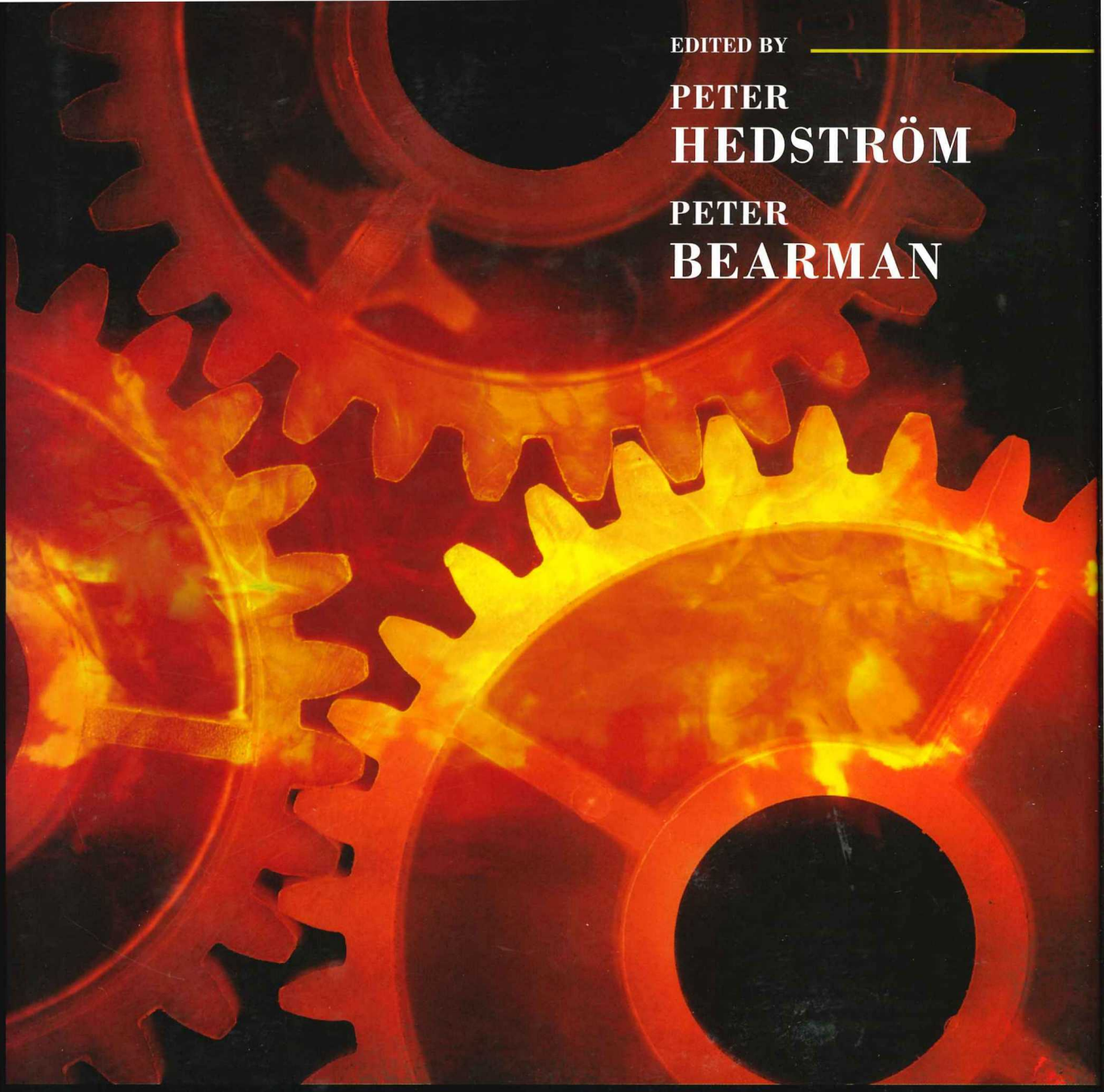

EDITED BY

PETER HEDSTRÖM

PETER BEARMAN

The Oxford Handbook *of* ANALYTICAL SOCIOLOGY

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2009

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
The Oxford handbook of analytical sociology/edited by
Peter Hedström and Peter Bearman.
p. cm.
Includes bibliographical references.
ISBN 978–0–19–921536–2
1. Sociology. 2. Sociology–Philosophy.
I. Hedström, Peter. II. Bearman, Peter.
HM585.0984 2009
301–dc22 2009013905
Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire
ISBN 978–0–19–921536–2
1 3 5 7 9 10 8 6 4 2

Acknowledgments

This book would not have come into existence without the generous support of several individuals and institutions. We wish to thank Fletcher Haulley at Columbia for his managerial and editorial work and for his careful attention to detail. We also want to thank the Yale School of Management for organizing one of the workshops at which drafts of the chapters were presented, the participants at these workshops for their valuable comments, and the Rockefeller Foundation's Bellagio Center for providing peace and tranquility at a critical juncture. Finally we want to thank our home institutions—the Institute for Social and Economic Research and Policy at Columbia and Nuffield College at Oxford—for partial funding, and Dominic Byatt at Oxford University Press for the detailed feedback we have received from him throughout the project.

CHAPTER 8

SIGNALING*

DIEGO GAMBETTA

INTRODUCTION

SIGNALING theory (ST) tackles a fundamental problem of communication: How can an agent, the receiver, establish whether another agent, the signaler, is telling or otherwise conveying the truth about a state of affairs or event which the signaler might have an interest to misrepresent? And, conversely, how can the signaler persuade the receiver that he is telling the truth, whether he is telling it or not? This two-pronged question potentially arises every time the interests between signalers and receivers diverge or collide *and* there is asymmetric information; namely, the signaler is in a better position to know the truth than the receiver is.

The state of affairs or event about which communication can occur could be anything. As it happens, in the literature it often has to do with the signaler's unobservable qualities, or, as it is sometimes said, her 'type'; for example, in a famous early study by Michael Spence (1974) the quality refers to a worker's productivity, which can be either of two types, high and low, and which the employer cannot directly observe. An 'honest' signaler would like to persuade the receiver that he is the good type, but the problem is that a 'dishonest' signaler too would like to do the same. How can the receiver decide and what should the signaler do?

Signaling theory, which is only a little more than thirty years old, has now become a branch of game theory. In economics it was introduced by Michael Spence in

* I am grateful to Macartan Humphreys, Ekaterina Korobtseva and Margaret Meyer for their specialized advice and to Michael Biggs, Valeria Pizzini-Gambetta, and the editors for their comments on an earlier version of this chapter.

1973. In biology it took off not so much when Amotz Zahavi first introduced the idea in 1975, but only since 1990, when Alan Grafen proved formally through game theory that 'honest' signals can be an evolutionarily stable strategy (ESS).[1] In both biology and microeconomics it has had considerable success and is now found in textbooks (e.g. Kreps 1990: ch. 17; Krebs and Davies 1998: ch. 14). Since its inception it has attracted a robust following in political science and more recently also among anthropologists, especially those with an interest in evolution. It is hard to think of any other theory that in recent times has been developing so fast across *all* behavioral sciences. Unfortunately, despite some lonely practitioners, sociologists are still largely oblivious to its existence. However, the theory's potential to make sense of many human interactions commonly investigated by sociologists is arguably large and the theory should become an important element in the toolbox of analytical sociology.

In the rest of the Introduction I will describe some key concepts. In Section 8.1 I will introduce the basic principles in nonformal terms. In Section 8.2 I will illustrate different ways in which the cost of the signal and the quality being signaled are related, and mention some of the theory's limitations. In Section 8.3 I will then briefly introduce some of its applications and lastly in Section 8.4 provide some information on its origins. This chapter is a mere introduction to a complex topic, and makes no claim to be exhaustive of all that scholars should know before using it properly.

Unobservable properties and signs

Much of what is important to know about others is not easily knowable. People have properties that affect how we decide to interact with them: loyalty, fidelity, trustworthiness, innocence, generosity, resilience, fighting ability, fertility, tolerance, determination, competence, intelligence—to name some examples. People also have intentions towards us that we would like to know before deciding how to interact with them.

The problem however is that these properties are not observable. We cannot directly see or otherwise perceive people's honesty or resilience. We cannot read their intentions as if they were written in an open book. They know and we do not. Sometimes we do not even know if they are who they say they are. Identity too can be an unobservable property. The problem extends to the goods that people trade and whose quality can be established only through experience. But how can we decide prior to experience whether to buy them or how can we persuade others that our goods are of good quality?

We can however perceive (and display) other things. We can see if people are handsome, if they smell nice, and if their voice sounds relaxed. We can observe how they move or dress, whether they have tattoos, and with whom they go out, how

much money they spend and on what items, whether their hands are smooth or callous, and whether they blush easily or look us in the eye. We can see what they do or gather evidence on what they have done in our absence. We can memorize their individual traits, such as face or email address, and we reidentify them on seeing those traits again. We can of course hear what they say and the accent with which they say it.

Sometimes what we perceive conveys no information on properties that interests us; sometimes we mistake an innocent gesture for a sign or a signal of some property. Still, our best chance to find out something about people's unobservable properties is by establishing a connection between their perceivable features and their unobservable properties. Whenever interests are potentially at odds, this connection is the object of signaling theory.

Signs and signals

People's perceivable features come in two forms: signs and signals. The difference and links between the two are complex, but in the limited space I have I will only mention the main distinctions, drawing on Bacharach and Gambetta (2001).

Signals are the stuff of *purposive* communication. Signals are any observable features of an agent which are intentionally displayed for the purpose of raising the probability the receiver assigns to a certain state of affairs. 'Features' of an agent that make up a signal could be anything: they include parts or aspects of his body, pieces of behavior by him, and his appurtenances.

Signs are a different concept from signals. Signs can be just anything in the environment that is perceptible and by being perceived happens to modify our beliefs about something or someone. But signs are dormant potential signals. They are the raw material of signals. The basic form of the sign–signal transformation is that a signaler takes steps to display the sign. We cannot take for granted that signs are noticed. A duelling scar may be not on the face but the thigh or chest. The crowded tables of a restaurant may be invisible from the street unless we devise some system for outsiders to see them. One way of signaling is to take steps to make apparent a sign that would not otherwise be observed: to bare the chest to display a tattoo, to glaze the restaurant façade to reveal the crowds inside. One trigger of this transformation is the bearer's realization of the meaning of certain actions in the eyes of an observer. I may be unaware that my accent is informing others of some quality of mine, until some observer acts in a way that makes me aware and at which point I may choose to display it intentionally. We often produce the raw material of signals innocently by going about our lives, without planning for them to become signals. Most of our actions and observable traits will never become signals, but some will. Even checking one's email on one's home computer at a certain time

can later become a verifiable alibi we can display, should we be accused of having perpetrated a crime somewhere else at that time.

The transformation goes also in the opposite direction and yesterday's signal can become today's sign. Consider a tattoo on my wrist, which induces beliefs in others that I am a sailor. Once I have had it done, in each subsequent encounter with someone it is a fait accompli and so a sign. I am not thinking about its effects every time someone sees it. But the day I chose to have it done, the action I took was a signal. Moreover, from the perspective of that day, on which I foresaw and intended an endless string of tattoo-induced beliefs in my nautical quality, it was a signal on all these future occasions too. So whether a tattoo is a signal or a sign on a particular occasion depends on the temporal perspective from which the agent regards that occasion.

> If the same signalling situation arises day after day, what is at first a conscious intention tends to become habitual and internalised. But it remains an intention for all that, on this simple test: were one to be made aware of the issue—for example, by losing a suitcase containing all one's clothes—one would choose to re-outfit oneself with similar items—for example, with new clothes in the same style—with the explicit intention of signalling one's qualities. In short, the intention does not have to be 'occurrent', but only 'triggerable'. So conforming to the dress codes and outward practices of one's group can be explained as a broadly intentional act, with the purpose of signalling membership of that group. (Bacharach and Gambetta 2001: 174)

Finally, it is important to notice that the understanding of signals in biological models does not require intentionality, but rests on natural selection and can be applied to all organisms (Searcy and Nowicki 2005: 3–5). While arguably many signals among humans are intentionally emitted and responded to, humans too like any other animals develop, display, and respond to behaviors and morphological features which have communicative value unthinkingly. Thus, for instance, it seems that human males prefer females with a higher waist–hip ratio (see e.g. Singh and Young 2001).[2] This is a 'preference' they manifest in very different human groups without knowing its source; namely, that it would be a signal of fertility with which the ratio would be positively correlated. Those with that preference would be rewarded by more offspring, which via genetic transmission are more likely to carry that preference, which will then spread in the population. In the biological sense, the ratio and the response are considered a 'signaling system.' Unlike other animals, though, humans can learn, and unconscious signals can potentially shift to a greater or lesser extent under the domain of culture and intentionality. Women, for instance, can learn that the waist–hip ratio makes them more attractive without knowing why and can take steps both to display it 'honestly' and more effectively by wearing figure-hugging dresses and to enhance it by squeezing their waist in a corset or bustle, in which case they would silence the biological signaling system as the waist–hip ratio would no longer be a reliable signal.

8.1 The Theory's Basic Principles

When we believe that our interests and those of others are identical, establishing a connection between signals and states of affairs can be relatively easy. In many of our exchanges we simply believe what others want us to believe by what they do or say. This is the case when we deal with an individual with whom we have reasons to believe that we share our interests either generally or in a given situation.

A child who is told by his father to jump from a wall—'Jump, I'll catch you'—has no reason to disbelieve his father. The father's words mean what he says; they describe his true intentions because of who he is. The underlying problem here is one of coordination: both need to understand given utterances or gestures in the same way. They need to know the conventions, whether linguistic or related to other gestures. The main threat to communication here is *misunderstanding*.

Yet, as a Sicilian anecdote indicates, even in this case one cannot be entirely sure. A Mafioso once said to his son 'Jump, I'll catch you!' and when the son did he let him fall to the ground. 'Now,' he said, 'you'll learn that you cannot trust anyone.' This grim lesson reminds us of the threat of misrepresentation, by which others wilfully lead us to make the wrong connection: they emit false signals. The son understood the meaning of the message correctly and acted accordingly, but the message was false. Whether people truly share their interests with us is also not directly observable, itself a property that may be misrepresented.

When our interests are not necessarily identical, as in courtship or in business, or worse they openly clash, as in conflicts, the risk of strategic misrepresentation is potentially rife. Can we believe the pitiful story of a beggar asking us for money in the street, a suitor promising marriage, or that Iraq has weapons of mass destruction? From the minor daily encounters to the grand interactions between nations, the question of what to believe is ubiquitous. We cannot trust what people say or do to persuade us that they are trustworthy by default; or when they try to persuade us that it is not to our advantage to fight them because they are stronger than us. And yet not all signalers are lying. The honest among them—by 'honest' here I simply mean that they mean what they say or signal to us—are eager for us to know that they are honest. Is there a way in which truth can be transmitted even when interests are not aligned and knowledge is asymmetrical?

Typical situations that signaling theory covers have two key features:

(i) There is some action the receiver can do which benefits a signaler, whether or not he has the quality k, for instance marry him, but
(ii) this action benefits the receiver if and only if the signaler truly has k, and otherwise hurts her—for instance, marry an unfaithful man.

This applies to conflict situations too: if we know that our opponent is going to win a fight, we may choose to yield without fighting at a lesser cost for both. Thus

k signalers and receivers share an interest in the truth, but the interests of non-k signalers and receivers are opposed: non-k signalers would like to deceive receivers into thinking they have k, in order to receive the benefit, while receivers have an interest in not being deceived. (The interests of ks and non-ks are also usually opposed because the activity of the latter damages the credibility of the signals of the former.)

The main result in signaling theory is that *there is a solution in which at least some truth is transmitted, provided that among the possible signals is one, s, which is cheap enough to emit, relative to the benefit, for signalers who have k, but costly enough to emit, relative to the benefit, for those who do not. If s is too costly to fake for all or most non-k signalers then observing s is good evidence that the signaler has k.*

If the cost relationship is such that all and only k signalers can afford to emit s, the solution in which they do so, often referred to in the literature as 'equilibrium,' is called 'separating' or 'sorting.' In this equilibrium signals are unambiguous, and the receiver is perfectly informed. All those with k will be perfectly separated from those without k by being able to emit s. No poisoner (normally) seeks to demonstrate his honesty by drinking from the poisoned chalice. Drinking is a signal that one is not a poisoner. At the opposite extreme both k and non-k signalers can afford to emit s relative to their respective benefits. In this case s is uninformative and the equilibrium is said to be 'pooling'; that is, once s is emitted the receiver knows nothing more about who has and who has not k than she did before, and should disregard s. (In Monty Python's *Life of Brian* there is a humorous rendition of a pooling signal. The Romans accept the mob request to free Brian, who is being crucified with a dozen others, and ask 'Who's Brian?' Brian answers 'I am Brian,' but so do all the others. The Romans are none the wiser and Brian is crucified.)

The cost condition may also give rise to intermediate equilibria, so-called semisorting ones. In a semisorting equilibrium there is a signal s which is emitted by all k signalers, but not only them; a certain proportion of non-k signalers emit it too. Here although observing s is evidence for the receiver in favor of k, it is not conclusive evidence; it makes it more likely that the signaler has k, but does not imply that he does. The higher the proportion of non-k signalers who use this signal the less conclusive is the evidence.

Signals that approach near perfection exist but are rare. Architects set up offices in the early skyscrapers to show that they trusted them not to collapse. Boatbuilders sailed on the first voyages to show they trusted their boats not to sink. Tim Spicer, a mercenary chief, said he always fought alongside his men to show his trust in their quality.[3] However, most signals in real life are only semisorting: they inform, but not perfectly. We rarely encounter a fully mimic-proof signal. Most of the time someone looking like a Hasidic Jew will be a Hasidic Jew. But occasionally he is a Palestinian suicide bomber disguised as one. Virtually everybody who boards a plane gives a signal, mostly unthinkingly, that he is not intent on causing it to

crash. But, as we know only too well, suicide terrorists may be prepared to do just that and can afford to mimic a normal passenger simply by boarding. Some can drink from a poisoned chalice.

When the contamination of the signal is partial it remains credible enough for the majority of honest signalers to keep using it and for a minority of mimics to gain from using it too. In it, at least some truth can be transmitted. In many instances weak signals induce receivers to probe and seek more credible signals, or k signalers to spend resources protecting their signals from the threat posed by dishonest non-k signalers. When the contamination is complete the signal stops being informative, and rational receivers should ignore it.

8.1.1 A classic example

To understand how signaling theory works I will now discuss a classic case of a perfectly discriminating signal reported by Livy, the Roman historian. The Etruscans were besieging Rome, and a brave man known as Caius Mucius infiltrated the enemy's camp aiming to kill Porsena, the King of the Etruscans.

> Afraid to ask which of the two was the king, lest his ignorance should betray him, Mucius struck as fortune directed the blow and killed the secretary instead of the king. . . . He was seized and dragged back by the king's bodyguard to the royal tribunal. Here, alone and helpless, and in the utmost peril, he was still able to inspire more fear than he felt.

Rather than being cowed, Mucius threatens Porsena, hinting at the fact that many more like him are queueng up to try and kill him.

> The king, furious with anger, and at the same time terrified at the unknown danger, threatened that if [Mucius] did not promptly explain the nature of the plot which he was darkly hinting at he should be roasted alive. 'Look,' Mucius cried, 'and learn how lightly regard their bodies those who have some great glory in view.' Then he plunged his right hand into a fire burning on the altar. Whilst he kept it roasting there as if he were devoid of all sensation, the king, astounded at his preternatural conduct, sprang from his seat and ordered the youth to be removed from the altar. 'Go,' he said, 'you have been a worse enemy to yourself than to me . . . I send you away exempt from all rights of war, unhurt, and safe.' (Livy 1912: 2. 12)[4]

The case of Mucius, who later gained the nickname Scaevola (left hand), shows us the basic principle at work, and makes very clear all the fundamental elements that need to be identified to describe a genuine signaling episode.

The main characters: Mucius is the signaler and Porsena the receiver.
The property: Porsena cannot observe Mucius' resistance to pain for the sake of loyalty to his Romans compatriots.

The signaler's gain if truth is transmitted: Mucius' interest is to avoid being tortured or killed.
The receiver's gain if truth is transmitted: Porsena's interest is to know Mucius' quality to avoid torturing him in vain.
The information asymmetry: Mucius knows whether he has the property, but Porsena does not.
The weak signal: Mucius could just say that he will not give in to torture. But words are cheap in this case, they do not meet the cost condition. Porsena knows that anyone could say that, and would in the circumstance. The question for Mucius is: Is there a signal that I can afford which is less costly than being tortured to death, but which is such as to leave Porsena in no doubt? Is there a signal that someone who merely pretends to have the resilience that I have could not afford?
The persuading signal: Incinerating his hand is just that signal. Mucius loses his right hand but keeps his life and honour, while Porsena, who is neither vengeful nor a sadist, avoids a pointless act. Mucius pays a high cost, in terms of pain and of inflicting a permanent handicap on himself, but still better than open-ended torture that could result in death. Both gain given the situation.
The cost condition: The high cost endured by Mucius is what persuades Porsena. 'If Mucius can do that to himself then there is little more that I can do to him,' he infers. 'He would die rather than betray his countrymen.' Cost is the crucial variable. Or, more accurately, it is the difference between that cost and the cost that a hypothetical mimic could afford to pay. When such a difference cannot be bridged, the signal is perfectly discriminating. No feeble man pretending to be tough could have endured Mucius' feat.

Any analysis that fails to identify *all* the above elements—and in the several sloppy applications that exist this is often the case—cannot with any certainty claim to be truly describing a signaling episode.

8.1.2 Differential costs

When considered carefully a signaling episode has at least two notional signalers rather than one. Even if the receiver faces just one person, he is uncertain whether that person is the one that has k or is the mimic, the bad or low-quality type who claims he has k but does not. The problem for the receiver is to distinguish between them, and for the honest signaler the problem is to make sure that he is not confused with the mimic.

Suppose there are two notional signalers, the Mucius type and a weaker type, let's call him Lucius. Porsena does not know whether the man in front of him is the Mucius or Lucius type. If he is the Lucius type then it pays Porsena to torture

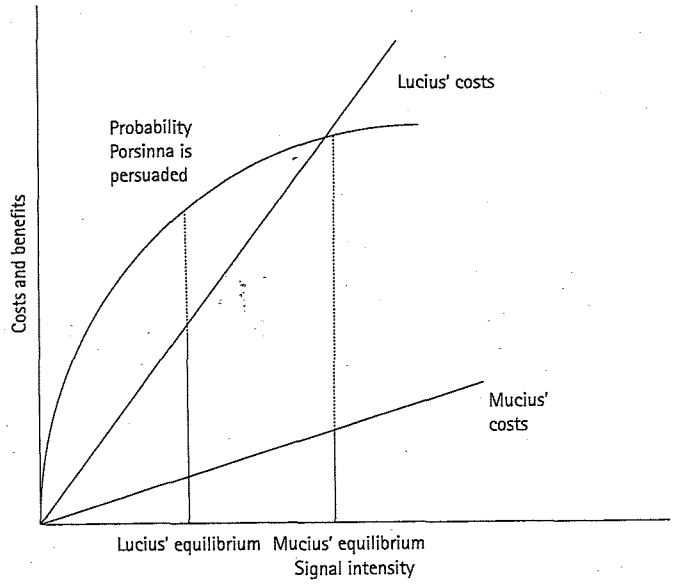

Fig. 8.1 A signaling model in which two signalers face different costs to emit the same signal intensity (after Johnstone 1997)

him, for he may obtain information on the Romans' plans, while it does not pay to torture Mucius. Both types would like Porsena to believe that they are resilient to torture.

Suppose there is a continuous signal whose cost for the signaler is positively correlated with the property the signaler wants to signal: the longer you can keep your hand in a burning fire the higher the cost and the more intense the signal becomes. The Mucius type is more resilient, and this implies that for him the same length of time with his hand in the fire is less costly than it is for the weaker type. Lucius can burn the tip of his fingers, Mucius a whole hand. Lucius and Mucius are on two different cost lines, as represented in Figure 8.1. For the same intensity they have different costs, Mucius can thus emit a more intense signal, a signal that Lucius cannot afford.

In Figure 8.1 I represent the signaler's benefits as a continuous variable for ease of exposition and generality. One can think of it as the probability that Porsena assigns to facing a type who is resilient to torture rather than someone who just pretends to be, a probability that grows as the signal intensity grows. The higher

that probability is, the higher are also the probability that Porsena will choose not to torture and the related expected benefit for the signalers. If both signalers are maximizers, they will stop at their equilibrium point, where the distance between their respective cost line and the benefit line peaks (the dotted lines in Fig. 8.1). In the situation depicted in the graph, even if Lucius pushed it to the limit, to the right of his equilibrium line, he would not be able to reach Mucius' signaling intensity at equilibrium, as it would be unbearable for him. On witnessing the highest level of signal intensity Porsena can rest assured that he is facing the Mucius type.

8.1.3 Differential benefits

Let us consider now the case in which the relationship between signal cost and signal intensity is the same for both types of signalers, whereas the benefit of signaling rises more rapidly for the signaler who has a higher need and receives more satisfaction from the receiver's response. Does it mean that the cost condition does not obtain and thus informative signaling cannot occur?

Suppose we have a loving creature, Don Ottavio, and a philanderer, Don Giovanni. Don Ottavio is more desirous of Donna Anna and in greater need than Don Giovanni to quell his loving thirst. Donna Anna is wondering who truly depends on her love most, and knows that both of them want to persuade her that they are the one. Following the conventions of their time, they manifest their desire by serenading Donna Anna. Both are equally skilled at serenades, so the cost is the same for them. However, Don Ottavio has an advantage for, given the higher benefits he expects, he can afford an extra cost and serenade her more and more frequently. We can think of the benefits in terms of the warmth of Donna Anna's response to her suitors which goes from ignoring to flirting and lastly to having sex.

In Figure 8.2 Don Giovanni could afford to do just as much, but if he is a maximizer, as we expect a Don Giovanni to be, he will serenade her only up to his equilibrium point in which the distance between the cost and the benefits for him is at its maximum. By contrast, Don Ottavio can serenade her more not because it is easier for him but because he gains more from her positive responses. The differential cost condition, the key to the reliability of the signal, is still achieved. So if Donna Anna wanted to yield more to the suitor who most depended on her love and from whom she can thus expect more loyalty, on observing the differential frequency of serenading she should decide that Don Ottavio is her man.

This is an illustrative example and it does not quite reflect how the story goes in Mozart's opera! It also does not take into account that persistence may sort the opposite effect and be taken as a sign of desperation. Yet there is some evidence that women looking for a long-term mate are responsive to 'serenading' persistence—in

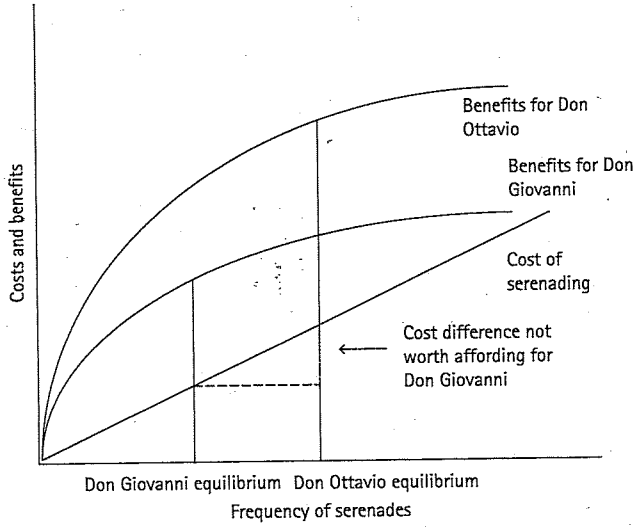

Fig. 8.2 A signaling model in which two signalers obtain different benefits from the same receiver's response (after Johnstone 1997)

the form of 'spending a lot of time with the woman, seeing her more often than other women, dating her for an extended period of time, calling her frequently on the phone, and writing her numerous letters' (Buss 2003: 102).[5]

8.2 Sources of Signal Cost

In order to identify a signaling episode it is essential to know the link between the property being signaled and the cost of the signal. There are several types of links, the main ones of which I will now describe.

8.2.1 Receiver's independent cost: 'wasting' key resources

In animal behavior 'the concept of signal costs requires that signaller fitness go down as the signalling goes up' (Searcy and Nowicki 2005: 14); in other words, the

signaling animal has to sacrifice or, as it is sometimes said, 'waste' resources, and the resources sacrificed have to be correlated with the information which is to be conveyed. A classic example is that of the stotting gazelle when detecting a lion on the hunt. Gazelles that stot, that is jump up and down, show to the lion that they have lots of physical energy, so much of it in fact that they can 'waste' it in display. The gazelle is signaling that it can outrun the lion should it choose to give chase (Zahavi and Zahavi 1997: 6–7). The lion, if persuaded, will be less likely to chase an energetic gazelle, for stotting is correlated with running potential. (The 'honesty' of the signal is maintained provided that lions occasionally put it to the test.)

A classic human example of a signal with this source of cost can be found in Ovid's *Fasti*, in which the poet describes how the Romans came to worship Jupiter Pistor or Jupiter the Baker. When the Gauls were besieging Rome in 387 BC Jupiter instructed the Romans 'to throw down among the enemy the last thing you'd wish to yield!' The Romans

> shook off sleep, and troubled by the strange command, asked themselves what they must yield, unwillingly. It seemed it must be bread: they threw down the gifts of Ceres [loaves of bread], clattering on the enemy helms and shields. The expectation that they could be starved out vanished. The foe was repulsed, and a bright altar raised to Jove the Baker. (bk. 6: June 9: The Vestalia)[6]

Here the cost of the signal to the signaler comes from the fact that he 'wastes' in display the precise resource he would need most were the receiver to attack, 'the last thing you'd wish to yield'—energy for the lion-stalked gazelle, bread for the Gaul-besieged Romans. (The difference is that the Romans took a desperate risk and bluffed!) These examples are about conflict, but wasting the resource the property of which one wants to signal applies more broadly. The classic Veblen's case, which I will mention below, comes under this heading: a rich man who wants to show that he is rich can choose conspicuous ways to 'waste' his money, for instance.

8.2.2 Receiver-dependent cost: exposure to risk

Another category of signal cost is, by contrast, *dependent* on the receiver's response. Here the signal does not require an expenditure of resources. Consider the alarm calls that many animals, birds, and marmots for instance, emit on spotting a predator. The traditional interpretation of this behavior is altruism: the animal alerts its companions of the impending danger. However, Amotz Zahavi (1997: ch. 12), one of the theory's founders, suggested a different interpretation of the call compatible with self-interest: the warning effect would be a by-product while the primary motive would be to inform the predator that one has seen it and that one is not

scared. By increasing the risk of predator detection and attack, the caller would signal that it has confidence in its ability to escape or fight back, thereby reducing the risk of an attack.

To understand the difference between receiver-dependent and receiver-independent costs, consider two types of signals aimed at dissuading opponents from fighting. One consists in threatening the opponent by moving very close to him or in showing one's back to him, either way signaling that one is not afraid of the opponent's attack, so much so that one makes it easier for the opponent to strike. Here there is no loss of resources, the cost is merely potential and due to the increased risk the individual incurs of suffering should the receiver attack. In the other case one can try to dissuade the opponent by tying one's hand behind one's back and thus depriving oneself of a resource which would be essential for one to fight. In both cases one shows confidence in one's ability to overcome the opponent, in the former by facilitating the opponent's attack, in the latter by handicapping oneself.[7] (The name used in zoology to refer to signaling theory, *the handicap principle*, is derived from this type of case and thus to receiver-independent costs.)

8.2.3 Third party-dependent cost

I do not know if it occurs among other animals, but among humans we also have signals the cost of which is dependent on a third party's response, which the communicating agents can exploit. Divine Brown, a Los Angeles prostitute who achieved her fifteen minutes of fame for administering oral sex to British actor Hugh Grant in 1995, revealed her particular kind of test to make sure that the prospective customer was not an undercover policeman. Before agreeing to trade, she asked Grant to expose himself in the street. The reason, she said, is that an undercover policeman would not do that. Divine believed that exposing oneself is an action with information value because only a real customer can afford to do it while an undercover policeman would not break the law for he would risk losing his job. In this case the signal cost for the 'dishonest' signaler is imposed by a third party, the law and the police authorities.

The same reasoning inspires a tougher test applied by drug dealers in New York. Since the mid-nineties

> as police have intensified their assault, the dealers have also adopted more perilous tactics. Five or six times each month, undercover investigators are now forced to use cocaine or heroin at gunpoint, to prove to dealers that they can be trusted. At least twice a month, an officer [who refuses] is shot or otherwise wounded during a staged purchase, say police commanders, who spoke on condition of anonymity.
>
> (*New York Times*, 21 January 1998; Gambetta 2009: ch. 1)

8.2.4 Multiple sources of cost

We often find signals that have more than one source of costs. These are harder to handle, for they can signal more than one property to more than one type of receiver. In animals the classic example of multiple costs is the peacock's tail: costly to grow in terms of developmental resources and nutrients, costly to fan and display, costly to maintain free of parasites, costly to carry, as it both makes the animal more noticeable to predators and hampers its ability to run away. It could serve to attract mates as well as to discourage predators. For humans, consider the tattoo as an example of multiple costs: when extensive and done according to traditional needle techniques, tattoos are produced at the cost of some considerable pain, a cost independent of receivers, which testifies to a certain endurance of the individual, à la Mucius. However, a tattoo can also generate cost from the receiver's response. If it cannot be washed out or veiled and if it is conventionally associated with belonging to a gang, the display of the tattoo can increase the likelihood of attacks by rival gangs, thus testifying to the signaler's loyalty to the gang and disregard for danger. (Here we can appreciate how even a name or a logo, once permanently etched on someone's body, can in certain conditions be costly to display. Not all words or symbols are necessarily cheap.)

8.2.5 Signals costless to honest signalers

A frequent mistake is to think that only signals that are costly for the honest signaler are informative. We often read sentences such as 'Signaling theory argues that expense of the signal ensures the honesty of that which is being signaled,' and the literature refers to 'costly' signals as being persuasive. In many cases, however, persuasive signals have little or no cost for the genuine possessors. We need to bear in mind that it is not the absolute cost per se that informs, but the cost differential between what the k signaler can afford relative to what the non-k signaler can afford (and both costs relative to the respective benefits obtained from the receiver's response if the receiver is persuaded). Only if this difference is large enough to discriminate between the two does the signal inform well. In many cases we obtain that difference perfectly well with costless signals, which work because it is the dishonest signaler who cannot afford them. Social life is interwoven with cheap yet reliable signals, so reliable in fact we do not even notice. If the property to be signaled is your identity, the cost of showing that it is you, if it is you, can be negligible. If I say on a low-quality interphone 'Hi, it's Diego' this is costless for me, but is also uninformative, for others could say that too. But if the sound transmission is good, then my statement will come with my voice and accent clearly attached, and this again is still costless for me but more informative, for it would

be more costly for someone else to imitate my voice and accent accurately. And if the medium of communication allows it I can just do what no one else can: show my face, which will be still costless for me and impossible to do for mimics. If being a woman works as a signal of nonaggression, for example, then the cost of being one and displaying it are high only for men who would like to mimic 'woman-ness,' while costless for genuine women. Drinking from a nonpoisoned chalice has no cost for the nonpoisoner. In short, not all honest signals imply a cost for the honest signaler.

8.2.6 Knowing costs, benefits, and incidence of deception

Signaling theory is a powerful tool that can deliver specific answers to the questions when and how much false signaling will occur. However, crucial for the existence of the solutions which it offers are strong assumptions about the players' background knowledge: they must know the sizes of the benefits and of the cost of emitting s, for both k signaler and non-k signalers; and a receiver must also know the base-rate probabilities that a signaler is k and non-k.

> When she peers through the Judas hole and sees a stranger in a UPS uniform, as she wonders whether to confer the benefit of opening the door she needs to have in mind the ease or difficulty of acquiring a UPS disguise, and the value to a criminal of what he would get from securing entry, and have some idea of the general incidence of criminality. (Bacharach and Gambetta 2001: 161)

These measures are of course not necessarily easy to obtain, and imperfect knowledge often explains the occurrence of successful deception.

What thus counts as a signal as well as what makes it more or less costly for different types of signalers is particular to the context in which it is used, and this is one of the reasons why the theory, while powerful in abstract terms, is often hard to apply without a fine-grained appreciation of the domain in which signals are emitted and received. To identify German spies British interrogators asked the results of famous cricket matches, the knowledge of which was costless to acquire for a genuine British man by simply living in the country, but hard to know and costly to memorize for even a well-trained German spy.

8.2.7 When signaling fails

Even without deception and even if the right cost conditions are present, there are still ways in which signaling can fail. In general, this is because signaling and asking for certain signals may reveal more than we would wish and thus prove confusing or backfire. I shall mention only two members of a larger family. To reflect on them is important, for it reveals some of the theory's limitations.

8.2.8 Social norms and conventions

Suppose you want to signal that you have a lot of energy. Does running around in circles persuade people that you do? It might, but it also persuades them that you are a very odd person. Or suppose that you want to persuade people that you have lots of financial resources: would, literally, burning a stack of banknotes persuade them? Once again it might, but it would also persuade them that you may also be mad. Or suppose further that you want to persuade a potential partner that you love them. Would giving them the equivalent in cash of the cost of an engagement ring serve the purpose? Which signals really work is not uniquely determined by the cost conditions posited by the theory. Signals that succeed need to be emitted within the constraints of what is acceptable, traditional, and considered normal. Nouveaux riches, for instance, by displaying their wealth prove they have it, but by doing so crassly they also display their uncouthness.

8.2.9 Receiver's demands

In the standard model the signaler chooses to send a signal somehow knowing the receiver has an interest in the information it conveys. But in a richer model the receiver herself when in doubt initiates the communicative exchange by probing and asking for further signals. But asking is not necessarily a neutral move. Consider the folk story of the woman who asks her fiancé to prove his love for her by bringing her his mother's heart. If he does, he signals that he is indeed prepared to do anything for her, but at the same time he gives a sign that he is a cruel man who stops at nothing to get what he wants. At the same time, by asking for such a love test she too gives a sign that she is a monster. Asking for a signal may lead us to say something about ourselves in the process which turns the signaler away. The signal can only work as intended if neither cares about the other's cruelty. To avoid side effects some shared understanding of what is permissible or acceptable to ask in the circumstance must exist. Often, when fishing for signals one pretends not to be, precisely to avoid offending the other.

8.3 Applications

Without aiming to be exhaustive, I will now briefly describe the various fields of applications in descending order of the theories' popularity.

The longest strides, bridging theory and empirical research with equal vigor, have been made in studies of animal behavior and morphology, concerning mating,

feeding, cooperation, and intra- and interspecific conflict. Here I have no space to dwell on this, while encouraging the reader to look up the excellent overview by Searcy and Nowicki (2005).

Signaling theory is now a fundamental part of microeconomics[8] and game theory. Various theoretical refinements have been published over the last twenty years (e.g. Cho and Kreps 1987; Kreps and Sobel 1994). Although to a lesser extent than in animal studies, several empirical tests have also been carried out: on Spence's early model of education as a signal of productivity (see Logren, Persson, and Weibull 2002 for a review and Kübler, Müller, and Normann 2005 for an experimental test), and on a variety of practices, such as product guarantees, financial markets, advertising,[9] charity donations (Glazer and Konrad 1996), and scientific publications funded by private firms (Belenzon and Patacconi 2008).

Recent theoretical developments in economics include[10] a model by Feltovich, Harbaugh, and To (2002), backed by some experimental evidence, which tries to explain a puzzle that is at once of interest to sociologists and may seem a challenge to signaling theory, so-called 'countersignaling':

> Why do the nouveaux riches flaunt their wealth, while the old rich scorn such gauche displays? Why do mediocre students often outperform talented students? Why are moderate quality goods advertised heavily, while high-quality goods rely on their reputation? Why do minor officials prove their status through petty displays of authority, while the truly powerful prove their strength by avoiding such displays?[11]

With respect to the standard model, two realistic modifications are introduced: first, in addition to low- and high-quality signalers there are medium-quality ones; next, receivers have additional incomplete information about the signaler's quality that is independent of the signal. For instance, the signaler's wealth is

> inferred not just from conspicuous consumption, but also from information about occupation and family background. This extra information is likely to be only partially informative, meaning that types of medium quality may still feel compelled to signal so as to separate themselves from low types. But even noisy information will often be sufficient to adequately separate high types from low types, leaving high types more concerned with separating themselves from medium types. Since medium types are signaling to differentiate themselves from low types, high types may choose to not signal, or countersignal, to differentiate themselves from medium types. While it might seem that the sender is just saving signaling costs by not signaling, we show that countersignaling can be interpreted as a signal of confidence that the extra information about the sender is favorable.[12]

In political science, apart from Jervis's early study (1970) written before signaling theory was first formalized, there have been both theoretical developments (e.g. Austen Smith and Banks 1998) and many applications.[13] The topics tackled using the theory range widely, and cannot possibly all be mentioned here. They include

ways of credibly signaling foreign-policy interests (Fearon 1997); trying to explain why sometimes mass political action affects politicians' choices (Lohmann 1993); how different political arrangements can favor more discriminating signals of high-quality politicians (e.g. Besley 2005; Humphreys and Weinstein 2008); under what conditions bargaining mediators are credible (Kydd 2003); whether the size of terrorist attacks can be a signal of terrorist organizations' resources (Overgaard 1994); and whether the theory can shed light on ethnic mimicry (Habyarimana et al. 2007).

Anthropologists too have started to use signaling theory, especially to make sense of seemingly 'wasteful' or 'inefficient' practices in premodern cultures, such as redistributive feasts, big-yam displays, and hunting difficult prey. Closer to the interests of sociologists, they have also used the theory to investigate the cooperative effects of differentially costly rituals and requirements in religious groups (see e.g. Sosis and Bressler 2003). Making sense of established human practices, an understandable ambition, is arduous, for they are often sustained by a whole gamut of other mechanisms. Still, in this literature there is now much attention paid to measuring the payoffs of both signalers and receivers, and attempts at testing the theory against other hypotheses. A good overview of studies in anthropology is in Bliege Bird and Smith (2005).

Finally, in sociology, as I said, the use of signaling theory has been scant and not of uniformly high quality. On the theory front, Breen and Cooke (2005; cf. Breen's chapter in this volume), in an article on the marriage market, develop a model of simultaneous signaling by three types of men and three types of women, before women decide whether to marry. Raub (2004) has a model of 'hostage posting' as a signal of commitment aimed at inducing trust. The late Michael Bacharach (who was an economist) and I developed the analytical links between trust decisions and signaling theory, and extended the signaling model to cover identity signaling (Bacharach and Gambetta 2001) and identity mimicry (Gambetta 2005).

On the empirical front, Hamill (2001) tried to explain, in terms of status signaling within their group, the peculiar attraction that a group of deviant youth in West Belfast displayed for the punishment beatings they received from the IRA. Kaminski (2004), in an ethnographic study of Polish prisons, employs signaling theory to make sense of prisoners' self-harm and initiation rituals. Podolny (2005) makes some use of signaling theory, with reference to status interpreted as a signal of a firm's quality (see also the chapter by Podolny and Lynn in this volume). Hamill and I (2005) try to test some signaling-theory predictions on the signals taxi drivers rely on when deciding whether to pick up hailers or callers in Belfast and New York, which are both dangerous cities for drivers in which paramilitary and ordinary criminals pose as bona fide passengers. On trust and signaling there is also a recent experimental paper by Diekmann and Przepiorka (2007). Finally, I use the theory to explain criminals' strategies to identify bona fide criminals; the patterns of prison fights; and the use of self-harm among criminals (Gambetta 2009: ch. 5).

8.4 Origins

Although as a formally defined and well-developed tool signaling theory is relatively young, in a loose sense the core idea has ancestors, the most notable of which is Thorstein Veblen's *The Theory of the Leisure Class* (see Fig. 8.3), published in 1899.[14] The idea came as a solution to a grand empirical puzzle: How can we explain the widespread occurrence of wasteful activities, self-harming practices, gifts, and nonanonymous altruism? How can we make sense of practices that do not produce material goods for the direct enhancement of well-being, but seem to waste useful goods, or waste the time that could be helpfully spent in productive activities? Why, was Veblen's question, should our fundamental instinct for workmanship, frugality, efficiency, and the pursuit of self-interest be violated on such a large scale by seemingly irrational and costly practices? (The question of waste is also at the centre of the theory in biology: 'The Handicap Principle is a very simple idea: waste can make sense, because by wasting one proves conclusively that one has enough assets to waste and more. The investment—the waste itself—is just what makes the advertisement reliable,' Zahavi and Zahavi 1998: 229.)

The analytical reach of the modern version of the theory has become broader, in that it includes, as we have seen, costs that are not produced only by wasteful use of resources, and its foundations have become deeper and more precise, capable of producing testable propositions. However, it is worth giving a brief account of the ancestors, not only because it is interesting as intellectual history, but because it reveals some of the pitfalls of using the theory which even some contemporaries still incur.

8.4.1 Veblen and his progeny

Veblen was at pains to make clear that he was using the term 'waste' as a 'technical term,' and 'for want of a better term': 'It is here called "waste" because this expenditure does not serve human life or human well-being on the whole, not because it is waste or misdirection of effort or expenditure as viewed from the standpoint of the individual consumer who chooses it' ([1899] 1994: 60).

His answer to the question of what waste is for relies on three assumptions: (i) we want more prestige rather than less; (ii) other people want to know how much of it we deserve; and (iii) prestige is, in contemporary societies, also a function of how much wealth we have. But since wealth is not an easily observable property of individuals and families, people display their real wealth by wasting time, goods, and money in a visible way, conspicuously that is:

> The growth of conspicuous leisure and consumption, it appears that the utility of both alike for the purposes of reputability lies in the element of waste that is common to

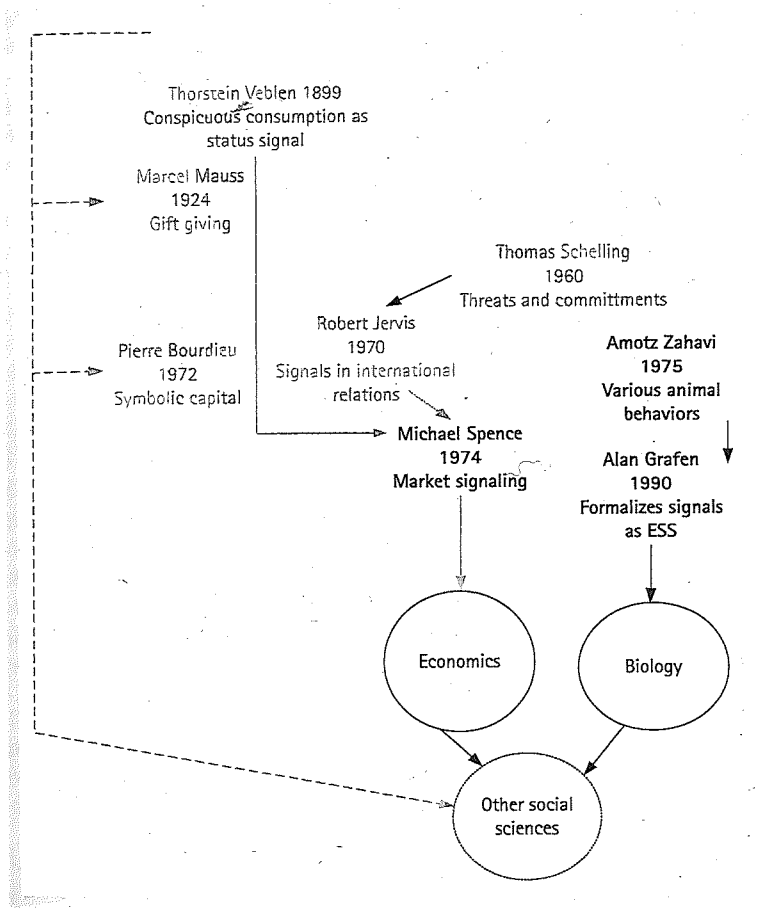

Fig. 8.3 Genealogy of signaling theory (ancestors in gray)

> both. In the one case it is a waste of time and effort, in the other it is a waste of goods. Both are *methods of demonstrating the possession of wealth.*
>
> ([1899] 1994: 53, emphasis added)

Veblen tried to fit into his model a monumental array of human practices, from dress to artistic efforts, from religious objects to leisurely pursuits, from rituals to hedonistic activities. Nearly anything that does not serve basic economic purposes becomes a conspicuous way of informing others of one's wealth. While his book is

still a brilliant read and an inspiration for new work (e.g. Bagwell and Bernheim 1996; Becker, Murphy, and Glaeser 2000), he has little to say on what his model cannot explain.[15]

Whether by influence or coincidence, Veblen's idea has progeny, which include Marcel Mauss and Pierre Bourdieu, and whose latest descendant is Eric Posner. In the *Essais sur le don* Marcel Mauss never cites Veblen, but his core observations bear a striking similarity to Veblen's theory:

> In some potlatch systems [in northwest America] one must spend everything one possesses and keep nothing. The highest prestige will be gained by the richest man who is also capable of squandering his wealth recklessly... Everything is conceived as if it was a 'war of wealth'... Sometimes it is not a question of giving and of receiving in return, but of destroying in order not even to appear desirous of receiving something back. Whole cases of candlefish or whale oil, houses, and blankets by the thousands are burnt; the most valuable coppers are broken and thrown into the sea in order to crush, 'flatten' a rival. In this way, not only does one improve one's position on the social ladder, but also that of one's family. (Mauss [1924] 1954: 35)[16]

Bourdieu never mentions Veblen or Mauss, but in *Esquisse d'une theorie de la practique* ([1972] 1977), as well as later in *La Distinction*, he moves onto the same turf. Just like Veblen, and declaring himself to be reacting to reductive 'economism,' he wants to 'extend economic calculation to *all* the goods, material and symbolic, without distinction' (p. 178). 'Symbolic capital, in the form of the prestige and renown attached to a family and a name is readily convertible back into economic capital' (p. 179). The acquisition of prestige, he writes in reference to his fieldwork in Algeria, implies

> substantial material and symbolic investments, in the form of political aid against attack, theft, offence, and insult, or economic aid, which can be very costly, especially in times of scarcity. As well as material wealth, *time* must be invested... *giving or squandering time* being one of the most precious of gifts... [S]ymbolic capital can only be accumulated at the expense of the accumulation of economic capital.
>
> (Bourdieu [1972] 1977: 180)

Grand theory relying on the core idea has reemerged recently with Eric Posner's *Law and Social Norms* (2000). Posner attempts something very ambitious; namely, to explain social norms as signaling equilibria. He believes that the key property humans want to signal is their ability to postpone gratification. In the language of economics, this is referred to as the 'discount rate'; that is, how much a good is valuable tomorrow relative to its value today. Those who signal a low discount rate are perceived as better partners in repeated cooperative endeavours, for they feel relatively less tempted to cheat. Posner interprets an extravaganza of practices as signals of a low discount rate: manners, fashion, gift-giving, conspicuous consumption, marriage and family, obedience to law, shaming of criminals, deferred sex, unprotected sex, voting, patriotic displays, self-censorship, race

discrimination, nationalism. He never cites Bourdieu, and mentions Mauss only once, in passing (p. 50), and Veblen in a footnote. However, as McAdams points out in an extensive critical essay, just like Veblen, 'at no point does [Posner] identify a particular norm or a type of norm behavior that his theories cannot explain' (2001: 654).

The manner in which the core idea of signaling is used by these authors shares a problem: they interpret anything that looks like waste of resources, anything that seems to contradict standard economic efficiency and has costs for the self-interested agent, as being motivated by the intent to signal either wealth or low discount rates. This tendency to explain too much before careful testing—aptly captured by the phrase 'when you have a hammer everything starts looking like a nail'—has a paradoxical consequence: it makes economic theory always true. This criticism does not apply to Veblen only but, as McAdams points out, to Posner too: 'If people refrain from costly behavior X, we say that they are maximising their utility by avoiding costs. But if people routinely engage in costly behavior X, we can now say that they are signaling their low discount rate and producing a net gain' (2001: 641). An analogous criticism has been voiced of Zahavi's rendition of the theory in biology: if a trait works cheaply and equally in everyone we can call it a standard Darwinian adaptation that maximizes fitness; if it is expensive and highlights individual differences we can call it a Zahavian adaptation that maximizes fitness via signaling. Together they 'explain everything and thus nothing' (Miller 1998: 345).

In theory there is nothing intrinsically wrong in the idea that waste can be a signal of wealth. Michael Spence deals with this in a chapter of his book: 'The status signalling model works, that is to say, signalling takes place, because the real opportunity costs of signalling via conspicuous consumption are negatively correlated with income, which is what consumption is supposed to signal' (1974: 68).

Generally, it is potentially fruitful to entertain the notion of symbolic capital as Bourdieu does; but one cannot proceed by fiat, for one needs criteria to separate what is investment in symbolic capital from costly human activities that have a different origin (cf. Cronk 2005: 609). We need criteria by which to establish how a practice can escape from this trap whereby everything that is real is, in one way or another, rational. What one can extend is not a conclusion, but a method that analyzes human action including waste with a presumption of rationality.

The solution to what Geoffrey Miller called 'the Panglossian temptation' is careful empirical design, which derives from the theory testable predictions liable to be disproved without assuming that all costly activities are *ipso facto* signals. We must obtain measures of the key variables—in particular the costs of the signal, and the benefits that accrue to both the signalers from the signal and the receivers in responding to a given signal. We could thus rule out, as McAdams writes, 'signalling claims when the measured returns were less than the measured costs' (2001: 641).

To test empirically whether people really spend their money on lavish objects to signal their wealth, whether they succeed and how, and who their audience really is and what it gains from responding in one way or another to the signal, for a start we could search for variations in Veblen's assumptions. For instance, we could test whether for domains in which wealth is not valued very much, and in which other properties confer prestige, as for example among academics or priests, we observe lower proportions of conspicuous expenditure. We could further measure whether domains in which wealth is more or less easily observable yield corresponding differences in conspicuous consumption. For instance, in Norway people can now see each other's income-tax returns online, while, say, in Southern Italy many are too engaged in the black economy (or, worse, in the 'black and blue' economy) for anyone to be able to easily know their real wealth, and people should thus spend proportionally a lot more than in Norway displaying it. Research along these lines is only just beginning (e.g. Kerwin, Hurst, and Roussanov 2007), and the extensive range of options that comes to mind with regard to wealth displays epitomizes the potential that this genuinely cross-disciplinary theory has in many other domains of human behavior.

Notes

1. An ESS is a strategy which, if adopted by a population of players, cannot be invaded by any alternative strategy. It is a Nash equilibrium which is 'evolutionarily' stable, meaning that once it is fixed in a population, natural selection alone is sufficient to prevent alternative (mutant) strategies from successfully invading.
2. This hypothesis is controversial; see, for instance, Buller (2005), who challenges the evidence for the cross-cultural applicability of the waist–hip-ratio preference.
3. Interview with Lt. Col. Tim Spicer, *Cambridge Review of International Affairs*, 13 (1) (1999), 165–71.
4. I am not the first to use this as an example of an extreme signal: see Austen Smith and Banks (2000).
5. 'One strong signal of commitment is a man's persistence in courtship. It can take the form of spending a lot of time with the woman [etc.] ... These tactics are extremely effective in courting women as permanent mates, with average effectiveness ratings of 5.48 on a 7-point scale, but only a moderately effective 4.54 at courting casual sex partners. Furthermore, persistence in courtship proves to be more effective for a man than for a woman because it signals that he is interested in more than casual sex' (Buss 2003: 102).
6. The translated text of *Fasti* can be found online at A. S. Kline's free archive of poetry website, at <http://www.tonykline.co.uk/PITBR/Latin/OvidFastiBkSix.htm#BkVIJune9>, accessed 2008.
7. A related category of signals draws its cost from defying not the risk of the receiver's attack, but the risk posed by impersonal forces. Some games of daring come into this category: teenagers in England for instance have been known to put their heads on the railway tracks and compete on who moves away last as the train approaches. Here they are all signalers and receivers of each other's actions. Modern technology allows youth to play this game of daring solo and film themselves, so as to build a lasting record-signal that they can display to a wider audience rather than only to witnesses. One such case ended up, briefly, on YouTube and was reported in the British press in January 2008; <http//:www.dailymail.co.uk/pages/live/articles/news.html?in_article_id=507592&in_page_id=1770>, accessed 2008.
8. An entertaining illustration from an economist's point of view can be found in Frank (1988: chs. 5–6).
9. See Bolton and Dewatripont (2005: 125–7) for an overview of these applications.
10. For another development concerning a phenomenon of interest to sociologists see Austen Smith and Fryer (2005) in which 'The key idea is that individuals face a two audience signaling quandary: behaviors that promote labor market success are behaviors that induce peer rejection.'
11. This quotation and the following one are taken from the authors' extended abstract available at <http://www.zhongwen.com/cs/index.html>, accessed 2008.
12. See also Daley and Green (2007), who generalize a signaling model in which the receiver is also partially informed of the qualities of the signaler by means other than the latter's costly signal, and has some idea of the signaler's 'grades'; that is, any imperfect public message about the signaler's type.
13. A sign of the theory's success is that a useful early book on asymmetric information models in political science by Banks (1991) is, at the time of writing this chapter, cited by nearly a hundred articles, according to Google Scholar.
14. Two ancestors of signaling theory are Thomas Schelling's *The Strategy of Conflict* (1960), in particular in his discussion of threats and credible commitments, and Robert Jervis's *The Logic of Images in International Relations* (1970). By contrast, Ervin Goffman—of whom most sociologists knee-jerkily think whenever the topic of signaling comes up—while of course dealing with communication, often brilliantly so, does not refer to the signaling principle, although he comes close. He circles around the problem of what he calls 'credibility' of statements (1969: 103 ff.), but does not offer the signaling principle as a solution. Following Schelling, he refers to 'commitment' (pp. 112–13), which is a different solution to the problem of credibility: rather than signaling that he has the property to do X, in commitment an agent finds a way to *bind himself* to do X at a later time. The closest Goffman got to signaling theory was to say the following (also quoted by Jervis):

> knowing that the individual is likely to present himself in a light that is favourable to him, the others may divide what they witness into two parts; a part that is relatively easy for the individual to manipulate at will, being chiefly his verbal assertions, and a part in regard to which he seems to have little concern or control, being chiefly derived from the expressions he gives off. The others may then use what are considered to be the ungovernable aspects of his expressive behaviour as a check upon the validity of what is conveyed by the governable aspects. ([1959] 1990: 18)

Goffman picks up one strand of ST by pointing out that people will not attend to easy-to-fake signals, such as words, but then points to their attempts to catch whether the others are truthful by observing giveaways, which would include emotional expressions (1969: 127).

15. For some penetrating criticisms of Veblen's theory see Veyne (1976: 94 ff.) and Elster (1983: 66 ff.).
16. Although I use the translation in the 1954 English edition as a basis, I have modified this quotation to make it more faithful to the original and correct some misinterpretations. For a modern study of gifts as economic signals see Camerer (1988).

References

Austen Smith, D., and Banks, J. F. (2000), 'Cheap Talk and Burnt Money', *Journal of Economic Theory*, 91: 1–16.

—— and Fryer, R. (2005), 'An Economic Analysis of "Acting White"', *Quarterly Journal of Economics*, 120 (2) 551–83.

Bacharach, M. O. L., and Gambetta, D. (2001), 'Trust in Signs', in Karen Cook (ed.), *Trust in Society* (New York: Russell Sage).

Bagwell, L. S., and Bernheim, B. D. (1996), 'Veblen Effects in a Theory of Conspicuous Consumption', *American Economic Review*, 86 (3): 349–73.

Banks, J. S. (1991), *Signaling Games in Political Science* (New York: Harwood).

Becker, G. S., Murphy, K. M., and Glaeser, E. (2000), 'Social Markets and the Escalation of Quality: The World of Veblen Revisited', in G. S. Becker and K. M. Murphy (eds.), *Social Economics: Market Behavior in a Social Environment* (Cambridge, Mass.: Harvard University Press), 84–104.

Belenzon, S., and Patacconi, A. (2008), 'Open Science as a Signaling Device: Evidence from Firm Publications', unpublished paper, March.

Besley, T. (2005), 'Political Selection', *Journal of Economic Perspectives*, 19: 3–60.

Bliege Bird, R., and Smith, E. A. (2005), 'Signaling Theory, Strategic Interaction, and Symbolic Capital', *Current Anthropology*, 46 (2): 221–48.

Bolton, P., and Dewatripont, M. (2005), *Contract Theory* (Cambridge, Mass.: MIT Press).

Bourdieu P. [1972] (1977), *Outline of a Theory of Practice* (Cambridge: Cambridge University Press).

Breen, R., and Cooke, L. P. (2005), 'The Persistence of the Gendered Division of Domestic Labour', *European Sociological Review*, 21 (1): 43–57.

Buller, D. (2005), *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature* (Cambridge, Mass.: MIT Press).

Buss, D. M. (2003), *The Evolution of Desire: Strategies of Human Mating* (New York: Basic).

Camerer, C. (1988), 'Gifts as Economic Signals and Social Symbols', *American Journal of Sociology Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure*, 94 (supplement): S180–S214.

Cho, I. K., and Kreps, D. (1987), 'Signaling Games and Stable Equilibria', *Quarterly Journal of Economics*, 102: 179–221.

Cronk, L. (2005), 'The Application of Signalling Theory to Human Phenomena: Some Thoughts and Clarifications', *Social Science Information*, 44: 603–20.

Daley, B., and Green, B. (2007), 'Market Signaling with Grades', unpublished paper, Graduate School of Business, Stanford University, December.

Diekmann, A., and Przepiorka, W. (2007), 'Signaling Trustworthiness: Evidence from Lab Experiments', unpublished sociology paper, ETH Zurich, November.

Elster, J. (1983), *Sour Grapes* (Cambridge: Cambridge University Press).

Fearon, J. D. (1997), 'Signaling Foreign Policy Interests', *Journal of Conflict Resolution*, 41 (1): 68–90.

Feltovich, N., Harbaugh, R., and To, T. (2002), 'Too Cool for School? Signalling and Countersignalling', *RAND Journal of Economics*, 33: 630–49.

Frank, R. H. (1988), *Passions Within Reason* (New York: Norton).

Gambetta, D. (2005), 'Deceptive Mimicry in Humans', in S. Hurley and N. Chater (eds.), *Perspectives on Imitation: From Neuroscience to Social Science* (Cambridge, Mass.: MIT Press), ii. 221–41.

—— (2009), *Codes of the Underworld: How Criminals Communicate* (Princeton, N.J.: Princeton University Press).

—— and Hamill, H. (2005), *Streetwise: How Taxi Drivers Establish Customers' Trustworthiness* (New York: Russell Sage).

Glazer, A., and Konrad, K. (1996), 'A Signaling Explanation for Private Charity', *American Economic Review*, 86 (4): 1019–28.

Goffman, E. [1959] (1990), *The Presentation of Self in Everyday Life* (London: Penguin).

—— (1969), *Strategic Interaction* (Philadelphia, Pa.: University of Pennsylvania Press).

Grafen, A. (1990), 'Biological Signals as Handicaps', *Journal of Theoretical Biology*, 144: 517–46.

Habyarimana, J., et al. (2007), 'Placing and Passing: Evidence from Uganda on Ethnic Identification and Ethnic Deception', paper presented at the annual meeting of the American Political Science Association, Chicago, August.

Hamill, H. (2001), 'Hoods and Provos: Crime and Punishment in West Belfast', D.Phil. thesis (Oxford University).

Humphreys, M., and Weinstein, J. (2008), 'Policing Politicians: Citizen Empowerment and Political Accountability in Uganda', unpublished manuscript, February.

Jervis, R. (1970), *The Logic of Images in International Relations* (New York: Columbia University Press).

Johnstone, R. A. (1997), 'The Evolution of Animal Signals', in J. R. Krebs and N. B. Davies (eds.), *Behavioural Ecology* (Oxford: Blackwell), 155–78.

Kaminski, M. (2004), *Games Prisoners Play* (Princeton, N.J.: Princeton University Press).

Kerwin, K. C., Hurst, E., and Roussanov, N. (2007), 'Conspicuous Consumption and Race', NBER working paper no. 13392, September.

Krebs, J. R., and Davies, N. B. (1998), *An Introduction to Behavioural Ecology* (Oxford: Blackwell).

Kreps, D. M. (1990), *A Course in Microeconomic Theory* (New York: Harvester Wheatsheaf).

—— and Sobel, J. (1994), 'Signalling', in R. J. Aumann and S. Hart (eds.), *Handbook of Game Theory with Economic Applications* (Amsterdam: North-Holland), ii. 849–67.

Kübler, D., Müller, W., and Normann, H. T. (2005), 'Job Market Signaling and Screening in Laboratory Experiments', *Games and Economic Behavior*, 64 (2008), 219–36.

KYDD, A. (2003), 'Which Side Are You On? Bias, Credibility and Mediation', *American Journal of Political Science*, 47 (4): 597–611.
LIVY (1912), *History of Rome*, trans. W. L. Roberts (New York: Dutton); at <http://www.perseus.tufts.edu/cgi-bin/ptext?lookup=Liv.++2.+12>, accessed 2008.
LOGREN, K. G., PERSSON, T., and WEIBULL, J. W. (2002), 'Markets with Asymmetric Information: The Contributions of George Akerlof, Michael Spence and Joseph Stiglitz', *Scandinavian Journal of Economics*, 104 (2): 195–211.
LOHMANN, S. (1993), 'A Signaling Model of Informative and Manipulative Political Action', *American Political Science Review*, 87 (2): 319–33.
MCADAMS, R. H. (2001), 'Signaling Discount Rates: Law, Norms, and Economic Methodology', *Yale Law Journal*, 110: 625–89.
MAUSS, M. [1924] (1954), *The Gift: Forms and Functions of Exchange in Archaic Societies* (London: Cohen and West).
MILLER, G. F. (1998), 'Review of "The Handicap Principle" by Amotz and Avishag Zahavi', *Evolution and Human Behavior*, 19 (5): 343–47.
OVERGAARD, P. B. (1994), 'The Scale of Terrorist Attacks as a Signal of Resources', *Journal of Conflict Resolution*, 38 (3): 452–78.
PODOLNY, J. M. (2005), *Status Signals* (Princeton, N.J.: Princeton University Press).
POSNER, E. (2000), *Law and Social Norms* (Cambridge, Mass.: Harvard University Press).
RAUB, W. (2004), 'Hostage Posting as a Mechanism of Trust: Binding, Compensation, and Signaling', *Rationality and Society*, 16 (3): 319–65.
SCHELLING, T. (1960), *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press).
SEARCY A. W., and NOWICKI, S. (2005), *The Evolution of Animal Communication. Reliability and Deception in Signalling Systems* (Princeton, N.J.: Princeton University Press).
SINGH, D., and YOUNG, R. K. (2001), 'Body Weight, Waist-to-Hip Ratio, Breasts, and Hips: Role in Judgments of Female Attractiveness and Desirability for Relationships', *Ethology and Sociobiology*, 16: 483–507.
SOSIS, R., and BRESSLER, E. R. (2003), 'Cooperation and Commune Longevity: A Test of the Costly Signaling Theory of Religion', *Cross-cultural Research*, 37 (2): 211–39.
SPENCE, M. (1973), 'Job Market Signaling', *Quarterly Journal of Economics*, 87: 355–74.
—— (1974), *Market Signaling* (Cambridge, Mass.: Harvard University Press).
VEBLEN, T. [1899] (1994), *The Theory of the Leisure Class* (New York: Dover).
VEYNE, P. (1976), *Le pain et le cirque* (Paris: Seuil).
ZAHAVI, A. (1975), 'Mate Selection: A Selection for a Handicap', *Journal of Theoretical Biology*, 53: 205–14.
—— and ZAHAVI, A. (1998), *The Handicap Principle* (Oxford: Oxford University Press).

CHAPTER 9

NORMS

JON ELSTER

INTRODUCTION

THERE is no general agreement among scholars about how to define social norms. I shall stipulate a definition, which will be justified, I hope, by its rough correspondence with ordinary as well as scholarly usage and by its usefulness in generating questions and suggesting answers.

Consider two statements:

(1) Always wear black clothes in strong sunshine.
(2) Always wear black clothes at a funeral.

The first injunction is a matter of instrumental rationality, since the air between the body and the clothes circulates more rapidly when the garments are black. The second expresses a social norm, which has no obvious instrumental significance. The existence and importance of social norms cannot be doubted. The proximate causes involved in their operation, which will be the focus of this chapter, are reasonably well understood. Yet their ultimate origin and function (if any) remain controversial.

There is a pervasive tendency in much of the literature to view social norms as socially useful. Norms are supposed to be society's way of coping with market failures (Arrow 1971), a mechanism for internalizing externalities (Coleman 1990), or a welfare-maximizing device (Ellickson 1991). There are no doubt cases in which norms have these effects, and probably some cases in which they owe their existence to these effects. Yet against this Panglossian view I believe that many norms are sources of pointless suffering. When a small girl comes home crying because her