

Cooperation and trust in conversational exchanges

Paul FAULKNER

BIBLID [0495-4548 (2008) 23: 61; pp. 23-34]

ABSTRACT: A conversation is more than a series of disconnected remarks because it is conducted against a background presumption of cooperation. But what makes it reasonable to presume that one is engaged in a conversation? What makes it reasonable to presume cooperation? This paper considers Grice's two ways of answering this question and argues for the one he discarded. It does so by means of considering a certain problem and analysis of trust.

Keywords: trust, cooperation, Grice, knowledge, reasons, belief.

1

Cooperation is a pervasive feature of everyday life. Take the quotidian example of our arranging to meet at a certain time and place: this requires us both to show up. And our cooperating in this can depend on communicating further agreements and information; our meeting, for instance, could hinge on my communicating train delays. Only I know of the train delay, and in general, given that we can only be in one place at one time, we are often in such a situation: we often possess information that would be good for another to share.¹ This sharing of information can be necessary for cooperative endeavour and is itself an instance of cooperation. Moreover, that interlocutors will share information and cooperate in conversation is something that is mutually expected: an essential feature of *conversations*, Grice observed, is that they are not a succession of disconnected remarks but are a cooperative effort. This observation Grice formulated as the supposition that conversations are conducted under a presumption of cooperation. Each interlocutor expects (*ceteris paribus*) the other to follow the *Cooperative Principle*: “make your conversational contribution such as is required... by the accepted purpose or direction of the talk exchange” (Grice 1967, p. 26). This presumption Grice then showed to be necessary for uncovering what information was communicated by *conversational implicature*. To give one of his examples: the recipient of the testimonial “Dear Sir, Mr. X’s command of English is excellent, and his attendance at tutorials has been regular. Yours, etc.” can work out that the letter writer intends to communicate Mr. X’s unsuitability for the advertised philosophy post, but only by presuming the Cooperative Principle is being followed (Grice 1967, 33).

Conversations are conducted against a background presumption of cooperation, and this presumption is necessary for understanding what information a speaker communicates. This remains true even when the conversation is between interlocutors who are knowingly uncooperative. For instance, Grice’s example of irony is A stating “You’re a fine friend” on learning that his close friend X has divulged his secret to a business rival.² In this example X’s actions demonstrate that there can be no presumption that relations

¹ We often have what Williams calls “purely positional advantage”, see Williams 2002, p. 42.

² Grice 1967 (p. 34) and changed ever so slightly.

are cooperative, but there still must be the presumption that conversation is so because without it X could not understand that A implicates the opposite of what he says. Nevertheless, it is possible for relations to be such that there could be no presumption that interlocutors are following the Cooperative Principle, but in this case interlocutors would not be engaged in a ‘conversation’ in Grice’s sense: the exchange would merely “consist of a succession of disconnected remarks”, and there would be an accompanying uncertainty as to what information these remarks conveyed (Grice 1967, p. 26).³ So not all talking is a conversation, not all exchanges are a cooperative effort, but this then raises the question: what makes it rational to presume that a talk exchange is a conversation? Grice states, “I would like to be able to think of the standard type of conversational practice not merely as something that all or most do *in fact* follow but as something that it is *reasonable* for us to follow, that we *should not* abandon” (Grice 1967, p. 29). This paper will consider Grice’s two ways of substantiating this claim, and argue for the one he discarded. The aim in doing so is to offer a general answer as to why it can be rational to presume cooperation.

2

What makes it rational for interlocutors to presume their talk is guided by the Cooperative Principle? In short, what makes it rational to presume conversational cooperation? This question is difficult because it raises a game-theoretic problem of trust. As Grice observed central conversational goals include “giving and receiving information, influencing and being influenced by others” (Grice 1967, p. 30). But then a speaker’s purpose in conversation can be to influence an audience and this can be at odds with the audience’s goal of receiving information. For example, when a nervous buyer asks a salesman about a used car’s history his purpose in doing so is to acquire information that will help him decide whether to purchase the car or not. When the salesman replies that the car has had one previous old lady owner his purpose in doing so is to influence the buyer’s behaviour and ensure a sale. The contributions of each are not disconnected but whether there is cooperation can be a hard question. Does the salesman really mean what he says? Is he making a joke in offering such a stereotypical response? The buyer wants to buy a car certainly, but he doesn’t want to buy a wreck. However, if the buyer thinks that the salesman would prefer to sell a wreck if he can, despite preferring any sale to none, then the question of whether to accept the salesman’s assurance—and so to continue in the accepted direction of the talk exchanges—is parallel to the hard question of whether to trust and cooperate that is faced in the Prisoner’s Dilemma.⁴

Of course not all talk exchanges take place in contexts where it is difficult to judge if there is cooperation. However, the point of the used car case is to illustrate how a divergence in communicative purpose can make it hard to explain why following the Cooperative Prin-

³ “[I]mplicatures do not presuppose language as simply a practice involving semantic and syntactic rules, together with the norm that certain kinds of utterances are taken to be true; they look to the use of language under favourable social conditions which enable it to be indeed cooperative. They are *conversational* implicatures, but not everyone who is talking with someone else is engaged, in the required sense, in a conversation” (Williams 2002, p. 100).

⁴ It is parallel in that trust in both games exposes the buyer and prisoner to exploitation and lack of trust produces a sub-optimal outcome. There is difference in that this game is one-way, see Hardin 2006, p. 46ff.

ciple and maxims is reasonable. The problem is that interlocutors' purposes in communication easily divide. They do so because what makes communication advantageous to speakers is that it allows them to influence audiences; through another's acceptance of what one says one can get them to think, feel and act in specific ways. On the other hand, what makes communication advantageous to audiences is that it can be a source of information about the non-immediate world. Arguably, communication has evolved and stabilized only because it can serve *both* these purposes.⁵ Cooperation represents an achievement because whilst it is always in an audience's interest to be informed, honesty needn't always best serve a speaker's interest of being believed. The used-car case thus illustrates a general game-theoretic problem of trust for conversation—or at least a problem for those conversations which involve an audience depending on a speaker for information. This problem could be set up as follows.

First assume that interlocutors contributions to a talk exchange are guided by self-interest. This is just to assume that the contributions of each are to be explained by their beliefs and desires. It is to assume a very broad notion of guidance by self-interest, where this could be made clear by distinguishing *raw* from *all-in* preferences. All-in preferences incorporate a subject's beliefs, desires and values, whilst raw preferences consider only self-interest narrowly conceived (Bacharach and Gambetta 2001). Thus the first assumption is not that interlocutors are selfish but that each is moved solely by their wants or all-in preferences. Second, assume that the audience is ignorant of the speaker's all-in preferences. Since the audience's only reason for believing what is said is the speaker's utterance, this ignorance implies that the audience is not in a position to assess the probability of what is said. In game-theoretic terms, the talk exchange takes place under *uncertainty* rather than *risk* because the audience cannot assess the probability of acquiring a false belief (Luce and Raiffa 1957, p. 13). Given these two assumptions, a talk exchange that purports to be informative will thereby present an audience with an *Acceptance Dilemma*. The audience might be in a position of needing to know something, but acquiring a false belief would leave the audience in a worse position than ignorance; and uncertainty as to the speaker's all-in preferences generates uncertainty as to the truth of what is said. Combining this vulnerability and uncertainty with the recognition that speakers can have motive to misinform makes the only rational course that of rejecting the speaker's offer of information. Thus, where these assumptions hold it is not rational to continue in the direction of the talk exchange and cooperate in conversation.

This argument proposes a certain scepticism concerning testimony. However, the Acceptance Dilemma is only a semi-sceptical conclusion because rejection of its two grounding assumptions is open. These assumptions are: (1) that interlocutors are guided by self-interest; and (2) that the audience is ignorant of the speaker's (all-in) preferences. As I interpret it, Grice's considered account of what makes it rational to follow the Cooperative Principle amounts to a denial of assumption (2).

So I would like to be able to show that observance of the Cooperative Principle and maxims is reasonable (rational) along the following lines: that anyone who cares about the goals that are central to conversation/communication (such as giving and receiving information, influencing and being influenced by others) must be expected to have an interest, given suitable circumstances, in participation in talk exchanges

⁵ This is argued by Sperber 2001.

that will be profitable only on the assumption that they are conducted in general accordance with the Cooperative Principle and the maxims. (Grice 1967, pp. 29-30)

Interlocutors don't always have an interest in being cooperative; interlocutors only have an interest in being cooperative *given suitable circumstances*. From an audience's perspective, these circumstances will be those where some judgement can be reached as to the risk of accepting the speaker's testimony. The weakness of the argument for the Acceptance Dilemma then lies in the fact that, contrary to assumption (2), we can nearly always reach such a judgement. In section three, I will develop the epistemology of this account of cooperation.

The main ambition of this paper, however, is to develop an alternative account of what makes it reasonable to observe the Cooperative Principle and maxims. This alternative develops an idea rejected by Grice.

For a time, I was attracted by the idea that observance of the Cooperative Principle and the maxims, in a talk exchange, could be thought of as a quasi-contractual matter, with parallels outside the realm of discourse. ... talk exchanges seemed to me to exhibit, characteristically, certain features that jointly distinguish cooperative transactions. (Grice 1967, p. 29)

Grice rejected this idea because "there are too many types of exchange, like quarrelling and letter writing, that it fails to fit comfortably" (Grice 1967, p. 29). However, a talk exchange can be an uncooperative quarrel and still be a conversation in Grice's sense that it can be presumed to be guided by the Cooperative Principle. Certainly there are limits to when this presumption can be rationally held, but the idea I want to pursue is that these limits, along with what rationally sustains this presumption, can be delivered by a general account of cooperative activity. This is to pursue Grice's discarded idea, with the basic idea being that it can be rational to presume that talk is guided by the Cooperative Principle because it can be rational to make certain presumptions about the trustworthiness of one's interlocutor. I will argue this in section five after highlighting some problems with Grice's favoured solution in section four.

3

The Cooperative Principle —"Make your conversational contribution such as is required"—subsumes various more specific maxims. These include maxims of Quality: "Do not say what you believe to be false" and "Do not say that for which you lack adequate evidence".⁶ And the maxim of Quantity: "Make your contribution as informative as required" (Grice 1967, p. 28). The satisfaction of these maxims of Quality and Quantity can be read as a definition of *trustworthiness*. Trustworthy testimony or testimony from a trustworthy speaker, and I will use these two interchangeably, should both address an audience's epistemic need for information and provide that information. Trustworthiness is not a factive notion, a speaker can do his best and still be in error, but it secures probable truth (reliability) and relevance (or at least speaker perceived relevance). Consequently, an audience's judgement of trustworthiness implies an assessment of the risk of accepting the speaker's testimony. However, assumption (2) of the Acceptance Dilemma is that audiences are ignorant of speakers' preferences, and so can only accept testimony under uncertainty. So one response to the Acceptance Dilemma —the response I'd like to consider in this section— is simply that this as-

⁶ Strictly these two maxims of Quality fall under the supermaxim "Try to make your contribution one that is true" (Grice 1967, p. 27).

sumption often lacks any plausibility. This, I suggest, is Grice's considered proposal: for the audience who cares about receiving information, it is rational to presume that the Cooperative Principle guides a talk exchange when one can form a judgement of trustworthiness, or equivalently —untrustworthiness, where this amounts to an assessment of risk.

Completing Grice's proposed explanation then requires an account of how we, as audiences, judge trustworthiness. To illustrate this consider again the purchase of a used car. Despite a potential buyer's wanting to purchase a used car, rational self-interest can counsel the would be buyer not to do so. This is its counsel when the salesman is judged untrustworthy; that is, judged not to be trying to say what is true. In terms of this case, this judgement is based on the possibility that the salesman's testimony could be rationally explained as an attempt to further the salesman's best interest of selling a wreck. So judgements of trustworthiness, or equivalently untrustworthiness, are based on judgements of an interlocutor's preferences or motivations for cooperation. Williams identifies four very general motivations people can have to cooperate, or allow themselves to depend in some way on another's actions. People can be motivated to cooperate by fear of sanctions, particular self-interest, a positive evaluation of cooperation and a positive evaluation of friendly relations (Williams 1988, p. 118). The first two of these motivations are egoistic and their combination provides standard resolutions to the Prisoner's Dilemma; it does so because punishment can change the structure of pay-offs that generate the dilemma making it a prisoner's best interest to cooperate. 'Punishment' can come in various guises. Given an indefinite series of exchanges, it could take the form of the other player adopting a tit-for-tat strategy, where this is the policy of cooperating on the first move and then doing whatever the other player did on the previous move.⁷ Given good communication networks, it could take the form of a tarnished reputation. A scientist who fabricates results can lose employment and struggle to regain it, and if word gets around that a used car dealer sells too many 'lemons' he will need to seek other employment, greengrocer perhaps.⁸ And given a centralised institution, sanctions could be officially implemented. In this way, professionals who fail in their professional roles can have their professional status stripped; doctors, for instance, can be struck off the General Medical Council (Blais 1987). By contrast, a positive evaluation of cooperation or friendly relations are non-egoistic motivations. In recognising that people can be motivated in these ways, we recognise that all-in preferences need not be raw.

As audiences we are sensitive to all these motivating reasons for cooperation. This sensitivity then allows us to reach judgements of whether or not a speaker is trustworthy. It can do so in two ways. First, it grounds the general distinctions we draw between testimonies. We distinguish testimony into types based upon its source and content and differentiate between the credibility of testimonial types, judging some to be credible and others not. Speakers' trustworthiness is inferred from these generalisations. For instance, the testimony of

⁷ Axelrod (1984) argues that tit-for-tat is the rationally optimal strategy. Though, the length of the series needs to be indefinite otherwise it is rational, in the terms of the dilemma, to always defect. See Luce and Raiffa 1957, p. 98.

⁸ See Kreps, Milgrom, Roberts and Wilson 1982. And maybe it is true that we are all "sensible to shame, when detected in falsehood". Hume 1777, §88, p. 112. And, in case it is not obvious, a 'lemon' is slang for "something which is bad or undesirable or which fails to meet one's expectations". Possibly from the phrase 'to hand a lemon' meaning "to pass off a sub-standard article as good; to swindle (a person)." Definitions taken from the OED.

doctors on medical matters is often judged to be credible. However, this cannot be because we have made the observations needed to empirically ground this generalisation.⁹ Rather, this judgement seems to be supported by our recognising that there are institutional controls on doctors and recognising that doctors also adopt a duty to care. The generalisation about the credibility of doctors on medical matters can then be rationally explained in terms of our recognising motivations for cooperation.¹⁰ The position is similar in the generalisation about the credibility of used car salesman on their cars offered for sale, as well as the general credibility we can attach to those with whom we have friendly relations. Second, our awareness of particular features of the testimonial context, where this includes our particular observation of a speaker, can change our expectations of trustworthiness. That there is a market for used cars implies that the reassurances of the used car salesmen are occasionally accepted. Again, it is arguably our sensitivity to what motivates cooperation which provides the rational grounds for modifying our expectations. For instance, one might imagine a potential buyer changing his expectation of trustworthiness on perceiving a salesman not to be the type to seek advantage. The buyer's knowledge that people can be motivated by more than raw self-interest can make this shift in expectations rational.¹¹

As audiences we have a battery of generalisations that support an inference to a speaker's trustworthiness. And our judgements of trustworthiness are responsive to our particular dealings with speakers. So assumption (2), which is needed to generate the Acceptance Dilemma, is often false: we are rarely ever in a state of uncertainty and can ordinarily form some assessment of risk. Maybe, (2) is only true for the occasional *one-off* encounter. And when it really is impossible to judge risk, as it might be in a one-off encounter, then what a speaker means by his utterance can become correspondingly opaque.

This, I suggest, develops Grice's considered solution. In the next section I raise some problems for this strategy before exploring the alternative Grice rejected.

4

Suppose an audience judges that a certain speaker is of a trustworthy type because the speaker is sensitive to the opprobrium that attaches to being wrong and so will 'try to make their contribution one that is true'. The correctness of this judgement is consistent with the audience also correctly thinking that the speaker does not extend his good will in any way. It need not matter to the speaker that the audience depends on his utterance, and it need not matter to the audience that the speaker is not moved by his dependence. All that matters to the rationality of the audience's acceptance is that the audience's judgement supports the inference that the speaker's utterance is informative or reliably indicates its truth.

This account of what makes cooperation reasonable satisfactorily explains many of our cooperative engagements. We expect "a great uniformity among the actions of all men"

⁹ The limits of observational data and our consequent inability to ground our generalisations about testimonial credibility in personal observation is often taken to be the central weakness of the reductive theory of testimony. See Coady 1992.

¹⁰ A related suggestion is Lipton's idea that what makes the acceptance of testimony rational is the audience's ability to infer that the truth of a piece of testimony is the best explanation of the speaker's uttering it. See Lipton 1998.

¹¹ I offer an account of contextual reasons for accepting testimony in Faulkner 2002.

(Hume 1777, p. 65). And we condition our behaviour on these expectations. I might know, for instance, that I must be ready on time because you are always punctual; or I might know that it is time to get out of bed by the clink of the milkman's bottles. Our responsiveness to uniformities in human behaviour is simply a special case of our responsiveness to the regularities we find in the world. The arrival of the milkman can function as my morning alarm clock and I expect it in the same way that I expect my car to start and the traffic lights to change after a certain pattern. The expectation here is simply one of inductive inference. Grice is correct to see these expectations as grounds for the rationality of the Cooperative Principle because, in general, it is rational to base cooperation on inductive predictions. Cooperation involves dependence and it is rational to be dependent when one expects things to work out. This conjunction of dependence and expectation then identifies a thin *predictive* sense of trust: we trust someone when we depend on them acting a certain way and predict that they will indeed act in this way. Thus, I trust you to be punctual in the same way that I trust my car to start: in both cases there is dependence and an inductive prediction. Where there is trust, there is cooperation.

Cooperation can be rationally grounded by inductive predictions about another's behaviour: one's having inductive grounds to predict cooperative behaviour is sufficient to make it reasonable to trust and cooperate. However, one might hope for a different answer as to what makes cooperation reasonable. For suppose that trustworthiness is defined, as suggested, in terms of the satisfaction of the maxims of Quality and Quantity. If this is the case, it is wrong to describe as trustworthy the speaker who tries to say what is true solely because he is sensitive to the opprobrium that attaches to getting things wrong. In not being moved by the audience's epistemic dependence, the speaker is not truly moved 'to try and make his contribution one that is true'; rather, the speaker is merely motivated to say what is true *or* what will not be recognised to be in error. The audience's judgement of this speaker's motivations is thus not sufficient for trustworthiness. This is illustrated by Williams's case of a speaker who says "Someone has been opening you mail" when they are the culprit (Williams 2002, 97). This speaker didn't lie, there is no lack of reliability and their utterance is even informative up to a point, but their utterance is also misleading in a way that renders the speaker *un*trustworthy; their utterance does not satisfy the maxim "Make your contribution as informative as is required". Moreover, the reason that this maxim of Quantity is not satisfied is just because the speaker is not motivated by the audience's need for information —the speaker is not motivated by the audience's epistemic dependence.

We trust one another to behave predictably in a sense that can be applied to our trusting that certain regularities in the world continue as they have done before. However, we also trust one another in a stronger or thicker sense; we expect things of one another and one of the things we can expect is that others be moved by our dependence on them. We think our dependence on another's acting in a certain way can impose an obligation on them to act in this way. Only if another is responsive to such obligations are they trustworthy, in the sense that their utterance is guided by the maxims of Quality and Quantity. This sense of trust might be labelled *affective* since it is associated with various reactive attitudes, and defined thus: Subject *A* affectively trusts subject *S* to ϕ if and only if *A* knowingly depends on *S* ϕ -ing and expects *S*'s knowing this to motivate *S* to ϕ (where *A* expects this in the sense that he expects it of *S* and will be susceptible to that reactive attitude characteristic of trust's being let

down were S not to attempt to ϕ for this reason).¹² The question of what makes it rational to presume that one's interlocutor is guided by the Cooperative Principle and associated maxims, I suggest, can then be phrased as the question of what makes it rational to trust in this affective sense.

5

Interlocutors can trust one another in the affective sense just outlined. The possibility of characterising the relationship between interlocutors as trusting in this way then allows for an alternative account of what makes it reasonable for interlocutors to follow the Cooperative Principle and its maxims. What makes it reasonable depends upon whether the interlocutor is, at that point, the speaker (trusted) or the audience (truster). I take the perspective of the audience first.

In the central problematic case I am concerned with the talk exchange is characterised by the audience needing information from the speaker. The speaker has purportedly told the audience what he needs to know, the question from the audience's perspective is then what makes it reasonable to continue this exchange in its mutually accepted direction and accept the speaker's offer of information. Faced with this question, the audience cannot choose to trust in the predictive sense. Predictive trust is not an attitude that can be adopted at will: the predictions we are willing to make are largely, if not entirely, a function of the evidence that we possess. By contrast, an audience can choose to trust in the affective sense and so choose to accept a speaker's offer of information. This need not amount to believing what is said: acceptance is the attitude of 'treating as true' rather than of belief (Cohen 1992). But since the audience's acceptance of the speaker's testimony will nevertheless guide action and further belief, the question of its rationality is still pressing. What would make acceptance reasonable, of course, is the belief that the speaker was cooperating. However, were such belief a pre-condition of affective trust, it would not be possible to choose to trust because it is not possible to choose this belief. What makes affective trust rational is then not the belief that the speaker is trustworthy but the presumption that he is so. This presumption is made in adopting the attitude of affective trust. In trusting a speaker, an audience places an expectation on them: the audience expects the speaker to be moved by recognition of their dependence to try and say what is true —to try and address their need for information. And in holding the speaker to this expectation in the trust-situation where the audience believes that the speaker can recognise his need for information, the audience presumes that the speaker has a reason to be trustworthy and so presumes that the speaker will indeed be trustworthy. Certainly, this implied presumption is defeasible, but when it is not defeated it gives the audience a reason to accept what the speaker says in the same way that the belief that the speaker is trustworthy would do so.

Affective trust thereby renders it reasonable for an audience to presume that the speaker being cooperative in what he says. It does so because it implies the presumption that the speaker is trustworthy. Moreover, affective trust is equally reasonable because it implies this presumption: it is reasonable to trust if one presumes the trusted is trustworthy. Thus affective trust is rationally self-supporting attitude: it implies a presumption that rationalises itself.

¹² I offer these definitions of trust in Faulkner 2007a and Faulkner 2007b.

Since the idea that we can adopt an attitude of affective trust thereby offers an account of what makes cooperation rational it equally offers a basis for reconstructing why it might be rational to be guided by the Cooperative Principle. On this reconstruction, the rationality of following the Cooperative Principle, from the audience's perspective, turns on the possibility of adopting this attitude of trust. This, in turn, rests on the possibility of making a certain presumption about the speaker and the relationship had with them, *viz.* that the relationship is such that the speaker will see the audience's dependence as a reason to be informative.

At this juncture it might seem that the rationality of affective trust comes down to the evidence had for believing that the relationship with the speaker really does support this presumption made in trust; that, for instance, the speaker is a friend or someone who has the audience's interests at heart. On this suggestion, what makes an audience's acceptance reasonable will then be the audience's evidence that this is the case and their assessment of the risk involved in acting on this belief. This suggestion thereby returns to Grice's considered account of what makes cooperation reasonable, outlined in section three. And this shows the suggestion to misconceive the nature of the trust that can bind interlocutors: it misconceives the nature of affective trust. Trust is a way of *creating* as well as sustaining trusting relationships: trusting relationships are created through the demonstration of a willingness, on the part of the truster, to put themselves at risk through dependence on the goodwill of the trusted. It is essential to trust that it can involve willed dependence because it is essential to trust that the nature of trusting relations is something that can be largely mutually determined, unlike the truth of belief. It is true that we do have beliefs about our interlocutors being certain types and that, for instance, we do discriminate in our trust practises between friends and strangers. And it is true that these discriminations can determine the nature of the reasons we have for trust and cooperation. But this is not because these discriminations constitute our reasons. Rather, it is because the truth of these discriminations determines the truth of the presumptions made in trust and so determines whether the reason trust provides is a good one or not.

Consider now how trust provides a speaker with a reason to be trustworthy. In affectively trusting a speaker, an audience believes the speaker can recognise their need for information and expects the speaker to be responsive to this and try to be informative —to try and follow Grice's maxims of Quality and Quantity. In holding the speaker to this expectation, the audience takes an evaluative stance towards the speaker. This evaluative stance is expressed positively in the presumption that the audience makes in trust, namely that the relationship had with the speaker is such that the speaker will see their need for information as a reason to be informative. So if the speaker agrees that their relation is, or should be, structured by this presumption, then the speaker will accordingly take themselves to possess a reason to be informative or trustworthy in the sense defined. In thus fulfilling the presumption made by the audience in trust, the speaker ensures that the audience does not merely have reasons for accepting what is said but is further justified in doing so. This is because the audience's trust, which is their reason for accepting the speaker's telling, in creating or sustaining the trusting relationship had with the speaker will also be an essential part of the explanation of the fact that the audience comes to accept a truth, when the audience does so. This is where there is agreement. However, trust also has a normative dimension which implies that a speaker has a reason to be trustworthy even if the presumption made by the au-

dience in trust is not shared. This is because rejection of the audience's trust and the presumption it involves will provoke the audience's resentment. The expectation characteristic of affective trust is defined in these terms. It is identified as that expectation whose defeat would provoke the resentment or sense of betrayal characteristic of trust being let down. This reactive attitude is a consequence of the audience wilfully taking a risk on the speaker's goodwill partly for the purpose of affirming the relationship is one that could sustain such expectations. Since the avoidance of resentment will be in any individual's self interest, the audience's trust thereby generates an instrumental reason for the speaker to be trustworthy.

These positive and negative reasons provided by an audience's trust are quite different to one another. The latter negative instrumental reason can be captured in terms of pay-offs and self-interested motivations, and so merely introduces further circumlocutions into Grice's favoured account. By contrast, the positive reason provided by sharing the presumption the audience makes in trust—the idea that the audience's dependence can be seen as a reason to be trustworthy—suggests that there is something wrong with assumption (1) needed to generate the Acceptance Dilemma or the problem of trust for cooperation. I outline this wrong in the next and final section.

6

For a time, Grice was attracted to the idea that observance of the Cooperative Principle and maxims could be thought of as a “quasi-contractual matter”. The ‘contract’ here is to abide by the Cooperative Principle and maxims provided and because the other interlocutor equally does so. Of course, interlocutors do not explicitly reach any such agreement, it is merely *as if* they do so. Conversation could be thought of in this way, Grice suggested, because it has certain features in common with other cooperative activities. In particular, it has the feature that “each party should, for the time being, identify himself with the transitory conversational interests of the other” (Grice 1967, p. 29). However, by ‘identify’ Grice does not mean that each sees the other's interest as their joint interest; rather, Grice means that the actions of each are strategic: each attempts to be responsive to the other, knowing that the other is attempting to be similarly responsive. Cooperative activity need not imply any identification of interest; for instance, Grice gives the analogy of cooperating to get a car mended where the interests of each party “may, of course, be independent and even in conflict — each may want to get the car mended in order to drive off, leaving the other stranded.” (Grice 1967, p. 29) The mutual responsiveness found in strategic action also characterises the dynamics of affective trust. However, in this case each party does genuinely “identify himself with the transitory conversational interests of the other”. And under the present proposal, the rationality of conversational cooperation can be seen to rest on just this identification. In presuming that the trusted will see his dependence as a reason to be trustworthy, the truster presumes an identification of reasons. This presumption implies that the reason each has is not ‘quasi-contractual’, as Grice suggested. That is, the reason provided by affective trust is not a conditional reason: it is not a reason to cooperate provided that the other does so. Rather, trust provides a reason to cooperate even if this renders the trusting party vulnerable. Affective trust implies the presumption that the trusted will be trustworthy and so this vulnerability will not be exposed, but the reason it provides is not conditional on

beliefs about how the other will behave. Hollis suggests that such a solution to a problem of trust, comparable to the Acceptance Dilemma, can be put in terms of ‘team work’.

How exactly does a team solution work? Formally, it replaces conditional advice with unconditional demands. Adam can decide what to do by reflecting that the good of the team requires Adam to do *a* and Eve to do *e*, and then treat this as an unconditional reason to do his bit. It no longer worries him that he does well to do *a* only if Eve will indeed do *e*, because Eve will have a similarly unconditional reason for doing *e* as her bit. (Hollis 1998, p. 137)

Thus the presumption of relationship made in affective trust is substantial: it is that the trusted is, in certain ways, a friend.

To the extent that trust can be thick enough to involve this substantial presumption, assumption (1) that grounds the Acceptance Dilemma, or the problem of trust in conversation, should be rejected. To distinguish raw from all-in preferences is to represent all motivation in egoistic terms. On this representation, selfless motivations like altruism nevertheless function as motivations because they provide a certain configuration of desire which then moves the subject to act in altruistic ways. Crucially, a subject is only moved by their individual wants and desires. However, if the presumption made in affective trust is fulfilled, what motivates is the common interest created by the truster’s dependence. It is not merely that beliefs about the trusted’s dependence in conjunction with beliefs about the relationship had provide all-in motivation. To conceive things this way is to offer an ‘encapsulated interest’ model of trust: the trusted cooperates because an interest in maintaining a relationship with the trusted provides an incentive to be trustworthy (Hardin 2002). To conceive things this way is to return to the solution to the problem of trust outlined in section three: we cooperate when we have evidence of trustworthiness, where beliefs about the relationship had and how this can change pay-off structures provide such evidence. However, on the present account describing a relationship as trusting describes a change in how things are viewed. Where there is trust, being informative is not merely the rational conclusion of weighted all-in preferences but the rational response to the trusted’s need irrespective of further all-in preferences. Where a relationship is trusting it comes to impose certain obligations and, within certain limits, one obligation is to respond to the trusted’s dependence. What the egoistic characterisation misses is how trust creates and sustains a relationship which then imposes normative constraints in the trust situation.

Now Grice rejected this strategy for explaining the rationality of cooperation for the reason quoted in section one, namely “because there are too many types of exchange, like quarrelling and letter writing, that it fails to fit comfortably.” However, I think that this simply fails to recognise a thicker notion of trust than predictive trust. Certainly a quarrel is evidence that one’s interlocutor will not cooperate, and if trust is a matter of prediction, one will not trust one’s interlocutor to be helpfully informative. But one can quarrel and still trust, and quarrels can be part of the life of a ‘team’. When a quarrel is situated, as it normally is, within the confines of a trusting relationship each can still reasonably presume that the other will still prove trustworthy even if the quarrel is evidence to the contrary. This is possible not merely because, as is usually the case, there is further and over-riding evidence supporting this conclusion but because the reason trust provides for this conclusion is, to a certain degree, insulated from this evidence to the contrary. This is shown by the fact that one can equally choose, in trust, to give someone trusted the benefit of the doubt even if the overall evidence might support the judgement that the trusted will prove untrustworthy. Provided

the evidence does not compel this judgement, there can be rational space not to make it. Of course, there are limits here. It cannot always be reasonable to ignore the evidence; that an individual can be driven by raw self-interest can be compellingly salient, as the purchase of a used car can unfortunately illustrate. But equally, if a quarrel is persistent or pervasive enough, one might be no longer reasonable in any presumption other than that one's interlocutor would view one's dependence as an opportunity to further their own interests. In this case, there could be no presumption of relationship and trust would be unreasonable. There are limits to when trust is reasonable and limits to whom we trust. The minimal hope is that these limits are roughly congruent. That the limits of our relationships, of our sense of community or 'friends', roughly mark the limits of our trust. We are not 'friends' with everyone, and do not share knowledge with everyone; and if Grice is right, these limits are also marked in understanding. The greater liberal hope is that this circle is extensive, even as extensive as mankind, and that all dialogues have the civility of conversations.*

REFERENCES

- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Bacharach, M., and D. Gambetta (2001). "Trust as Type Detection", in C. Castelfranchi and Y.-H. Tan, *Trust and Deception in Virtual Societies*. Dordrecht: Kluwer, pp. 1-26.
- Blais, M. (1987). "Epistemic Tit for Tat", *Journal of Philosophy* 82 (7), 335-349.
- Coady, C.A.J. (1992). *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
- Cohen, L.J. (1992). *An Essay on Acceptance and Belief*. Oxford: Clarendon Press.
- Faulkner, P. (2002). "On the Rationality of Our Response to Testimony", *Synthese* 131 (3), 353-370.
- (2007a). "On Telling and Trusting", *Mind* 116 (464), 875-902.
- (2007b). "What is Wrong With Lying?", *Philosophy and Phenomenological Research* 75 (3), 535-558.
- Grice, P. (1967). "Logic and Conversation", in P. Grice, *Studies in the Way of Words*. Cambridge (MA): Harvard University Press.
- Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage Foundation.
- (2006). *Trust*. Cambridge: Polity.
- Hollis, M. (1998). *Trust Within Reason*. Cambridge: Cambridge University Press.
- Hume, D. (1777). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Clarendon Press.
- Kreps, D.M., P. Milgrom, et al. (1982). "Rational cooperation in the finitely repeated prisoners' dilemma", *Journal of Economic Theory* 27 (2), 245-252.
- Lipton, P. (1998). "The Epistemology of Testimony", *Studies in the History and Philosophy of Science* 29 (1), 1-31.
- Luce, R.D., and H. Raiffa (1957). *Games and Decisions*. Mineola (N.Y.): Dover Publications.
- Sperber, D. (2001). "An Evolutionary Perspective on Testimony and Argumentation", *Philosophical Topics* 29 (1 & 2), 401-413.
- Williams, B. (1988). "Formal Structures and Social Reality", in B. Williams, *Making Sense of Humanity and other philosophical papers 1982-1993*. Cambridge: C.U.P., pp. 111-122.
- (2002). *Truth and Truthfulness*. Princeton, Princeton University Press.

Paul FAULKNER is a lecturer at the Department of Philosophy in the University of Sheffield. He did his PhD at University College London. Among his recent publications on testimony, we find the following ones: "On Telling and Trusting", *Mind*, 2007; "What is Wrong With Lying?", *Philosophy and Phenomenological Research*, 2007; "On the Rationality of Our Response to Testimony", *Synthese*, 2002; "The Social Character of Testimonial Knowledge", *Journal of Philosophy*, 2000.

ADDRESS: Department of Philosophy, University of Sheffield, S10 2TN. E-mail: paul.faulkner@sheffield.ac.uk.

* This paper was largely written whilst I was a fellow at the Institute of Philosophy, University of London. Particular thanks are owed to Bob Stern.