afford such a car. Generically, Zahavi's idea is that the signal means that the sender has a certain quality Q that costs a degree of Q to produce, that only the honest signaler can afford. For this reason, so-called "Zahavi-handicaps" are often called "quality-handicaps." You have to *waste* some money (a quality) – and so "handicap" yourself – to signal you have money (the quality signaled), otherwise the signal is too easy to fake.

A signal S is then a *quality-handicap* of quality Q that means high-Q when S costs a degree of Q that only the honest signaler can afford. Quality-handicaps prove the degree of the quality they mean by pricing signals lacking that degree out of the market.⁷

When Zahavi first proposed his idea in the mid-70s, a lot of people were skeptical. Zahavi's idea was not simply the idea that nature sexually selects for traits that give one an advantage at reproducing that might trade-off with your chances of surviving. Rather, Zahavi's idea was the seemingly "paradoxical" idea – the idea that "really sticks in the throat," as Richard Dawkins put it in *The Selfish Gene* (Dawkins 2016: 207, 412) – that nature sometimes sexually selects *for* traits that gives one a biological advantage *through* giving one a biological *disadvantage*, so that giving one a disadvantage is, as it were, part of the point of the trait. Traits might then be selected for precisely *because* they handicap their bearer. This is then selection *for* a handicap.

A real proof of possibility would come from a formal model. Initial efforts to construct mathematical models failed, suggesting that nature wouldn't follow such a path (Maynard Smith 1976; Kirkpatrick 1986). The breakthrough occurred when Alan Grafen (1990) successfully published models showing that quality-handicaps can indeed remain stable.⁸

The point about the importance of differential cost has been made by Grafen (1990), Johnstone (1995, 1997), Maynard Smith and Harper (2003), Searcy and Nowicki (2005), Saunders (2009) and Frazer (2012). Once we see that the idea is about differential cost, the *amount* of the strategic cost taken in isolation does not matter for the handicap mechanism to do its work. The combined efficacy and strategic costs for a handicap could be very low and hard to detect, provided that the dishonest signaler still cannot afford even that low cost. The strategic costs – high or low – need only be such that *only* the honest signaler can afford to pay the extra, strategic costs.

⁷Quality-handicaps are sometimes classified as indices. According to Maynard Smith and Harper (2003: 15), an index is a signal (and so part of a stable evolved sender-receiver system) whose intensity is causally related to the quality being signaled, and so cannot be faked. For example, when a tiger leaves its claw marks on a tree, the height of the signal causally correlates with the size of the tiger. The tiger – though he may benefit by signaling that he is taller than he is – can't reach any higher, and so cannot be faked. But that's not quite right, as quality-handicaps can be faked. Mutations – defections – can occur. It is just that it is not worth producing a dishonest copy of a quality-handicap, as the costs are too high for the mutant to maintain the handicap. The mutant can produce the fake signal, he just can't afford it. Defections may occur. They just won't evolve.

⁸Though Zahavi originally proposed his Handicap-Principle to explain how sexual selection stabilizes ostentatious traits, he later generalized to *all* animal signals. *All* signaling systems, Zahavi argues, involve handicaps: First, Zahavi asserts that all stable signaling systems – even chemical signaling between cells in an organism – face temptations (evolutionary pressures) – for defection; a stability problem looms

who is signaling who about what, when they are signaling and how – efficacy can be very low, or even possibly very high, in absolute terms, e.g. number of calories required to maintain or produce the signal. These are arguably high efficacy costs, not necessarily strategic costs. The *strategic* costs of a signal are the *extra* costs involved in Zahavi's mechanism for ensuring the honesty of signals; it is the strategic costs of a signal that make it a handicap. They are the extra costs honest signalers pay to produce a version of the signal that dishonest signalers cannot afford.

After the publication of Grafen's models, Zahavi's idea won over the field. "It is widely considered a truism in biology today," Peter Hurd wrote in 1995 while completing his doctorate, "that signals must be costly to be reliable. This view stems from the so-called handicap principle as proposed by Amotz Zahavi" (Hurd 1995: 217). "Over the past quarter-century," write Lachman, Számádo and Bergmann, Zahavi's "costly signaling" hypothesis "has emerged as the dominant explanation" for honest signaling despite the temptation to deceive (Lachman et al. 2001: 13189). "The problem of dishonest mutants [defectors] seemed intractable," Mark Laidre and Rufus Johnstone write, "until Amotz Zahavi suggested a solution" (Laidre and Johnstone 2013). Since the publication of Grafen's models, James Higham writes, Zahavi's Principle "has become one of the most enduring and well known of all theories in animal behavior and behavioral ecology, and has been adopted by other fields, such as evolutionary psychology and human evolution" (Higham 2014: 8; cf. Scott-Phillips 2010: 122; Zollman 2013: 127). Recent books on human evolution that make heavy use of the principle include Dario Maestripieri's (2012) Games Primates Play, Mark Pagel's (2012) Wired for Culture: Origins of the Human Social Mind, Steve Stewart-William's (2018) The Ape that Understood the Universe: How the Mind and Culture Evolve, and William von Hippel's (2018) The Social Leap: Who We Are, Where We Come From, and What Makes Us Happy.

Given the popularity of Zahavi's proposal, it should then not at all be surprising that a philosopher might harness the Handicap Principle to explain assertion. If biologists have discovered what explains the stability of non-human animal signals, then wouldn't that also explain the stability of human animal signals? Why not handicaps for pragmatics?

3. Deterrents: another solution

Despite the evident popularity of quality-handicaps as *an* explanation for the stability of a signaling system over the last three decades, theorists now are less likely to agree that handicaps are the *only* solution. Here is Kevin Zollman:

Despite its ubiquity, [in recent years] there has been a growing dissatisfaction with the Handicap Principle ... These concerns ... suggest the Handicap Principle may not be the primary explanation for the stability of signaling systems when interests occasionally conflict. (Zollman 2013: 128)

Another mechanism has emerged: deterrents. When it comes to signals, a deterrent is a cost that a dishonest signaler is likely to pay if caught making a dishonest signal. That potential cost deters dishonesty. Deterrents can do the trick.

Deterrents require three elements. First, a receiver needs a way to determine whether a signal is dishonest. Deterrents require verification. Second, the receiver needs a way to dole out punishment. Third, the costs to the dishonest signaler must be high enough to offset the potential benefits from deception. A ten-cent fee for double-parking won't deter bad behavior. With these three elements in place, deterrents can stabilize signaling systems, given conflict of interest. It's how we keep people from parking their cars

for all signaling systems. Second, Zahavi also asserts, quality-handicaps are the *only* solution to the stability problem. Hence, he concludes, *all* signaling systems involve quality-handicaps (Zahavi 1977, 2007, 2008, 2010; Zahavi and Zahavi 1997). Grafen agreed: "Persuasive [honest] signaling necessarily involves waste as only [waste] can enforce honesty" (Grafen 1990: 532).

wherever they want, among countless other forms of behavior. Theoretically speaking, this is a pretty obvious possibility, and now widely recognized (Fitch 2010; Számadó 2011; Laidre and Johnstone 2013; Számadó and Penn 2015).⁹

We now have two mechanisms on the table that both involve costs that explain stability. Since both mechanisms involve costs – counter-incentives to deception – both fall under "costly signaling theory" (Frazer 2012; Higham 2014). Both mechanisms undermine dishonesty by placing counter-incentives on the scale; the conflicting interests that would otherwise undermine the system are counter-balanced by costs the dishonest signaler cannot afford. Either the signal is too costly to fake (quality-handicaps), or faking the signal incurs penalties (deterrents). Either way, dishonesty doesn't pay; that's why honesty is the stable strategy.

When a quality-handicap stabilizes honesty, costs are paid *within* the stable strategy; the costs are actually paid by honest signalers within equilibrium. When a deterrent stabilizes honesty, on the other hand, costs are paid by the dishonest signaler who strays from the stable strategy, *outside* of equilibrium.¹⁰

When deterrents stabilize a system, there can be little to no observed cost to signaling. For the better the deterrent, the less likely anyone will pay any actual costs at all. If no one cheats, only potential cheaters pay potential costs. Stable signals can then be very cheap signals overall (Scott-Phillips 2010: 126; Számadó 2011: 5; Zollman 2013: 129; Zollman *et al.* 2013: 1). Given deterrents, "even unrelated individuals with conflicting interests can communicate honestly by using cost-free or very cheap signals ... [Waste] is not required to create honest signals" (Lachman *et al.* 2001: 13189, emphasis removed).

Theoretically this seems straightforward. Are there empirically established cases of honesty due to deterrents in the animal signals literature? Yes. The existence of "status badges" on male house sparrows and on female paper wasps (the wasps that nest around your house if you live in the United States or Europe) is the standard case in the literature of stability due to deterrents (Rohwer 1977; Møller 1987; Strassmann 2004; Tibbetts and Dale 2004; Tibbetts and Izzo 2010; Injaian and Tibbetts 2015; Tibbetts *et al.* 2015; Webster *et al.* 2018). Sparrows and wasps live in dominance hierarchies. The status-badges on some sparrows and wasps then signal higher resource potential – the ability to win fights. Learning who has higher status can help prevent unnecessary fighting for dominance, like our example involving red deer from before. Honest signals would then benefit both sides. But there's still be an incentive to cheat. What stops dishonesty from evolving and undermining the effectiveness of status badges?

The consensus seems to be that in both cases when a competitor has some means for telling that the badge is misleading as to dominance, some form of punishment such as harassment or fighting will occur. The harassment has costs for foraging and caring for offspring, and thereby makes it less likely that deceptive status badges would evolve and destabilize the reliability of status badges. Status badges in the wild are then like black belts worn by martial artists. If you wear one, even those without black belts are likely to

⁹Though many theorists continue to recognize handicaps as a stabilizing mechanism, doubts have emerged. For criticisms of the Handicap Principle as an explanation (empirically, theoretically and methodologically), see Hurd (1995), Bergstrom and Lachmann (1997), Lachman *et al.* (2001), Huttegger and Zollman (2010), Grose (2011), Számadó (2012), Zollman (2013), Zollman *et al.* (2013) and Számadó and Penn (2015).

¹⁰Someone, of course, has to dole out the punishment, and punishing someone can prove costly. So a "higher-order" problem for the evolution of punishment now emerges. For discussion in the human case, see Richerson and Boyd (2005). For discussion in the animal kingdom, see Nakao and Machery (2012).

test you to see if it is genuine. If it is, they will back away. If not, they'll harass you until you take it off, or until they throw you out of the dōjō.

4. Green's mistake

Provided human assertive communication, at least in some contexts, faces a stability problem that it nevertheless solves, what mechanism solves the problem? Quality-handicaps or deterrents?

Green claimed that handicaps stabilize assertive communication, that handicaps make it possible for assertions to reliably provide evidence of what the speaker believes and has evidence for, and so reliably provide evidence for what is mostly likely to be true. Hence the title of his paper.

To say why assertions are handicaps, Green first reviewed Zahavi's idea, citing metabolic and predation costs to producing a signal, as I've done here, as examples of costs paid. He then defines handicaps as "signals ... that can only be faked with great difficulty as a result of being costly to produce" (Green 2009: 150–1). He then rehearses the stability problem:

[I]n the absence of some mechanism for vouchsafing honesty, [constative] speech acts will be prone to abuse by those who take assertion and other speech acts lightly ... This temptation threatens to make assertions, conjectures, etc., less worthy of our belief: Liars and those who say things on insufficient evidence threaten to undermine sincere speakers in the way that brightly colored but non-noxious frogs threaten the credibility of signals sent by those frogs who are both brightly colored and noxious. (Green 2009: 157)

Green then claims that handicaps "in our technical sense" stabilize assertions. "Assertions, conjectures, suggestions, presumptions, and the like ... carry a cost" that ensure the honesty, often enough, of our assertive speech acts (Green 2009: 157).

But what *handicaps* – what "waste" of a quality – does he identify? What are the strategic costs that the honest signaler can afford to spend at equilibrium that the dishonest signaler cannot? What are the costs that honest signalers pay that price dishonesty out of the market? It's certainly not metabolic costs – true assertions cost no more calories that false assertions.

This is where, Green says, norms governing assertion come into the picture. "In performing a speech act I incur a liability – a handicap in our technical sense" (Green 2009: 157). When we assert, our assertion is governed by a norm to the effect that we must assert that P only if we believe that P with very good evidence for P. If we assert that P but we're not in that condition, then we are liable to suffer a loss of credibility. The loss of credibility is then a cost that ensures reliability. The loss of credibility is the handicap that stabilizes assertion.

Our credibility comes under threat every time we perform a speech act requiring adequate justification, or whose content could be incorrect, or that could be insincere. Friendships, marriages, careers, company stock values, even governments can be brought down through a loss of credibility, and I will suggest that it is the threat of such loss that enables us to discern a connection between speech acts and handicaps. (Green 2009: 153–4)

When norms govern assertion, assertions "will be difficult to fake precisely because of the cost involved." When norms govern assertions, assertions are handicaps (Green 2009: 154).

If I have done my job, it should be evident that Green has made a mistake. The norms he's referring to – the norms of assertion – do not impose *strategic* costs on *honest* signalers that *honest* signalers *pay* in equilibrium that make dishonest signals of the same form too costly to produce or maintain. The norms he's referring to don't identify a *waste* of a *quality* designed to signal *quality*. Norms don't create handicaps *in the technical sense*; the norms of assertion are not *quality*-handicaps. Rather the norms of assertion are *deterrents*. They only impose costs on *dishonest* signalers paid *outside* of equilibrium; they do not impose strategic costs on honest signalers paid *in* equilibrium.

There's a good reason assertions and other speech acts in general can't be qualityhandicaps, for quality-handicaps can only reliably (stably) signal the quality they waste; that's how the mechanism works. But when we use language, we can signal – we can *mean* – just about anything. Zahavi's mechanism can't be the mechanism that explains the stability of human communication, at least not in general. Many commentators agree (e.g. Fitch 2010: 197–8). Lachman *et al.* put this point in terms of *arbitrariness*. The connection between signal and meaning in human languages is largely arbitrary, but the connection between quality-signals and the quality they mean is not (Lachman *et al.* 2001: 13192). The logic of quality-handicaps doesn't fit the category of linguistic communication.

Speech acts are also very cheap to produce (beyond the already paid costs required to develop the language faculty and to learn the local language). We only pay *efficacy* costs, which are rather low when the person you are talking to is standing right in front of you. That's another clear reason to doubt that speech acts are quality-handicaps. There seem to be no obvious candidates for *strategic* costs that honest signalers pay. No wonder, then, that Zahavi himself never applied his principle to human communication. He even said speech acts were not handicaps (Zahavi and Zahavi 1997: 80), that human communication is an exception to his rule. Though I admire Green's turn to the animal signals literature, the Handicap Principle isn't the place to look to explain the stability of assertion.

5. Social norms as deterrents

If handicaps rarely, if ever, contribute to the stability of assertive communication under conflict of interest, might deterrents play an important role? I think they do. In concluding I shall discuss one deterrent, though I think there are others. The deterrent I have in mind is the social norm for truth-telling (and related epistemic and relevance norms), with accompanying sanctions for violations. Green was right, I believe, to think there's a *norm* that, buttressed by sanctions, helps motivate the stability of assertion, given conflicts of interest. He was wrong to think the norm created a handicap instead of a deterrent.

What do we mean by norms? Here we are not talking about objectively valid principles that perhaps no-one may endorse or follow, at least not in the first instance. Rather we are talking about *our* norms – *social* norms – *our* principles that we embrace through normative attitudes, principles that we guide ourselves by and enforce through sanctions, whether internally when we sanction ourselves or externally when we sanction others (Bicchieri 2006, 2017; Brennan *et al.*, 2013). When norms exist in this sense, they enter into the calculus of costs and benefits that influence our decision-making, and so our behavior. Social norms in this sense often provide a counterweight to what otherwise might seem to be the thing to do. If I really need to pee while walking through an unfamiliar city, I might consider ducking into an alley. But if I also know of – and also embrace – the norm against urinating in public (even in alleys), then I'll stay on the hunt for a public restroom. This way I avoid potential internal sanctions (I won't feel guilty or ashamed) or potential external, third-party sanctions (Bicchieri 2006, 2017; Boyd and Matthew 2015; Boyd 2018).

With Green, among many others, I think social norms for honesty and competence weigh on us. Though it is true that we frequently have no reason to lie and don't even think about it – truth, as it were, directly flows from our lips – that is not always the case. After all, that's the very point we're addressing in this essay: we're not always angels. Norms prescribing honesty – norms that carry the cost of sanctions for non-compliance – then place their thumb on the scale. We may have so internalized them that we don't see them as a counterweight – we don't even think of lying. But when the advantages of lying or misleading start to go up, we often bring counter-considerations to mind, even if only in a murky way. They often make a difference, if not always.

For we punish violations for dishonesty. Green rightly emphasized loss of credibility. We frequently tarnish reputations through gossiping about other people. "Don't believe him, he can't be trusted," we might tell a friend. We might even be expected to gossip about others who we think can't be trusted (Enquist and Leimar 1993; Dunbar 1997). We punish in other ways as well. We might refuse help. We might require restitution. We might even banish the violator from our group. It happens.

Sometimes there are, of course, very good reasons to lie or mislead. Sometimes we should do it for moral reasons. Sometimes we can get away with it for selfish, baser reasons. Sometimes the scale favors dishonesty. But for all that norms, very frequently, enter the picture and make a difference. Norms for honesty enforced by sanctions then help deter dishonesty, and thereby help stabilize human communication (Silk *et al.* 2000; Faulkner 2011; Graham 2015, 2019). They don't always work, but even so they often do a lot of work.

Social norms can even change our preferences, through internalization. That is, we may not only come to follow a social norm as a cognitive shortcut to deliberation (Bicchieri 2017), we might also come to positively embrace the content of the norm as worth pursuing, as the right thing to do, as something we value for its own sake (Miller 2001; Sripada and Stich 2006; Bowles and Gintis 2011; Graham 2015). The *social* norm becomes our *personal*, individual norm as well. Potential conflicts of interest that pose the problem for the stability of communication may then become even less of a problem overall.

Green is to be applauded for turning to the animal signals and evolution of communication literature to find inspiration for understanding the stability, and so the reliability, of human communication, and so for understanding just how humans are able to express their attitudes and thereby perform assertive illocutionary acts. But he was wrong to turn to handicaps. Deterrents, not handicaps, explain how we pull it off. Among the class of deterrents stabilizing assertion, social norms play a central role.¹¹

References

Bach K. and Harnish R.M. (1979). Linguistic Communication and Speech Acts. Cambridge, MA: MIT Press.

Bergstrom C.T. and Lachmann M. (1997). 'Signaling Among Relatives: Is Costly Signaling Too Costly?' *Philosophical Transactions of the Royal Society of London B* 352, 609–17.

¹¹I am grateful to the referees for many helpful comments that led to improvements. I am also grateful to an audience at the University of Glasgow for equally helpful conversation that also led to improvements. I wrote this paper while a fellow at the Institute for Advanced Study in the Humanities at the University of Edinburgh. I am grateful for its supportive and collegial environment that helped me complete this paper.

Bicchieri C. (2006). The Grammar of Society. New York, NY: Cambridge University Press.

- Bicchieri C. (2017). Norms in the Wild. Oxford: Oxford University Press.
- Bowles S. and Gintis H. (2011). A Cooperative Species: Human Reciprocity and its Evolution. Princeton, NJ: Princeton University Press.
- Boyd R. (2018). A Different Kind of Animal: How Culture Transformed Our Species. Princeton, NJ: Princeton University Press.
- Boyd R. and Matthew S. (2015). 'Third-party Monitoring and Sanctions Aid the Evolution of Language.' Evolution and Human Behavior 36, 475–9.
- Brennan G., Eriksson L., Goodin R. and Southwood N. (2013). *Explaining Norms*. Oxford: Oxford University Press.
- **Cronin H.** (1991). The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today. Cambridge: Cambridge University Press.
- Darwin, C. (1871). The Descent of Man, and Selection in Relation to Sex. London: John Murray.
- Davis W. (2002). Meaning, Expression and Thought. New York, NY: Cambridge University Press.
- Dawkins R. (2016). The Selfish Gene. 40th Anniversary Edition. Oxford: Oxford University Press. [Originally published 1976.]
- Dunbar R. (1997). Grooming, Gossip, and the Evolution of Language. Cambridge: Harvard University Press.
- Enquist M. and Leimar O. (1993). 'The Evolution of Cooperation in Mobile Organisms.' Animal Behaviour 45(4), 747–57.
- Faulkner P. (2011). Knowledge on Trust. Oxford: Oxford University Press.
- Fitch T. (2010). The Evolution of Language. Cambridge: Cambridge University Press.
- Frazer B. (2012). 'Costly Signaling Theories: Beyond the Handicap Principle.' Biology and Philosophy 27, 263–78.
- Goldberg S. (2015). 'Recent Work on Assertion.' American Philosophical Quarterly 52(4), 365-80.
- Grafen A. (1990). 'Biological Signals as Handicaps.' Journal of Theoretical Biology 144(4), 517-46.
- Graham P.J. (2015). 'Epistemic Normativity and Social Norms.' In J. Greco and D.K. Henderson (eds), Epistemic Evaluation: Purposeful Epistemology, pp. 247–73. Oxford: Oxford University Press.
- Graham P.J. (2019). 'The Function of Assertion and Social Norms.' In S. Goldberg (ed.), *The Oxford Handbook of Assertion*. Oxford: Oxford University Press.
- Green M. (2009). 'Speech Acts, the Handicap Principle and the Expression of Psychological States.' Mind and Language 24(2), 139-63.
- Grice H.P. (1975). 'Logic and Conversation.' In P. Cole and J. Morgan (eds), *Syntax and Semantics*, Vol. 3, pp. 22–40. New York, NY: Academic Press.
- Grose J. (2011). 'Modeling and the Fall and Rise of the Handicap Principle.' *Biology and Philosophy* 26(5), 677–96.
- Higham J.P. (2014). 'How Does Honest Costly Signaling Work?' Behavioral Ecology 25(1), 8-11.
- Hurd P.L. (1995). 'Communication in Discrete Action-Response Games.' *Journal of Theoretical Biology* 174, 217–22.
- Huttegger S. and Zollman K.J.S. (2010). 'Dynamic Stability and Basins of Attraction in the Sir Philip Sidney Game.' *Proceedings of the Royal Society B* 277, 1915–22.
- Injaian A. and Tibbetts E.A. (2015). 'Advertised Quality and Resource Value Affect Aggression and Social Vigilance in Paper Wasp Contests.' *Animal Behaviour* **102**, 259–66.
- Johnstone R.A. (1995). 'Sexual Selection, Honest Advertisement and the Handicap Principle: Reviewing the Evidence.' *Biological Reviews* **70**(1), 1–65.
- Johnstone R.A. (1997). 'The Evolution of Animal Signals.' In J.R. Krebs and N.B. Davies (eds), *Behavioral Ecology: An Evolutionary Approach*, pp. 155–78. Hoboken, NJ: John Wiley and Sons.
- Johnstone R.A. and Grafen A. (1993). 'Dishonesty and the Handicap Principle.' Animal Behaviour 46, 215–34.
- Kirkpatrick M. (1986). 'The Handicap Mechanism of Sexual Selection Does Not Work.' American Naturalist 127(2), 222–40.
- Knight C. (1998). 'Ritual/Speech Coevolution: A Solution to the Problem of Deception.' In J. Hurford, M. Studdert-Kennedy and C. Knight (eds), Approaches to the Evolution of Language: Social and Cognitive Bases, pp. 68–91. Cambridge: Cambridge University Press.
- Lachman M., Számádo S. and Bergstrom C. (2001). 'Cost and Conflict in Animal Signals and Human Language.' Proceedings of the National Academy of Sciences USA 98(23), 13189–94.
- Laidre M.E. and Johnstone R.A. (2013). 'Animal Signals.' Current Biology 23(18), R829-33.

- Laland K.N. (2017). Darwin's Unfinished Symphony: How Culture Made the Human Mind. Princeton, NJ: Princeton University Press.
- Maestripieri D. (2012). Games Primates Play. New York, NY: Basic Books.
- Maynard Smith J. (1976). 'Sexual Selection and the Handicap Principle.' *Journal of Theoretical Biology* 57 (1), 239–42.
- Maynard Smith J. and Harper D. (2003). Animal Signals. Oxford: Oxford University Press.
- Mercier H. and Sperber D. (2017). The Enigma of Reason. Cambridge, MA: Harvard University Press.
- Miller S. (2001). Social Action: A Teleological Account. Cambridge: Cambridge University Press.
- Mokkonen M. and Lindstedt C. (2016). 'The Evolutionary Ecology of Deception.' *Biological Reviews* 91, 1020–35.
- Moller A.P. (1987). 'Social Control of Deception Among Status Signaling House Sparrows Passer domesticus.' Behavioral Ecology and Sociobiology 20, 307–11.
- Nakao Y. and Machery E. (2012). 'The Evolution of Punishment.' Biology and Philosophy 27, 833-50.

Pagel M. (2012). Wired for Culture: Origins of the Human Social Mind. New York, NY: Norton.

- Richerson P. and Boyd R. (2005). Not by Genes Alone: How Culture Transformed Human Evolution. Chicago, IL: University of Chicago Press.
- Rohwer S. (1977). 'Status Signaling in House Sparrows: Some Experiments in Deception.' *Behavior* 61, 107–29.

Saunders S. (2009). 'Costly Signaling: A Work in Progress.' Biology and Philosophy 24(3), 405-16.

- Scott-Phillips T.C. (2010). 'Evolutionary Stable Communication and Pragmatics.' In A. Benz, C. Ebert, G. Jäger and R. van Rooij (eds), Language, Games, and Evolution: Trends in Current Research on Language and Game Theory, pp. 117–33. Berlin: Springer-Verlag.
- Searcy W. and Nowicki S. (2005). *The Evolution of Animal Communication*. Princeton, NJ: Princeton University Press.
- Searle J. (1969). Speech Acts: An Essay in the Philosophy of Language. Cambridge: Cambridge University Press.
- Silk J.B., Kaldor E. and Boyd R. (2000). 'Cheap Talk When Interests Conflict.' Animal Behavior 59(2), 423–32.
- Spence M. (1973). 'Job Market Signaling.' Quarterly Journal of Economics 87(3), 355-74.
- Sperber D. (2001). 'An Evolutionary Perspective on Testimony and Argumentation.' *Philosophical Topics* 29, 401–13.
- Sperber D., Clément F., Heintz C., Mascaro O., Mercier H., Origgi G. and Wilson D. (2010). 'Epistemic Vigilance.' Mind and Language 25(4), 359–93.
- Sripada C. and Stich S. (2006). 'A Framework for the Psychology of Norms.' In P. Carruthers, S. Laurence and S. Stich (eds), *The Innate Mind II: Culture and Cognition*, pp. 237–56. Oxford: Oxford University Press.
- Stewart-Williams S. (2018). The Ape that Understood the Universe: How the Mind and Culture Evolve. Cambridge: Cambridge University Press.

Strassmann J.E. (2004). 'Rank Crime and Punishment.' Nature 432, 160-2.

- Számadó S. (2011). 'The Cost of Honesty and the Fallacy of the Handicap Principle.' Animal Behavior 81(1), 3–10.
- Számadó S. (2012). 'The Rise and Fall of the Handicap Principle: A Commentary on the 'Modeling and the Fall and Rise of the Handicap Principle'.' *Biology and Philosophy* **27**(2), 279–86.
- Számadó S. and Penn D.J. (2015). 'Why Does Costly Signaling Evolve? Challenges with Testing the Handicap Hypothesis.' Animal Behaviour, 110, e9-e12.
- Tibbetts E.A. and Dale J. (2004). 'A Socially Enforced Signal of Quality in a Paper Wasp.' *Nature* 432, 218–22.
- Tibbetts E.A. and Izzo A. (2010). 'Social Punishment Mediated by Signal Incongruence.' *Current Biology* 19, 1637–40.
- Tibbetts E.A., Forrest T., Vernier C., Jinn J. and Madagame A. (2015). 'Socially Selected Ornaments and Fitness: Signals of Fighting Ability in Paper Wasps are Positively Associated with Survival, Reproductive Success, and Rank.' *Evolution* **69**, 2917–26.

Tomasello M. (2008). The Origins of Human Communication. Cambridge, MA: MIT Press.

- **Turri J.** (2016). *Knowledge and the Norm of Assertion: An Essay in Philosophical Science*. Cambridge: Open Book Publishers.
- Veblen T. (1899). The Theory of the Leisure Class. London: Macmillan.
- Von Hippel W. (2018). The Social Leap: The New Evolutionary Science of Who We Are. New York, NY: Harper Collins.

- Webster M.S., Ligon R.A. and Leighton G.M. (2018). 'Social Costs are an Underappreciated Force for Honest Signaling in Animal Aggregations.' *Animal Behaviour* 143, 167–76.
- Williamson T. (1996). 'Knowing and Asserting.' Philosophical Review 105(4), 489-523.
- Zahavi A. (1975). 'Mate Selection: A Selection for a Handicap.' *Journal of Theoretical Biology* 53(1), 205–14.
- Zahavi A. (1977). 'The Cost of Honesty.' Journal of Theoretical Biology 67, 603-5.
- Zahavi A. (2007). 'Sexual Selection, Signal Selection and the Handicap Principle.' In B. Jamieson (ed.), Reproductive Biology and the Phylogeny of Birds, Part B: Sexual Selection, Behavior, Conservation, Embryology and Genetics. Boca Raton, FL: CRC Press.
- Zahavi A. (2008). 'The Handicap Principle and Signaling in Collaborative Systems.' In P. d'Ettorre and D.P. Hughes (eds), *Sociobiology of Communication: An Interdisciplinary Perspective*, pp. 1–10. Oxford: Oxford University Press.
- Zahavi A. (2010). 'Profile: The Handicap Principle and Social Behavior.' In T. Székely, A.J. Moore and J. Komdeur (eds), *Genes, Ecology and Evolution*, pp. 535–7. Cambridge: Cambridge University Press.
- Zahavi A. and Zahavi A. (1997). The Handicap Principle: A Missing Piece of Darwin's Puzzle. Oxford: Oxford University Press.
- Zollman K.J. (2013). 'Finding Alternatives to Handicap Theory.' Biological Theory 8(2), 127-32.
- Zollman K.J.S., Bergstrom C.T. and Huttegger S.M. (2013). 'Between Cheap and Costly Signals: The Evolution of Partly Honest Communication.' *Proceedings of the Royal Society B: Biological Sciences* 280, 20121878.
- Zuk M. & Simmons L.W. (2018). Sexual Selection: A Very Short Introduction. Oxford: Oxford University Press.

Peter J. Graham is Professor of Philosophy and Linguistics at the University of California, Riverside. He has published articles on proper names and singular terms, testimony and assertion as speech acts, testimonial knowledge transmission and testimonial warrant, the role of social norms in explaining the reliability of communication, the role of social norms in explaining our epistemic judgments about what we ought to believe and do, the role of social norms in explaining our epistemic behavior, the nature of functions and their role in explaining the reliability of communication and perception, and the nature of epistemic warrant, especially epistemic entitlement. He is a co-editor of the *Routledge Handbook of Social Epistemology* (Routledge, 2019) and *Epistemic Entitlement* (Oxford, 2020).

Cite this article: Graham PJ (2020). Assertions, Handicaps, and Social Norms. *Episteme* 17, 349–363. https://doi.org/10.1017/epi.2019.53