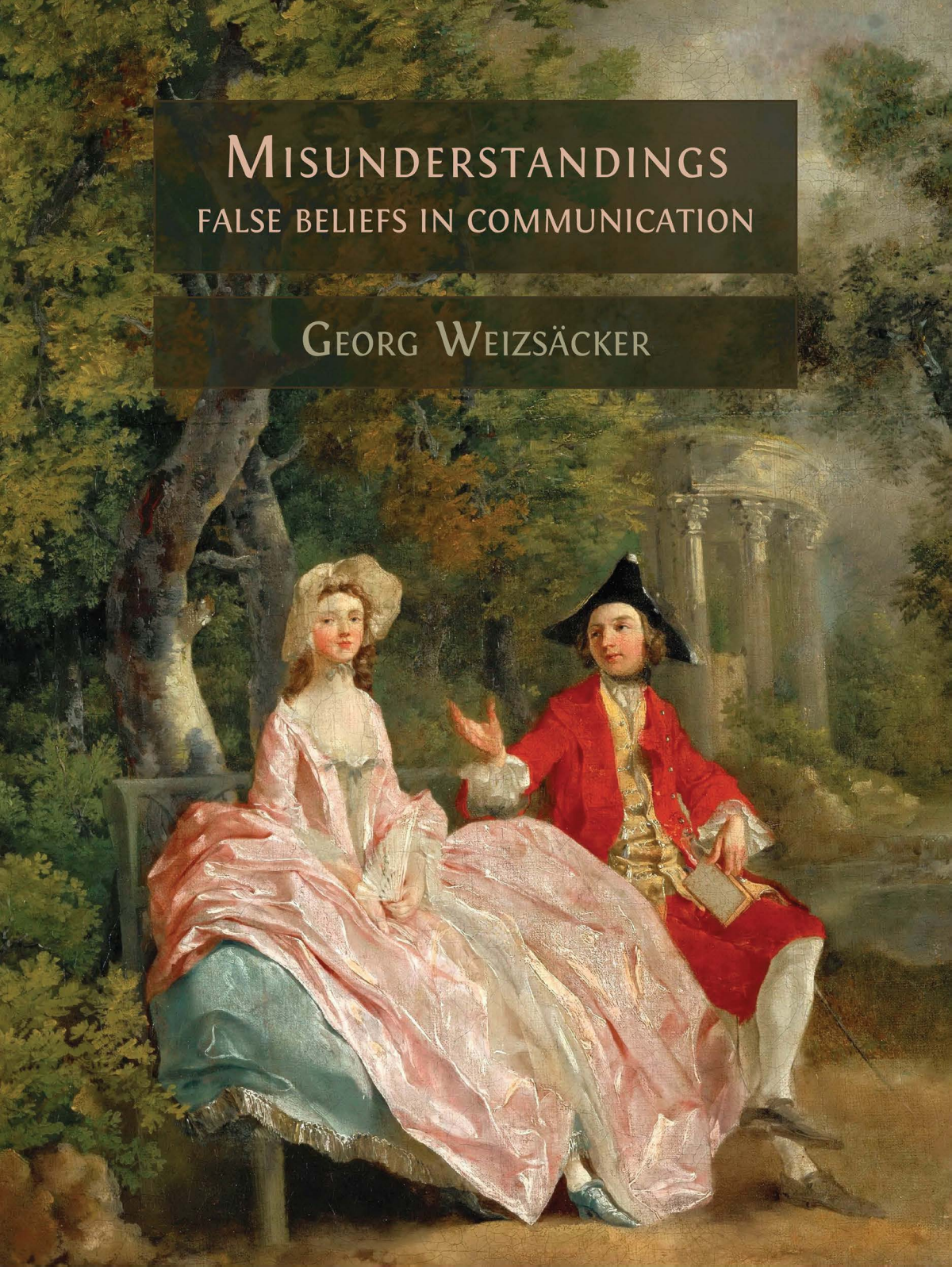
Misunderstandings

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 7

Perceiving how they talk

Continuing to apply the book's main trick, we now ask about our view of how they talk. That is, the questions in this chapter are analogous to those in Chapter 4 ("Talking") except that we consider them from the other side of the table. We listen, by mentalizing about their talking.

Some communication scientists regard the ability to listen to others as the most fundamental social skill of all. This is, in part, due to the observation that the success of a conversation lies not only in a mutual understanding of what is said, but also in the acknowledgement thereof. If the listener can signal to the talker that she is understood, both may feel that the conversation was worth its while.

As in other parts of this book, we focus on smaller steps, namely on listening beliefs. They are only one component of this feedback process: our understanding of what they say. We learn about the two uncertain aspects that their statement may indicate: their future actions and the state of the world.

The other person's challenge is quite different from ours. The only uncertain aspect that they try to predict, when talking, is our reaction to their statement. This chapter therefore asks about our belief about their belief about our reaction to their possible statements.

 https://doi.org/10.11647/OBP.0367.07

Question 13: Do we think they think we don't listen?

Earlier questions in this book focus on how our talking is explained by our first-order talking belief. Now we apply this reasoning to their talking: do we correctly predict their first-order talking belief?

Let us give this a name: our subjective expectation about their first-order talking belief is *our second-order listening belief*.

Our aim is to have an accurate version of this belief, meaning that we understand their belief with sufficient precision. This sounds doable – or not? Recall that a first-order talking belief is a rich collection of beliefs, with a separate belief for each combination of a possible statement and a possible information set that the talker could be in. Accuracy of our second-order listening belief thus asks for a lot: that we correctly predict how they predict our reaction to every possible statement, not just the one that they actually make, and that this correctness of prediction applies to each of their possible information sets.

We interpret their statement in light of this collection of beliefs. Returning to a word used earlier, this interpretation is how the statement "means" something to us.

Let us say this again, and more slowly. We have an understanding of what they expect us to do in reaction to their statement. How important this expected reaction is for them, and what they believe about the reaction, depends on their circumstance. We therefore have an understanding of the circumstances in which they would likely say what they say, and in what other circumstances they would likely not say it. Hearing their statement therefore changes our expectation about their circumstance. We learn something about their type and about the state of the world.

Notice how these arguments are intertwined with our knowledge about their incentives. What do they *want* us to do? When we interpret their statement, we consider their incentives to influence us.

The previous paragraph implies that our (first-order) beliefs about their type are an important part of our listening, too. When we assess their perceived relevance of a piece of information – and generally, when we assess

their incentive to influence us – we differentiate by their possible types. In some cases, we have a pretty good idea about their type anyway. In other cases we do not, but we can analyze the situation type by type, and piece together by aggregating over the type-specific predictions. In any case, we need to consider each of their possible types separately, one at a time.

Let us first consider the case that our incentives are aligned well enough. This means that the same actions that benefit us, relative to other actions that we could choose, also benefit them. If so, then they may want us to learn something. They may believe that if we knew about a certain circumstance, we would follow our common (aligned) incentives and choose the action that is optimal in this circumstance.

They may therefore choose a message that is informative: a message that is distinctive for the circumstance that they want us to learn about.

This sounds nice and easy, but it asks for a lot of consistency between their beliefs and our beliefs. Perhaps, it asks for too much? How would they know which of their statements leads to what possible reaction from us? Do we know the extent to which they believe it? Only if their belief about our reaction coincides with our belief about their belief about our reaction do we indeed understand them. Quite a requirement! If, in contrast, the beliefs do not coincide, our understanding fails. In other words, the extent to which our second-order listening belief is accurate influences the extent to which their statement is informative to us.

What if our incentives are not aligned enough? In this case, we cannot both gain from such a coordination. They may try to induce us to choose something that is good for them – but bad for us. We, in turn, may or may not understand this. Notice how, perhaps surprisingly, this amounts to the exact same question that we discussed in the previous paragraph: our understanding of their circumstance relies on the accuracy of our second-order listening belief. This property holds even if our incentives are not aligned. And, the same reasoning also applies in all cases where our incentives are partially aligned, i.e., aligned for only some of our actions or only for some of their circumstances.

Careful: in this discussion, there does not exist a "truth" of statements. Statements can be freely associated with the circumstances in which they are uttered. What we, as the listener, understand to be the meaning of a statement is merely the connection between the statement and the circumstance

in which we believe them, the talker, to make the statement.

But words like "truth" or "lies" pop up in our mind: we wonder about the correlation between statements and circumstances – and the fact that we wonder about it helps us, in particular, to understand whether they want to trick us.

Always, let us keep in mind that accuracy of our second-order listening belief is different from accuracy of our first-order listening belief. Accuracy of our second-order listening belief is about our understanding of what they believe. Accuracy of our first-order listening belief describes our expectation of their actions and the state of the world directly, by saying that this expectation conditions accurately on the statement and on our information.

The two concepts are not even nested. An accurate first-order listening belief may occur together with an inaccurate second-order listening belief – we may correctly understand the implication of a statement even if we do not understand what they think about our reaction. Likewise, an accurate second-order listening belief may occur together with an inaccurate first-order listening belief – we may understand what they expect from us in response to a statement even if we don't know what it indicates.

But the two beliefs are connected in a simple way: the inner layer is the basis of the outer layer. Our belief about their future actions and about the state of the world depends on our belief about their belief. In this sense, the second-order listening belief *causes* the first-order listening belief.

Causation is a strong claim, so let us be slower and more precise, once again. We assumed in Chapter 2 that both interlocutors not only maximize subjective expected utility, but also know this property about each other. This implies that person i's first-order belief about person j predicts a best response to person i's second-order belief: the talking strategy that person i predicts for person j maximizes person j's payoff, given person i's knowledge about this payoff and given person i's second-order belief. The first-order belief is, thus, tied to the second-order belief. May this connection – strictly speaking, a logical implication – suffice for using the word causation.

(But notice also that the second-order belief does not fully determine the first-order belief. Even with a given second-order belief, our first-order belief about the other person's type, or our a-priori belief about the state of the world, may both be more or less correct. This variability – a degree of free-

dom in the analysis – exists independently from second-order beliefs and it therefore enables the possibility that first-order beliefs are inaccurate despite the presence of accurate second-order beliefs, or vice versa.)

Summing all of this up, it is plausible in many conversations that an inaccuracy of second-order listening beliefs (at the inner layer) leads to an inaccuracy of first-order listening beliefs (at the outer layer): if we do not understand what they expect from us, then we likely fail to interpret well what they say.

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

While Ralph understands that Steve expects him, Ralph, to make a decision about the ball game, he misperceives Steve's first-order talking belief about what would have happened after a statement that enquires about Ralph's state of despair. Ralph therefore does not understand that Steve's question is a defensive statement that he uses to avoid incurring Ralph's wrath.

For the purpose of measurement, a big plus of asking about second-order belief accuracy – the congruence of person i' second-order belief and person j's first-order belief – is that we compare like with like (belief with belief).

Strictly speaking, the previous paragraph is utterly mistaken: the second-order belief is a probability distribution of a probability distribution of an action, which is harder to describe than a first-order belief. The latter has one "a-probability-distribution-of" less.

But we can rescue the mistaken paragraph. To make the elicitation of a second-order belief manageable for the participants in an experiment, one can ask for the means of all relevant distributions, as point beliefs. This can be done in simple words ("What do you think is their prediction about your reaction, on average?") that do not even require that the participants realize the complications of distributions over distributions. It is a quick-and-dirty solution, but one that is unlikely to distort the results.

Another practical issue is that the number of beliefs that are to be elicited is getting larger and larger. Talking beliefs include not only the (believed) reaction to the statement that the talker actually makes, but also those regarding the alternative statements that she could have made. If the talker,

on top of all this, also faces separate possible circumstances (constellations of information sets) then separate second-order beliefs need to be elicited from the listener, one for each circumstance. This may require a lot of data collection.

A nice trick to reduce the amount of data is to use a stylized game where everything is mirror symmetric. That is, to use a game where symmetric constellations of the states of the world, the talker's statements, and the listener's reactions to the statements, lead to identical payoffs. In such a game, it suffices to measure fewer beliefs and make inferences about the "mirrored" constellations.

Grabova et al. (2023) conduct a constant-sum game where the talker knows the state of the world, which could be A or B. The listener only knows that the two states are equally likely to occur. The talker's two possible messages are also labeled ''A'' and ''B'' and she sends one of these messages at will. Messages have no direct payoff consequences but the listener chooses one of two payoff-relevant actions, which are also called ''A'' or ''B''. The symmetry is completed by allowing only two payoffs in this game, such that one player is the winner and earns the high payoff, whereas the other loses and receives the low payoff. The identity of winner and loser is state-action dependent: in each state, the listener wins if his action matches the state and the talker wins if the listener's action does not match the state. The game thereby induces a very transparent incentive for the talker to try and mislead the listener about the true state of the world. Since the game is symmetric regarding the labels ''A'' and ''B'', the entire data collection can be done under the scenario that one particular message, ''A'', is been sent, and all answers can be applied also to the other possible message. That is, the sender is only asked: ''Suppose that you sent the message 'A'. How likely do you think it is the receiver chooses 'A'?'' For second-order beliefs, the listener is asked to predict the talker's answer to this question. All beliefs about states/messages with ''B'' can be inferred from this, under the reasonable assumption that the participants have beliefs that do not depend on the label. In effect, the experiment provides a simple set of measures for first-order and second-order beliefs about truth telling and trust. The results show that second-order listening beliefs are fairly accurate on average: the average predictions of the listener's probability

of trusting the message is close to one half, in both the talker's first-order talking belief and the listener's second-order listening belief. One half is also the equilibrium prediction -- a babbling equilibrium. A data pattern that contradicts game theory is, however, that the heterogeneity in listeners' second-order beliefs moves in the wrong direction: listeners who believe that the talker expects them to trust the message do indeed trust the message with higher likelihood. This is inconsistent with the incentives in the game because a talker who first-order believes that the listener trusts his message is more likely, not less likely, to lie. (This is theoretically and empirically true.) A listener who has such a second-order belief should therefore be less, not more, trusting. The data thus indicate a limitation of asking for second-order listening beliefs: at least for some of the experimental participants, the reported beliefs apear to be poor proxies for the participants' second-order beliefs. An alternative explanation -- one that the book's next question will pick up and that is even less consistent with the theory -- is that these participants do not base their decisions on second-order beliefs at all.

Question 14: Do we think they talk to us as strangers?

Our type is always with us; it is near impossible to forget. Can we nevertheless take the talker's perspective, with limited knowledge about ourselves, while listening to them? Otherwise, we may take things personally when they are not.

Notice how this can also have a warm, positive ring to it: we may *like* them talking to us in a personal way.

One such possibility are white lies, i.e., deceptions that are for our, the listener's, benefit. Many white lies are about our own type: they flatter us. We enjoy hearing them.

Do we also believe them? Not in the sense of learning about us. We have superior knowledge about our type. This book even takes the extreme, simplifying view that we have perfect knowledge of our type. But we do learn about their belief about our type. The white lie may indicate that they have

a positive view of us. We may not have been sure of this before, and enjoy the update.

Is it justified, though? Under what conditions may we believe a white lie? To find an answer, recall the nature of the second-order listening belief: our belief about their belief about our reaction. Recall also that our second-order listening belief is a wide collection: it specifies what we expect them to expect from us for every possible statement that they may make and for every possible information that they may have about us. If we update from the white lie, then we must believe that there is a connection between our reaction and their circumstance, in their minds: the way in which they expect us to react must co-vary with their belief about our type. Upon hearing their statement, we can thus infer something about how they view us.

For concreteness, let us suppose that a stylized information structure governs all beliefs. We know our type and they, the talker, could be in two possible information sets: one that indicates to them that we are "for them" (i.e., that our incentives are aligned with theirs) and one that indicates to them that we are "against them" (...not aligned with theirs). Our second-order listening belief specifies what we expect them to expect us to do in each case.

Let us also consider the case that we believe that they, in both information sets, believe us to show a positive reaction – one that increases their utility – to the particular statement that we call a white lie, and a less positive reaction to not telling it. This would be a natural belief for us to have, given that the white lie flatters us.

But this second-order listening belief is not suitable to justify our credulity: since we believe them to best respond to their first-order beliefs, we must believe that they tell the white lie in both of their information sets. In each of them, they expect a positive reaction from us, so their incentives do not change across the two sets. Hearing them tell the white lie is therefore not informative for us.

That is, we cannot believe the white lie, in this case. A second-order listening belief of the above-described kind is too simple. We need to find a more involved justification. What other second-order belief would do the job?

Well, involved or not, the belief would have to be somewhat peculiar. It would need to prescribe that their belief about our reaction differs between

the two information sets – although by assumption, we *cannot* differentiate between the information sets (otherwise we would not need to learn).

Likely, such a peculiar second-order listening belief would be inaccurate: it is plausible that their actual first-order talking belief does not differentiate much between their information sets. If so, then we have a misunderstanding – a mismatch between our second-order beliefs and their first-order beliefs.

The accuracy is, of course, an empirical question. It may well be that their first-order talking belief *does* differentiate between their information sets, at least to some extent. Also, recall that this is all a stylized example with a particular information strucuture. Other situations may make different predictions.

Yet, the stylized example makes a general point: thinking through the logic of white lies makes them less believable. We should realize that they have an incentive to tell the lie, and that we should not update too much because their incentive to tell the white lie may not change much with the information that they have. It is therefore hard to justify our credulity with plausible second-order listening beliefs. Rather, they may be off target, or we may disregard them altogether.

Importantly, the same reasoning often applies to "black" lies, i.e., those that do not benefit us if we believe them. Here, too, we should consider their beliefs about our reaction. Here, too, we may fail to notice that these beliefs may not depend much on their information. The talker often has, very simply, an incentive to lie to us. We often have, very simply, no way of detecting the lie.

Lie detection has other interesting aspects, too. We may believe that they do not like to lie. That is, we may attribute a preference type to them that gives them a low utility if the statement is misleading (or uninformative) about their beliefs. We think of them as being truth tellers.

This rather optimistic view is perhaps a more plausible description of our view of them, but it does not describe second-order listening beliefs. The belief that they are truth tellers is a belief about their direct utility from making a certain statement in a certain circumstance. This belief does not rely on their beliefs about our reaction.

We can summarize the discussion by saying that it is plausible that we

believe a lie – nothing keeps our first-order listening beliefs from prescribing that we update – but if this credulity stems from our second-order listening belief, then it may well be inaccurate, i.e., we plausibly mispredict the other person' first-order talking belief.

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

Agniezka hates aggression and feels vulnerable when the gloves come off. Dimitri has observed ample evidence of her sensitivity in the past, and is aware of it. He does not, however, expect Agniezka to judge their collaboration based on the aggressiveness of his tone. Agniezka, in contrast, wonders how aware Dimitri is of her sensitivity. She expects that if Dimitri was aware of it, he would have considered his aggressive battle speech to prevent her from staying on the team. Based on this second-order listening belief, Agniezka interprets Dimitri's talking about "blood" as revealing his unawareness of her sensitivity, which makes the possibility of staying on the team even less attractive to her.

Lying and lie detection are tightly connected to the two interlocutors' second-guessing of each others' types. Misunderstandings about preferences can induce misunderstandings about the credibility of an utterance and make it to be perceived as a lie. All this is not only a qualitative discussion (lie or not; white lie or black lie): misunderstandings about preferences also determine the size of the harm, or the size of the benefit, that a lie causes. Dishonesty and distrust may be more or less hurtful, and the pursuit of a perceived mutual goal may make us feel more or less good.

Measuring this in a laboratory experiment is difficult. Many of the larger effects of a lie are emotional. One cannot install a non-trivial knowledge structure about the participants' emotional types, and one cannot even induce emotions sufficiently well. The experiments are simply too small.

Psychologists have developed a technique that nevertheless succeeds in detecting the interlocutors' emotions about each other's actions: perspective taking. All participants receive the same description of a realistic but hypothetical situation. This situation can describe any interaction of two people, including an emotionally loaded interaction. Half of the participants are asked to write a text about the situation, written in first person singular from the angle of one person in the described situation. The other half of

participants do the same, but from the other person's perspective.

After the participants have taken their perspectives in this way, the experimenters can ask them about views, beliefs, preferences, etc., and measure the extent to which all of these vary between the two groups.

It is reasonable to assume that real-life differences in perspectives are more extreme than what one may find in such an experiment. After all, the participants are not really part of the described situation. Any effect of perspective taking is therefore likely to be a lower-bound estimate of the effect in a naturally occuring situation of the same kind.

Gordon and Miller (2000) describe to their participants a situation of a couple who are together for about a year. One partner, the ''lie receiver'', observes that his/her loved one -- the other partner, or ''lie teller'' -- has a restaurant lunch with a former boyfriend/girlfriend. The lie teller had not told the lie receiver about the lunch beforehand and does not even admit to it later, when asked explicitly how he/she had spent the lunch hour. The lie receiver reacts negatively and, so the story goes, the relationship's continuation is now in danger. The partipants in the experiment all read the same text, creating the same knowledge about the facts of events, including the lies that were told. They are randomly assigned to take the perspective of the lie teller or the lie receiver. (These labels are not used in the experiment -- all individuals in the story have names.) A questionnaire then elicits the participants' views about the two partners' actions, in 14 questionnaire items. Via a factor analysis, the 14 items are projected on four main factors, three of which describe whether the actions were justified, misunderstood, and common. A regression analysis shows that being randomly assigned to the perspective of the lie receiver creates the view that the lie was significantly less justified, less misunderstood, and more common, compared to the case that one is randomly assigned to take the perspective of the lie teller. The experiment thus illustrates how the judgements of the appropriateness of statements can vary between the talker and the listener even if they have identical information about the facts of a situation. In this story, the lie teller is unlikely to realize the psychic damage that the lie does and the lie receiver is unlikely to be aware of this. However, the analysis refers to a comparison of first-order beliefs and second-order beliefs only

`somewhat vaguely and indirectly, by comparing the perceptions of how much the lie is viewed as justified, misunderstood, and common.`

Question 15: Do we think they predict our reaction independent of context?

Context adds a vast richness to language, which makes listening hard: much harder than deciphering a code or following a convention. No code or convention is rich enough to cover all possible meanings that arise through combinations of utterances and contexts. Understanding means interpreting their statement in light of the context.

Our second-order listening belief is suitable to capture this – because it is rich, too. They, the talker, consider the context in their first-order talking belief; our belief about their first-order talking belief allows us to consider exactly this fact. We interpret their statement by guessing how they expect us to react, depending on the circumstances that they may find themselves in.

The richness of their context-dependent talking implies, however, that our listening is ambiguous. There are too many dimensions of context that we may or may not consider in our beliefs about their beliefs. (Ambiguity may not be a bad thing. Poetry's meaning is up for grabs but this does not make poetry less valuable – quite the opposite.)

In any case, listening is more valuable if our second-order belief considers context well, i.e., if we realize what dimension of the state of the world their first-order talking belief conditions on. How do we identify this dimension?

To answer this, it is useful to recall Chapter 4's discussion of perceived relevance. As talker, we choose our statement because we anticipate that in light of a particular context, the statement generates a response from them that enhances our utility. Now, as listener, we understand that their first-order belief is analogous. They choose their statement because they anticipate, in light of a particular circumstance, a utility-enhancing response from us. They perceive some piece of the context as relevant. We listen well if our second-order listening belief reacts to this piece, too.

In this search for context, it helps that we already know their statement.

We know that they, as talker, choose this statement to generate the best possible outcome for themselves. We examine our belief about their belief to assess why they regard this statement as optimal.

This reasoning helps us to reduce the ambiguity. They could have said so many other things – the opportunity cost of talking. We can assess this opportunity cost: our second-order listening belief indicates the extent to which they believe that each of the alternative statements would have generated an alternative benefit. Our second-order listening belief thereby justifies their choice of the particular statement that they make, and it reduces the set of candidate contexts along the way: the talker must perceive their statement to yield a higher benefit than each of the alternative statements, and the context that they condition on must contribute to this perception.

Paraphrasing this in simpler words: We may know what they are talking about because our second-order belief, in connection with the statement that we hear, indicates what context is perceived relevant for them.

Or, we may fail to know it. As always in this book, it is an empirical question.

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

The governor does not take the issue of stolen artifacts very seriously. His second-order listening belief fails to examine a particular dimension of context: Rachel and he already have a previous thread of conversation, namely the once-started initiative on cultural policies. Normally, it would be in Rachel's interest to talk about this topic. (She could ask him to fund it, too.) Nevertheless, she decides to talk about the stolen artifacts, which signals her sense of urgency for this new topic. It is likely that she has reasons to regard it as more relevant than the old one.

Context variation can provide a benchmark for comparison. "To put something in context" often means that we compare the size, or importance, of an item with another item that the context provides. In a laboratory study, the experimenter can measure the reaction to a stimulus in the presence or absence of such a quantitative context.

This simple trick can also be used in experimental studies of language evolution. A sizable set of research pursues this approach, some of them by

studying non-human animals or children.

Barner and Snedeker (2008) provide an example that demonstrates how 4-year-old children's understanding of words reacts to context information. The authors run an experiment that investigates how the children connect the context-dependent adjective ''tall'' with a novel set of objects that come under the made-up name ''pimwit''. When is a pimwit tall? The task of a participating child is to examine 9 such novel objects, the pimwits, and classify each of them as tall or short. (To connect this situation to our discussion, we view the child as the listener and we view the experimenter as the talker who says: ''Point at pimwits that are tall.'' The child's reaction reflects his or her second-order belief about what the experimenter expects him or her to do.) In the experiment, pimwits are little figures of varying height between 1 and 9 inches, with a one inch difference in height among each pair of adjacent pimwits. The average pimwit has, thus, a height of 5 inches. To create context, a second group of four distractor objects is placed alongside the nine target pimwits. These distractor object are of a similar physical style as the pimwits but they are not labelled at all and the children are not asked to evaluate them. Instead, the distractor objects only sit on the table, acting as a potential reference set for the nine target pimwits. In the SHORT DISTRACTOR treatment, the distractor objects have an average height of only 1.25 inches, whereas in the TALL DISTRACTOR treatment, the distractors' average height is 8.75 inches. Indeed, the children classify many more of the nine target pimwits as ''tall'' in SHORT DISTRACTOR than in TALL DISTRACTOR: the average heights of the marginal (smallest) pimwits that are classified as tall in the two treatments are 5.4 and 8.4 for SHORT DISTRACTOR and TALL DISTRACTOR, respectively. The experiment thus shows that even at this early age, the interpretation of language is context dependent. (Or especially at this age?) A further variation shows that the children also use the precise specification of the context: in a variant of SHORT DISTRACTOR, the distractor objects are endowed with their own name, ''tulvets'', and are painted in a different color from that of the target pimwits but have the same heights as the distractor objects in SHORT DISTRACTOR. In this condition the marginal height of the pimwits that the children classify as tall is significantly higher, at 6.9 inches. This result is similar to another experiment where no distractor objects are

shown at all, yielding an average height of the marginal ‘‘tall’’ pimwit of 7.2 inches. That is, the 4-year olds appear to use the distractor objects as a comparison set only if their specification suggests to do so. Overall, the children show remarkable reactions to context in their second-order beliefs. The experiment cannot give evidence, however, on whether these second-order beliefs are accurate, as it does not measure the talker's first-order talking belief.