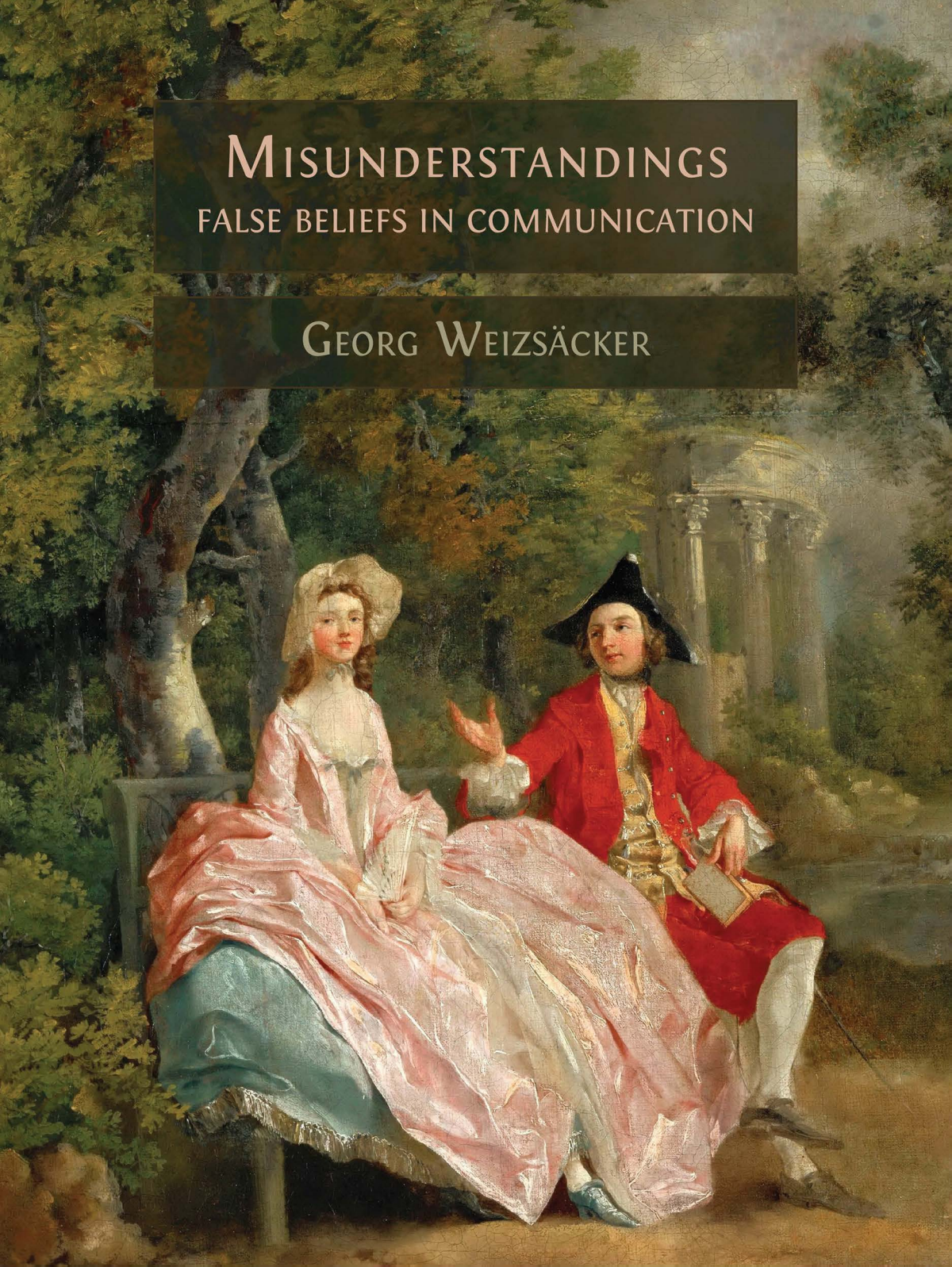
Misunderstandings

False Beliefs in Communication

Georg Weizsäcker



https://www.openbookpublishers.com/

ISBN Paperback: 978-1-80064-881-4
ISBN Hardback: 978-1-78374-051-2
ISBN Digital (PDF): 978-1-80511-138-2
DOI: 10.11647/OBP.0367

Cover: Thomas Gainsborough, Conversation in a Park (1746), Louvre Museum,
https://commons.wikimedia.org/wiki/File:Thomas_Gainsborough_-_Conversation_in_a_Park_(1746).jpg
Cover design: Jeevanjot Kaur Nagpal

The publication of this work was supported by the Open Access Publication Fund of Humboldt-Universität zu Berlin, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491192747. The publication of this work was further supported by the CRC/Transregio 'Rationality and Competition: The Economic Performance of Individuals and Firms' (TRR 190), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 280092119.

Chapter 6

Seeing what they don't see

Now we turn the table, and mentalize. We consider *their* view of the conversation and ask about their expectations. It is the book's main trick: we dig deeper by re-asking everything that has been asked before, but now about our view of the other side of the table.

The book therefore returns to square one and first considers the initial perspective – about the conversation before it even starts. What do they know when they enter the conversation, and what would they like to know?

The answer is that they know neither our actions nor the state of the world, and that they would like to know both. (So that they can pick a better action.) They will therefore consider their information about our type, which helps them to predict our action, and they will consider their information about the state of the world, which helps them to predict both of their unknowns.

Does our mentalizing reflect this?

The problem has two layers. The inner layer is that we need to form beliefs about their beliefs, in order to understand their view on the conversation. These second-order beliefs are the basis for the outer layer, which consists of our first-order beliefs about the things that *we* care about: their actions and the state of the world. The first-order beliefs are the basis for our actions and statements.

Here is the good news: the previous chapters have already dealt with the outer layer. They already asked what biases of our first-order beliefs may be prevalent and detectable.

 https://doi.org/10.11647/OBP.0367.06

As discussed earlier, the previous chapters asked all of these questions as empirical questions and did not go any further. They did not ask for the reasons of the biases, which we do now at least to some extent: if the inner layer – our beliefs about their beliefs – is off target, it can explain why the outer layer – our beliefs – is off target, too.

(Notice, here we go on re-naming: The talking beliefs and listening beliefs of the previous chapters are subsequently called *first-order talking beliefs* and *first-order listening beliefs*.)

Question 10: Do we fail to see ourselves in the mirror?

They do not know our type and therefore cannot predict our statements and actions. They cannot read our minds and cannot know our fears and desires. They have some information about all of it; luckily, much less than we do.

These are obvious facts but it is less obvious how well we factor them in. A first obstacle is that we may fail to know what information they may have about us. A second obstacle is that we may mis-interpret how they use their information. We may fail to understand how they, given what is manifest to them, predict our actions.

Ugh, what a messy, two-headed insufficiency: our knowledge is imperfect and, on top of it, our way of dealing with it is imperfect. To make things simple, the book only deals with the second obstacle. How is information used? In the case of the present question: how do we think that they use their information about our type?

(Thinking back to the previous chapters, we note that they, too, were all about the second type of obstacle. Nowhere in these chapters did we ask whether, how, or why the information that we have may be wrong. Of course it may indeed be wrong; but this is for another book.)

The first question about second-order beliefs, Question 10, asks whether we believe their belief about our actions to condition too little on their information about our type. That we may think of them as not using their information as much as they actually do. Or, the opposite bias, that we

think of them as using it more than they do. More generally, that we fail to see the degree to which they look at a conversation with us as being different from a conversation with someone else.

Expressed in the previous chapters' words: do we think that they do not discriminate well? Do we overestimate or underestimate the extent to which Question 1 is answered affirmatively, for their belief?

A possible reason for overestimating how much they regard us as special is that we know our type and it is hard to forget it. This knowledge of our type may serve as an anchor for our beliefs – here, our second-order belief. We may suffer from an "illusion of transparency", meaning that we may project our knowledge about ourselves onto the other person. We may think they know exactly how we feel.

In line with the previous chapters, however, we formulate the question by asking for a possible underappreciation – here, one that appears in second-order beliefs.

$$\text{Is } P^i_{P^j_{a^i}(\cdot|I^j_{\theta^i})} \text{ too close to } P^i_{P^j_{a^i}(\cdot)}?$$

Rachel fails to see that the governor has come prepared for the possibility that she would make a plea for funding. He has interacted with many people whose aim it is to get him to support them. Her own public profile, the governor's personal knowledge of her, and her role during the institute visit, are all consistent with the possibility that he may expect such a plea. Rachel could have anticipated that he has a tested strategy to deflect it.

Conversely, the governor does not think much about how Rachel views his personal type. He does not need to do so, anyway, since she makes no use of her knowledge about him.

On method: how does one elicit a second-order belief?

The simple answer is that the experimenter can first elicit person j's first-order belief and then ask person i what they think about the response of person j to the first-order-belief question.

Clarity of the instructions is, unsurprisingly, always important. For the procedure to be clear to the participants, person i should receive a full de-

scription of how exactly the first-order belief of person j was elicited. Ideally, person i should have the full instructions of person j available.

The measurement of second-order beliefs about ourselves considers a particular kind of meta-accuracy: we ask how well they predict us, given what they know – and how it changes if they know different things about us. A comparison of actual and believed correlations across different people can be a good measure of such second-order discrimination. Whenever different people actually view us differently, our second-order beliefs about their beliefs about us should also differ from each other, i.e., the second-order beliefs should show a non-unity correlation. If we discriminate too little, then the correlation between our (second-order) assessment on how they assess us tends to be larger than the correlation between their actual (first-order) assessments of us.

Several studies have found that people can make pretty accurate guesses about what other people think on average, but far less accurate guesses about how individual others deviate from the average. Carlson and Furr (2009) measure a respondent's assessments about how people whom they know from different contexts view them. Their questionnaire study addresses the ''main participants'' as well as several ''informants'', who are either parents or friends of the main participants. Each informant fills in a personality questionnaire about a main participant, with questions like ''On a scale from 1 to 7, can he/she easily resist temptations?''. The researchers translate these questionnaire responses into values on the Big-5 personality scales, separately by informant. The main participants fill in the same questionnaires, also separately by informant: they estimate how each informant assesses them on each question. The results show that the main participants discriminate somewhat in their second-order assessments, but not enough. For example, while they expect their parents to view them differently compared to how their friends from college view them, there is a clear under-discrimination of informants: the correlations between the two second-order assessments are far too high. Averaging across the Big-5 personality traits, the actual correlation of assements between a parent and a college friend is 0.30, whereas the main participants' meta-assessments predict that this correlation is 0.61.

Question 11: Do we underacknowledge their knowledge?

The next bit of mentalizing is about their knowledge about the state of the world. We have information about ω, and they have information about it, too. Our respective information sets are different, however, and if we do not put ourselves in their shoes and consider their view, then we may form inaccurate talking beliefs or inaccurate listening beliefs, or both.

Just like in Question 10, it is a two-step challenge. First, we need to imagine their information sets, and second, we need to imagine how they deal with each possible information set.

Our difficulty is, then, to appreciate how they use their information. Out of the countless items that are manifest to them, which ones shape their beliefs? What is the context that they focus on?

One possibility is that we may neglect that something is salient to them. We may simply not think about it and may therefore miss what they see as the point of the conversation. Another possibility is that we may expect the opposite, that they give something more attention than they actually do. Both of these effects can arise if, for instance, we simply project onto them what is salient to ourselves.

An important special case is that we may have *more* information than they. We may know something and have to lead the conversation in a way that acknowledges that they do not know it. If we do not want to give our information away, we have to be considerate about how they may learn it from the conversation. If we do want to give it away, the same is true.

Analogous arguments apply for things that are manifest to both of us but where we suspect that we pay more attention to them than they do. We can benefit from considering this asymmetry in attention and preparing our conversation accordingly. We may want to show them what things we pay attention to.

Another important case is that we may *not* know something but have to deal with the fact that they know it. Here, we have to imagine, counterfactually, each possible kind of knowledge that they could have.

All other cases of their information set being different from ours (not larger or smaller, just different), and other constellations of asymmetric attention to things that are manifest to them and us, also amount to this type of mentalizing. Depending on how carefully we imagine their information sets and how carefully we consider their beliefs for a given information set, we master our challenge of "thinking it through".

Notice a difference between first-order belief and second-order belief: the former is simpler than the latter in terms of the information conditions. The utility-relevant part of our first-order belief relates only to the information set, or circumstance, that we are actually in. There are many circumstances that we could be in, but to maximize our utility it suffices if our belief in the actual circumstance is accurate. This is different, however, for our second-order belief. There, many different information sets that the other person could be in are relevant for us. We need to consider beliefs about their belief in each one.

So, it's complicated, but we have to bite the bullet. Our best hope of making a good prediction of their belief is, indeed, to understand their possible information sets and to gauge their interpretations for each of them.

An easy special case is that a relevant piece of information is available to both interlocutors, in a way that both obviously pay attention to it. An example is the fact that Rachel and the governor are not alone in the room. This fact is important – one behaves differently in private versus in public – and it is obvious to both Rachel and the governor. It is also obvious to them that it is obvious to them, and perhaps even on higher orders of beliefs. We may call this possibility a co-presence of context.

In such a co-presence of context, there is no question about the salience of the context and we only need to assess how the other person may condition on it. However, we may still be off target in this assessment. Relevance of a context is not a black-and-white phenomenon. Recall the definition of perceived relevance, which the previous chapters introduced as a utility difference. If we fail to assess how large they perceive the utility difference to be, then we likely misperceive how they condition on the circumstance. That is, co-presence of context is a useful benchmark, but more as a qualitative idea about the coincidence of information, not so much as a quantitative property of beliefs. (And for most contextual factors, there is no co-presence anyway.)

But the idea of co-presence is a good anchor. Starting from this simple case, we can extrapolate and ask, in what ways may their information, or their beliefs, be different from ours? In what way may they not pay attention? How do they interpret what is manifest to them?

$$\text{Is } P^i_{P^j_\omega(\cdot|I^j_\omega)} \text{ too close to } P^i_{P^j_\omega(\cdot)}?$$

When judging the situation, Dimitri pays very little attention to Agniezka's beliefs. He wants something from her – that she stays on the team – and he should therefore be especially careful about her judgment of, and knowledge about, the options that the team has when going forward.

Agniezka, in fact, knows a lot. For instance, she talked to the members of the other team multiple times. Now, in the elevator, she worries about what Dimitri may know about these contacts and how he may judge her less-than-complete commitment to their own project.

A major disadvantage of laboratory research, as a method, is that one cannot be sure about external validity. This criticism applies to communication quite forcefully: humans may understand other humans better, and certainly differently, if they interact in the context-rich real world. The rather sterile laboratory environment may influence the way in which signals are used, in hard-to-predict ways. The co-presence of context may be amiss in the laboratory.

This concern may be even stronger for a study that elicits second-order beliefs. Is such an elicitation not far from real life? Do humans really think about what others think? In terms of probabilities? For each of a whole list of information sets that they may be in? All of this seems highly unlikely.

Chapter 9 will return to the issue. As a brief preview, indeed the concern is serious. It may be hard, and sometimes impossible, to elicit good proxies of the interlocutors' mentalizing about each other. But an imperfect measurement may be better than none.

Another line of defense against this criticism: laboratory studies are at least *replicable*. If one has reason to suspect that a particular feature distorts the results of a study, then one can replicate the study and leave out, or modify, the critical feature. In particular, one can modify the context. This is, of course, a variant of the main advantage of laboratory studies: control.

The researcher's imagination can be at work in a flexible playground.

Question 11 is a good example where the experimenter's control is important. Measuring how a participant views the way in which other participants use *their* information requires much transparency about what information each participant has. Experiments can provide such clarity.

Notice also that the above concern presents, in itself, an interesting set of research questions. If interlocutors indeed do not usually think through other people's information sets, then it is all the more plausible that interesting belief biases may be found.

The so-called hindsight bias is an example of information projection (''I knew it all along...'') in that we find it hard to imagine that other people, when they made their decisions in the past, did not know what we know now. With hindsight, we judge their actions unfairly or otherwise inaccurately. Camerer et al. (1989) show that the hindsight bias can appear in markets. One group of experimental participants predicts the market price of an asset at the end of 1980, being informed of all previous end-of-year prices up to 1979. A second group was informed about the event that the first group had to predict: in addition to being informed about prices up to 1979, they also saw the realized price at the end of 1980. The members of the second group were then asked to predict what the participants in the first group would predict about the 1980 price. These predictions of predictions are far too close to the realized price. The second group cannot un-know the 1980 price and tends to behave as if others knew it, too.

Question 12: Do we ignore that they judge us in context?

When predicting our actions, they factor in what they know about the world. With the same reasoning as in Question 3 we can argue that, indeed, they can predict our actions better if they discriminate between different contexts. Do we consider that they do so, to an appropriate extent?

The context helps us in this. Consciously or sub-consciously, we may "just know" what other people, in a given context, expect from us. And

even if intuition cannot be trusted, the context is still prone to influence our second-order belief. The desire to fulfill what is expected from us may guide our actions – it may be a strong motivator and it may lead us to form accurate second-order beliefs.

The literature on psychological game theory investigates the role of second-order beliefs analytically: in deviation from standard game theory, it lets the payoff that an agent receives from an interaction depend directly on beliefs. For instance, if we believe that they expect us to be kind to them, then we feel guilty if we do not live up to this perceived expecation, and we obtain a lower utility because of this feeling of guilt.

The nature of second-order beliefs is thus an important part of the equilibrium in a psychological game: guilt aversion may lead to kindness. The second-order belief may, however, differ from context to context. In a new circumstance we may think that something else is expected from us – and guilt aversion may thus be irrelevant, or lead to other behaviors.

Likewise, context determines the politeness of statements. If our second-order belief is inaccurate in a particular context (i.e., we do not understand what is expected from us) then we may say something inappropriate, or perceive their statements as putting us on the spot.

(A side note: Linguists use the word *power* to describe the possibility to influence each other's actions in a conversation. Power is still another important part of context: the social relation between talker and listener determines who can steer the other's actions. Power, in turn, is not written in stone. It is surprisingly fluid, especially during a conversation.)

Summing this discussion up, we note how widely second-order beliefs set the stage for the conversation and that they, and their accuracy, depend on context. Similar to its open-ended view on first-order beliefs, the book views these possibilities, once again, as empirical questions.

Recall, finally, that the previous two questions have already indicated a pattern of failing to realize how the other person may differentiate between their information sets: we may be egocentric and project our view of the world onto them. Our information set is only one – it is not differentiated. In contrast, their information set is variable, from our perspective. We should view their expectation about us in a differentiated way, too.

$$\text{Is } P^i_{P^j_{a^i}(\cdot|I^j_\omega)} \text{ too close to } P^i_{P^j_{a^i}(\cdot)}?$$

Steve expects that Ralph views his, Steve's, part in the conversation in light of the awkward context. He expects that Ralph expects him to make a statement that refers to Ralph's miserable state of mind, which would violate their power relation by inducing Ralph to talk about something that he does not want to talk about. Steve also expects that Ralph believes that Steve will likely gossip to other schoolchildren later, telling them about the unusual encounter. All of these second-order beliefs are a misunderstanding, as Ralph did not really expect anything from Steve.

For Ralph, in turn, Steve is nothing more than a nuisance in the present situation. He does not give much thought to Steve's beliefs about his, Ralph's, actions. He thus also fails to notice that Steve is unusually afraid of a violent outburst.

A simple way of manipulating the knowledge about others' knowledge is to create co-presence of events, as described in the discussion of Question 11. In an experiment, one can easily manipulate the context in a way that both person i and person j not only know about the manipulation, but they also know that the other knows about the manipulation, and so on.

Such a manipulation should, if possible, keep everything else constant. In particular, it should keep the material incentives constant. The manipulation of context is, then, a pure *framing manipulation*.

Dufwenberg et al. (2011) have their participants play a public-good game with different frames. In their game, each of three players chooses how much of her cash endowment to allocate to a private account, and how much to a public account that is jointly owned by all three players. Both accounts are paid out at the end of the experiment. Individually, each player has an incentive to

free ride on the other players and contribute nothing to the public account. But the public account receives a subsidy from the experimenter, who increases the account balance by 50 percent before the account is shared. This implies that it is collectively better for all players if each player contributes to the public account, the more the better. To measure expectations and their connection to the players' actions, the authors ask the participants to state both first-order beliefs about the contributions of the other two players, and second-order beliefs about the other two players' beliefs about their (respective) co-players' first-order beliefs. The authors also induce a framing manipulation where in one treatment the game is described as a ''giving'' game, whereas in another treatment it is described as a ''taking'' game. In a ''giving'' game, the positive-sounding frame may induce a more generous social norm -- which indeed appears. First-order beliefs are more optimistic in the ''giving'' than in the ''taking'' game, with expected contribution levels of 7.5 tokens versus 4.8 tokens, out of an available budget of 20 tokens. That is, the mere labeling affects how much the participants expect others to contribute. This also shows in their second-order beliefs, with an average of 8.1 tokens in ''giving'', versus 5.3 tokens in ''taking''. On average, second-order beliefs are fairly accurate, even in the way that they react to the labeling. The experiment also shows that while all beliefs are too optimistic, the differentiation between the two treatments' beliefs translates into actions: participants actually contribute 5.2 tokens on average in ''giving'', significantly more than the 3.8 tokens that the participants contribute in ''taking''.