

# Understanding, Charity, and Interpretation <sup>1</sup>

DAVID ROSENTHAL

To determine whether one disagrees with somebody, one must have a sound understanding of what that person is saying. But it is sometimes difficult to distinguish cases of misunderstanding others from cases of substantive disagreement. One can best resolve that quandary by making every effort to understand accurately, and that requires construing what the other person says as charitably as possible. But such charity is more complicated than often recognized, since it applies in several ways: in respect of how others use individual words, what assertions they make, and what inferences they regard as acceptable. And applying charity must rely on how one sees the relevant things; charity is unavoidably by one's own lights. An additional type of charity applies to the background explanatory pictures that sometimes figure in what people say. Construing in respect of one factor can affect how we construe in respect of another. So tradeoffs in how one construes the several factors are sometimes necessary, which can even lead to equally good alternative construals. Charitable construal of our own thoughts also occurs, often unconsciously, in resolving difficulties and, more generally, in guiding and regulating our thought processes.

## 1. Misunderstanding and Disagreement

We typically think about verbal communication in terms of the everyday conversations we have with one another. Such conversations typically flow smoothly and effortlessly, and rarely require or involve conscious reflection and

---

<sup>1</sup> Extensively revised from a presentation to the Hamilton College Humanities Forum on Translation and Cultural Exchange, October 17, 2011.

what to say. So we often pay relatively little attention to the difficulties or failures in understanding one another that can occur in everyday conversation.

Still, it is not all that unusual for mutual misunderstanding to occur in daily conversation. Such misunderstanding may well just pass unnoticed, since the main desideratum in such conversations is often successful social interaction, and that encourages the parties simply to ignore misunderstandings that do not matter in some special way. It can be easier and more comfortable to let minor misunderstandings pass and continue as though mutual comprehension is unimpaired. The parties will often just move on when they find what somebody said to be unclear or possibly mistaken, again unless getting things right matters for some reason. In these ways we tend to treat conversing as largely effortless and transparent.

Getting things right does often matter when people are discussing something that they especially care about, and then potential misunderstanding or disagreement can compel attention. This often happens with political issues or with matters of moral or aesthetic judgment. But it may also happen just with everyday decisions, for example, about what to do. And disagreements occur strikingly often in academic exchanges, especially in the humanities and social sciences, where there may seem to be fewer accepted ways to settle an issue.

But it can sometimes be difficult, in any of these cases, to tell whether people are disagreeing about something or just failing to understand one another. It might seem to one that another person would readily agree with one if only they understood what one is saying. But it can also be tempting to see somebody as disagreeing with one in a substantive way, even when a more careful and sympathetic exploration would reveal that the other person is simply using words in a relevantly different way.

And no matter how thoughtful one is about such things, it is not always all that easy or straightforward to distinguish accurately whether somebody disagrees with one or instead misunderstands what one is saying. That is so not only with relatively abstract matters of politics, aesthetics, or morality, but also with relatively mundane matters. It can even happen that one person sees misunderstanding where the other sees substantive disagreement.

The difficulty in distinguishing disagreement from misunderstanding in interpreting somebody else's written work, say, in a political or legal or academic context, can be especially intractable because then there is no back-and-forth exchange that could help resolve things. The frustration that arises from reading

something that seems wrong can impede any effort to understand the other person's remarks as generously as one might when one actively interacts with somebody. So there is a tendency to see substantive disagreement in the written work of others. It is somewhat unusual that commentary on written work explores the possibility that something that seems wrong might be due simply to the author's way of putting things is different from one's own.

Looked at from the perspective of somebody not engaged in a commentary or other exchange, this can seem unfortunate. When viewed by somebody who is independent and reasonably knowledgeable about the matter at hand, it may well seem that avoiding misunderstanding would not be all that difficult if the people involved were not so dug in on construing any apparent difference as due to a failure on the part of the other person.

## **2. Identifying Misunderstanding**

How should we go about distinguishing misunderstanding and disagreement? When two people seem to differ in what they are saying, how can we determine, in both mundane and more controversial contexts, whether that difference is due to their not understanding one another or, instead, to their actually disagreeing with one another? These questions arise both for one looking at an exchange from the outside and for those who are party to the exchange. And the difficulty can be challenging. Indeed, the interplay of misunderstanding and disagreement may even make it seem that people agree when they do not.

It is especially important to resist jumping to any conclusion about disagreement vs. misunderstanding, since irrelevant factors can bias one in one direction or the other. One natural approach would be to see if one of the two possibilities can be eliminated, thereby establishing the other. This can happen with patient continued conversation, which sometimes shows that the appearance of disagreement was illusory, and that the apparent difference was solely due to misunderstanding. But if success in dispelling disagreement remains elusive despite sustained effort, it may be natural to conclude that the difference does reflect genuine disagreement.

But any such conclusions will be reliable only to the extent that there is a good-faith effort to avoid misunderstanding. And that can be more complicated and challenging than might sometimes appear. A serious source of difficulty is that misunderstanding can result from several different factors. Perhaps the most

straightforward and obvious factor is that people sometimes use words in ways that differ, and it is not always transparent when that happens. The ways the parties use the relevant words may overlap enough to conceal the operative differences.

But though differences in the way words are used may well be the most salient source of misunderstanding, it is by no means the only potential source. People may use the individual words in the same way, but still assign different truth conditions to various sentences that contain those words. The two parties differ instead about what would make various sentences true. And that would lead each party to misconstrue the assertions that the other party makes in using such sentences.

The assignment of truth conditions not only affects speech acts of assertion, but also affects all the other types of speech act so brilliantly studied by J. L. Austin (1961/1979), including but not limited to questions, requests, and expressions of doubt or wonder. All these require interpretation that rely on the truth conditions that a speaker associates with the corresponding assertoric sentences.

It may seem that assigning different truth conditions to the same assertoric sentences would be somewhat unusual. But it very likely happens a lot more than we recognize. Many cases are doubtless simply written off impatiently as substantive disagreements, rather than as a semantic divergence about how to evaluate assertoric sentences for truth. Differences in the way words are used are a lot easier to detect than semantic differences in the assignment of truth values. But they can be crucial in revealing and disarming cases of genuine misunderstanding

Semantic differences in the use of individual words and in connection with assigning truth conditions to sentences do not exhaust the potential for causing misunderstanding. Assertions are often linked together inferentially, and people can differ about which inferential connections they accept. And those differences can be altogether independent of how the parties use individual words or assign truth conditions to sentences.

Differences about which inferential connections the parties accept can be more difficult to detect than the other two, because it typically emerges only when one person offers reasons for some belief, relying on some inference, and the other person rejects that inference and so denies the cogency of those reasons. But however difficult to detect, differences about whether to regard an inference as acceptable can readily result in mutual misunderstanding.

So misunderstanding can result from differences of three distinct types, about the way the parties use individual words, about how they assign truth conditions to sentences, and about how they evaluate the acceptability of inferences. And these three ways in which people's use of language can differ will often interact in ways that make identifying each more difficult and less accurate, thereby amplifying the risk of mutual misunderstanding.

### **3. Construing Others**

I'll return to these three sources of misunderstanding in section IV, and an additional source in section V. But it will be useful in the meantime to highlight the crucial tie between understanding, on the one hand, and interpretation and translation, on the other. When I say something and you grasp the meaning of what I say, you in effect make what I say your own in a particular way.

It is not that you thereby endorse what I say or even think it is sensible. But you cast what I say in your own terms, using words in ways that you would in talking about the matters under consideration. You in effect translate what I say into words that are natural to you. Sometimes the words you use to render what I say will be exactly the same as the words I used. But they need not be; indirect discourse often uses words different from those that would figure in a direct-discourse quotation.

All understanding of what somebody else says, even in the most basic and mundane cases of communication, is in this way unavoidably a matter of translation of what that person says into one's own terms. One cannot grasp what somebody else says except in terms that are natural and make reasonable sense to one. Understanding what somebody else says is interpreting what that person says, by putting it into words one could oneself use to talk about the matter at hand. It is a type of translation, even if only natural language is involved. We can think of it as translation from one idiolect into another.

Because we typically think of translation as being from one natural language to another, it typically goes unnoticed that we must translate from one idiolect to another when both belong to a single natural language. But all mutual communication and mutual comprehension consists in translation between idiolects, whatever the natural language. And all failures of understanding are failures of such translation between idiolects.

The only anchor one has about what words mean, what the truth conditions of sentences are, and what inferences are acceptable is our own usage in respect of words, assertions, and inferences. When things go smoothly and there are no differences to be resolved, we each tacitly take for granted that others operate with word meaning, truth conditions, and inference validity exactly as we ourselves do.

But that assumption is reasonable and appropriate only so long as there are no differences in what the parties say. When differences do arise, so that we seem to disagree or misunderstand the other person, we must do our best to render what that person says in terms we ourselves would comfortably use and readily understand. There is no other way. We cannot interpret what others say except by casting what they say in our own terms. The only way to understand what others have in mind by what they say is to find the most sympathetic and successful translation of what they say into terms we find natural enough to be comprehensible.

Our own use of words, assignments of truth conditions, and acceptance of inferences are the inevitable anchors for grasping what others say. That does not undermine at all the thoroughgoing social nature of our use of language. The way we use words, assign truth conditions, and assess inferences are all deeply influenced by how others use language. Our linguistic interactions with others form the dispositions we have in respect of those three aspects of how we ourselves use language. And the way that interacting with others effects our linguistic dispositions will interact with the way one construes others and the way other seem to construe us. Because of all that, we end up using words, assigning truth conditions, and assessing inferences in ways that overlap very heavily with the way others do.

Still, those ways of using language will rarely if ever be exactly the same from one individual to another. The linguistic dispositions each of us forms by interacting with others will always be to some degree idiosyncratic, leading to the differences that obtain among idiolects. And if the resulting idiolectal idiosyncrasies are not recognized and adjusted for, they can readily result in failures of mutual understanding.

When people have fluent command of the same natural language and engage in casual conversation, it will typically seem that they do not differ in their use of words and the way they treat sentences and inferences. They easily and naturally react to one another's remarks, saying what comes to mind.

But when differences that matter in some way do occur, so that there is reason to assess whether the difference is due to misunderstanding or substantive disagreement, each party must interpret the other into their own words. And all anybody in the end has to go on in is one's degree of success in rendering others' speech into words we would use ourselves and they will take to be accurate.

#### **4. Charity of Translation**

To maximize understanding what another person says, one must construe how that person uses words, assigns truth conditions, and evaluates inferences. These three factors reflect three units of meaning that operate our thinking: term-sized units, sentence-sized units, and inference-sized units. Maximizing success in understanding somebody requires construing that person's remarks so that those three factors differ as little as possible, ideally so little as not to be noticeable by either party.

Any such construal is inevitably rendered in one's own terms. How then does one seek to minimize divergence? To understand the other person, we want to do maximum justice to what that person is saying and has in mind. But since one cannot get directly inside the other person's mind, what can guide the way one construes the other's remarks? The considerations discussed in the last two sections will help.

Consider first the term-sized factor of now the other person is using individual words. We all tend to use words in ways we believe conform to standard norms for those words. So it is natural to construe another person's use of words as reflecting such norms. But it can sometimes be that we can make the best sense of what somebody says by seeing that person as using words in a somewhat different way. And we can seek to maximize successful understanding by ascribing to the other person the use of words that seems to make the best sense of what that person says. We can in this way do our best to put ourselves in the other person's communicative shoes in respect of the way that person uses words.

But how does one determine what makes the best sense of the other person's assertions? The only way is what seems to one to make the best sense. One must rely on what makes the best sense by one's own lights. There is nothing else one can go on. We must assume that the other person intends to use words in a way that makes the best sense. So it is charitable to construe that person's use of words in that way. But because we cannot get into the other person's thought processes,

our only access to what makes the best sense is in the end how we see the relevant matters.

Often the best way to construe another person's words will be the same as the way one would oneself use those words, what W. V. Quine describes as "the normal tacit method of homophonic translation" (1960, p. 59). But when it seems that what the person says makes better sense if some relevant words are construed differently, we should go with that construal. This constitutes charity for the construing of the other person's words.

Much the same points hold for how we can best assign truth conditions to the sentences another person uses in making assertions. We should assume that people intend to use words in a way that makes good sense. So we can best construe somebody's words in whatever way makes the best sense of what they say. Similarly, we should assume that people also intend that their assertions are true. Charity of construal will require that we take the assertoric sentences others use to be true as often as we reasonably can.

Again, that is truth as one sees the relevant matters. One should, as much as possible, construe others' assertions as being true by one's own lights. One has in the end no access to what is true apart from one's own best estimate about what is true. There is no other standard by means of which one can construe as true the sentences others use to make assertions.

The tendency to construe as true the sentences others use to make assertions is reasonably natural and automatic. When somebody says something that strikes us as wrong, we often think, "What that person must really mean is such-and-such." We in effect slot in truth conditions for the other person's sentences that would make the relevant assertions true by our lights.

This natural urge to maximize the truth of what other say is often referred to as the principle of charity (Wilson 1959; Quine 1960; Davidson 1973-1974). But it is only one of charitable construal. Construing another person's use of words as making the best sense of statements that contain those words is another, term-sized form of charitable construal.

And things can get complicated with two distinct ways of construing charitably, one pertaining to term-sized units of meaning and the other to sentence-sized units. When another person's assertions do not seem true to one, one can adjust the meanings of some constituent words or the truth conditions one assigns to that person's sentences, or even both. Sometimes it will be obvious what the most charitable adjustment is, but in many cases it may not be. One has then simply



to do one's best with the two available types of adjustment, and be sensitive to the tradeoffs that can arise between adjusting for word meaning and for truth conditions.

The complications in implementing an overall charitable construal increase dramatically when we take into consideration the reasons people give for saying things. Such reasons rely on inferences from reasons to the statements for which they are reasons. So we must then add a principle of charity for the construing of others' inferences. We must not only construe words so as to make the best sense of what others say and truth conditions so as to maximize the truth of others' assertions; we must also construe the inferences they make as being acceptable as often as possible. And again, our only measure of acceptability for inferences will rely on the way we see things.

We must accordingly figure out how to balance all three types of charitable construal. Inferences can be good even if not all their constituent sentences are true. So maximizing acceptability of inferences and maximizing truth of assertions are independent. And it may be easiest to see a person's inference as acceptable if one takes some constituent sentence not to be true, or the sentences as all true if one sees the inference as faulty. Indeed, that is often what happens when others seem to agree with us for what we think are the wrong reasons.

Charitable construal is often taken to be a matter of maximizing simply the truth of another's assertions or simply the acceptability of the other person's inferences. Apart from leaving out the need to construe meanings of the other's words, any one-factor version of charitable translation inevitably overlooks the need to balance the three types of charity so as to produce most charitable overall version of what the other person says.

Indeed, the need to balance those three factors has to my knowledge not been explicitly noted in the literature. But once one notes that charity must apply in these three independent ways, it is obvious that tradeoffs between the three will sometimes be necessary. This makes it crucial also to recognize that charity does apply in all those ways.

When it does seem necessary to strike balance among the three, moreover, there may be alternative ways to do so that are equally charitable, so that there is no single best construal. It can happen that we get a good overall construal by taking all the assertions to be true but sacrificing the acceptability of some inferences, but that we get an equally charitable overall construal by saving validity and giving up some truth.

If there are alternative equally charitable construals, continued conversation or written exchange may point to one overall construal that is preferable to all the parties concerned. But it can also turn out that such additional interaction does not suggest a single best construal.

The occurrence of alternative best construals may sometimes explain how there can be equally compelling but incompatible readings of texts, something familiar to anybody who has worked extensively in the history of philosophy (Rosenthal 1989, §4). And it can likely also explain incompatible readings of arguments and positions in academic, political, or legal contexts that one cannot decisively adjudicate among.

Relying on charitable construal is at bottom seeking to do the greatest justice possible to the other person's remarks. This idea is by no means new, and arguably has strong roots in antiquity. John Stuart Mill's famous remark that "[h]e who knows only his own side of the case, knows little of that" (1859/1991, 42) is in effect an injunction to maximize charity in construing the views of those with whom one disagrees. And Mill traces this approach to Cicero, presumably because Cicero, like other Academic skeptics, held that one should be able to argue either side of any issue, doing justice to each (Remer 2022, esp. p. 212; Brittain and Osorio 2022). Mill does not note, nor to my knowledge does Cicero, that this requires balancing the charitable construal of words, sentences, and inferences. But the foregoing considerations make it plain that it must.

Because charitable construal must rely on how the person doing the construing sees things, the interpretation of historical texts, whether in philosophy, law, politics, or other matters, will always in that way be "Whiggish," to use Richard Rorty's useful term (1979; see also Rosenthal 1989, §§3-4). Such Whiggish interpretation in no way suggests or warrants simply reading one's own views into historical texts. It is only that the way we understand historical texts unavoidably relies on what sentences we take to be true and what inferences valid, and our use of individual words.

It is sometimes urged that interpreting historical texts requires extensive immersion in the writings of the relevant period and tradition. Such immersion can indeed be useful, and perhaps sometimes indispensable. But such immersion requires one to interpret everything one reads in the relevant corpus. And it unavoidably relies on charity in respect of word meaning, truth, and inference. Immersion in a historical period must itself proceed by charitable interpretation.

## **5. Charity and Background Theories**

The three forms of charity described so far are characterized by the semantic unit they apply to, words, sentences, and inferences, respectively. But there is a fourth factor that can figure in the way we understand what other say, and which calls for a type of charitable interpretation. This fourth way pertains to the general beliefs people have, often somewhat theoretical in nature, which sometimes inform and underlie what people say and think. People's assertions and inferences often reflect some theoretical picture about the phenomena under discussion, and then one may not accurately understand what they say without taking into consideration that theoretical picture.

This requires a fourth type of charity in interpreting what others say. We must determine what background theoretical picture is most likely operative, relying on which background picture makes the best sense of what people say. Charity here may be more exploratory and speculative than with the use of words, truth conditions of sentences, and validity of inferences. In those three cases, people explicitly use particular words and sentences, and typically make their inferences explicit as well.

By contrast, the background picture that informs their statements and inferences often is not explicitly stated. In that case one must be sensitive to whether some theoretical picture is operative, and if so what it is. It will seem sometimes that the most charitable construals of words, sentences, and inferences do not by themselves make good sense of what somebody says, and that can suggest that the person has some tacit background picture in mind.

One can then charitably consider likely hypotheses about what picture would make the best sense of what the other person is saying. A person's remarks will sometimes clearly suggest some underlying theoretical picture, and sometimes an explanatory picture will be front and center in what the person says. But often it will be only tacit. And in those cases, charity can provide a strong and useful guide about what picture is operative. One ascribes the tacit background picture that makes the best sense of what is said.

Charity here as always inevitably rests on how the person doing the interpretation sees things. One must ascribe whatever background theoretical picture seems to one to make the best sense of the other person's remarks. One might not endorse the background picture one ascribes to the other person; indeed, it may well be a picture that one roundly rejects. But it must seem to one to make

the best sense of what the other person says. And once again, there may be tradeoffs between the background picture one ascribes and the charitable construals of words, truth conditions, and inferences.

If one sees such background theoretical pictures as simply sets of assertions related by inferential connections, they would not require a distinct type of charitable interpretation. One could then simply be charitable about the other person's assertions and inferences. But there is more to such background pictures than simply the truth of a number of assertions that are bound by inferential ties. Such pictures also play a distinctive explanatory role in connection with the relevant phenomena. It is that explanatory role that unifies and individuates such background pictures. And it is such explanatory roles that figures in the charitable detection of what background picture makes the best sense of what another person is saying. It is in that way that appeal to such background picture calls for a distinct type of charitable interpretation.

An example may help. In describing others in folk-psychological terms, we take for granted the posits, assumptions, and explanatory framework of whatever folk-psychological picture one has. And the specifics of what is assumed in somebody's folk-psychological picture can vary from one person to another.

Some today might tacitly fold in neo-Freudian ideas that others would roundly reject. And some religious convictions might also affect one's folk-psychological picture, thereby making a difference to how one describes others folk-psychologically. In talking with somebody, one will sometimes recognize that they simply see how people operate folk-psychologically in ways different from our own take on those things. Charity in determining another person's take on the explanatory framework of folk psychology may then be needed for a correct construal of what that person is saying.

There are other examples. Some pertain to how we see the workings of social groups, and to how we think social groups should operate. Some people think things work best politically and socially when government has a minimal role in people's lives, whereas others believe that government and other organizations should provide support to people when needed. Some see political and social matters in a largely conservative way, interpreting things in distinctively traditional ways, whereas others interpret social and political matters in a more open-ended way to accommodate and even encourage change. Some tend to picture social interactions on the model of life in small towns, whereas others picture them in a large and diverse metropolitan environment. Failing to take

account charitably of such differences can impede an accurate understanding of what others say.

Yet other examples figure in academic contexts. Theoretical considerations can underlie what somebody says in an academic dispute, and understanding those background considerations can then be pivotal to an accurate grasp of what that person says. Doubtless that sometimes also occurs with the aesthetic and moral judgments people make. And it is said by many that people who grew up in different generations sometimes have a background picture distinctive of their generation, and which affects many things they say and think. In all such cases it can be important in fostering mutual understanding to take charitable account of differences in background pictures.

Interpreting what somebody says is inextricably bound up with interpreting whatever theoretical approach they have about the things under discussion. One often cannot do one without doing the other. This observation recalls the argument Quine calls pressing from above for the indeterminacy of translation, which involves interpreting some relevant theory held by the person being interpreted (1970, 183). And it is of a piece more generally with Quine's insistence that no serious line can be drawn between semantic and theoretical considerations (Quine 1951). Meaning and belief interact in ways that require that one construe the two together as a package.

## **6. Some Observations**

As noted in section IV, the first three types of charity operate on distinct units of meaning, term sized, sentence sized, and inference sized. But the fourth can also be seen as pertaining to a distinctive unit of meaning, that which can sustain a fabric of theoretical explanations. Theorizing can be formulated inferentially, with postulates that lead to the phenomena to be expressed, but it need not be. It can instead simply reflect theoretical claims with explanatory connections to the relevant phenomena.

One can see the contrast between inferences and a more general set of explanatory connections in the way Plato distinguishes in Book VI of Republic between the type of intellectual activity he sees as occurring in connection with the third level of his divided line and that which is characteristic of the fourth. The third level involves inferences from hypotheses that are treated there as beginning points (511a-b). The fourth level, by contrast, embeds hypotheses in an

explanatory fabric that treats them not as beginning points, but as genuine hypotheses to be explored and evaluated: “truly as hypotheses—stepping stones to take off from” (Plato 1992, 511b5, p. 195). If we put aside Plato’s ontological commitments, we have a clear contrast between inference and a fabric of explanatory connections.

The structural analogy between types of charity and Plato’s four levels also holds for the first two forms of charity. We apprehend the things that occur at Plato’s second level by way of beliefs, which represent in a sentence-sized way. By contrast, the first level is populated by images (*εἰκόνας*; 510a1), which represent in a term-sized way (509d-510a).

The most common way to represent how words are used is with the lexical entries in a dictionary. But those entries are in effect abbreviations of sentences that say that the target word means what the lexical entry specifies. And any other way of specifying the use of individual words is likely also to involve sentences. It can be explicit, as in saying ‘The word ‘dog’ refers to domestic canines’ or even ‘Dogs are domestic canines’, or tacit, as in uttering ‘dog’ while pointing at a dog, which is a nonverbal way of saying that ‘dog’ refers to that type of thing.

But that does not affect the distinction between word-level charity and sentential charity. The sentences that specify how an individual word is used are readily distinguishable from sentences in which that word figures in saying something else. So we can apply charity about individual words independently of charity about truth conditions.

Charity of construal is a tool for maximizing mutual understanding. But mutual understanding may not be always one’s overriding goal in how one construes what somebody else says. In competitive exchanges, such as debates or legal proceedings, for example, the desire to score points can override the wise injunction of Mill and Cicero to arrive at an accurate understanding of what one’s opponent says.

There are, however, other more interesting and intellectually challenging ways in which one’s primary goal may not involve having an accurate understanding of others. Harold Bloom (1997) has advanced the striking and provocative idea that writers often tacitly, though also actively and creatively, misread those of the predecessors who most influence them. Bloom has mainly the literary world in mind, and in particular poetry. And he sees this process as driven in large measure by one’s feeling that those who most influence one have

already scaled the relevant artistic heights, leaving one unclear how one might make one's own mark. Hence Bloom's idea of the anxiety of influence.

Bloom sees literary figures as often in effect engaging in a kind of intellectual struggle with those who most influence them, which leads them to misread those predecessors, but in ways that can be highly compelling. Such agonistic but creative misreading of others in seeking to make one's own mark likely occurs in other intellectual enterprises as well. Academics in the humanities and social sciences often focus on aspects they reject in the writing of those who have most strongly influenced them, sometimes resulting in new and revealing approaches to the relevant topics.

As noted at the outset, when somebody says something that strikes one as wrong, it can be unclear whether to see oneself as misunderstanding what is said or as simply disagreeing in a substantive way. Bloom casts readers as simply settling agonistically for disagreement with what others write, especially those who most strongly influence one, and work off those disagreements to develop their own views.

But even in their writings of those who most deeply influence one, there is typically more than enough for one to disagree with. So even if one interprets those works charitably, resulting in an accurate understanding of them, there should be ample opportunity, as well as significant motivation, for one to elaborate one's own views. Bloom's creative misreading should not be necessary. And working off of real disagreements, as against those that one agonistically misreads, would likely result in developing views that are more stable and compelling. It may even be that doing so is a lot more common than Bloom imagines.

In ordinary conversation as well as in more formal interactions, we expect that others will construe us charitably in the ways described above. So if somebody were to respond to what one says in a way that suggests total lack of concern with charitable construal, it would seem at best strange, and likely leave one with a sense that something is significantly amiss.

But that is actually the way it often seems when in the exchanges one has with current artificial intelligence programs (AIs). The large-language models (LLMs) on which AIs run do generally produce responses that address the topic at hand, but those responses often do not seem specifically directed to what one has said. However relevant the machine's response may be to the subject matter, there is often an odd sense of disconnectedness. The machine does not seem actually to understand what one says, as against simply picking up on what topic one is talking

about. Even when the machine replies to a direct question, there is often something off about the response.

Given the way current AIs operate, that is to be expected. The gigantic libraries of texts that constitute LLMs are structured to generate new text that pertains to the topic raised by whatever input it receives. Things are moving fast, and there is now some hope that AIs can simulate some cases of human reasoning (e.g., Metz 2024).

But generating a sense of human communicative interaction is another matter. A machine could appear to have a genuine understanding of what a person says to it only if it can perform charitable assessments of the inputs it receives. That would require evaluating inputs for both truth and validity, and current LLMs do not have anything like the resources needed to do so. Without an ability to interpret inputs charitably, the machine's responses will seem detectably artificial in a significant way.

Alan Turing's classic article (1950) urged testing a computer program for intelligence by seeing whether people would take its responses to be those of another person. Turing was careful to specify that this test, now generally known as the Turing test, should be thought of as only testing whether a machine exhibits intelligence, and not also be whether it can think. Turing held that the question whether machines can think is "too meaningless to deserve discussion" (1950, p. 42; see Mitchell 2024).

But perhaps we can reduce skepticism somewhat about that latter question. Charitable construal may not be necessary in testing for intelligence, but it likely is pivotal in testing for thinking. So if a machine did generate responses to human inputs that led humans to see those responses as reflecting a charitable construal of those inputs, that might encourage a sense that the machine is thinking, or at least doing something very close.

Developing programs that could engage in charitable construal is by no means impossible. But it would not be trivial. The program would begin with need a database that enabled it to identify things commonly taken to be true and false. But it would also have to be able to evaluate commonsense inferences for validity. And doing so would require access to the connections among words that sustain commonsense inferences, which are relatively holistic in nature. Constructing a database for that would be no mean feat. So the appearance of thinking in machines, as against mere intelligence, is likely not around the corner.



## **7. Self-construal and Thought Processes**

Charitable construal maximizes success in understanding what another person says. And since what somebody says expresses their thoughts, charitable construal enables one to grasp the other person's thoughts. What about one's own statements and thoughts? Does charitable construal have any application there?

We rarely have any need to construe our own individual words. If I say 'The bank is nearby' I know which meaning of 'bank' I have in mind. And there is typically no need to construe one's own speech acts. Every thought that is expressed in a speech act is conscious (Rosenthal 2005, ch. 10), and being conscious means that the thought appears in one's stream of consciousness (Rosenthal 2022). And knowing in that way what thought is expressed by what one says typically obviates any need for one to construe one's own speech acts.

But things are not always so straightforward with one's conscious thoughts. One can be confused about what one consciously thinks. And one may sometimes have a sense that one's conscious thoughts about a particular matter do not all fit together smoothly or that what one thinks about something is not quite right, even if one cannot see quite what the difficulty is. One's conscious thinking may also sometimes lead to a problem or dead end, which one cannot see how to resolve or get around.

It sometimes happens in such cases that taking a break or getting a good night's sleep results in resolving such difficulties even without one's giving the matter any further conscious thought. One will then not be aware of what mental process led to one's seeing clearly how to think about the relevant things; evidently unconscious thought processes enabled one to see how things fit together or to solve the problem. Often this leads to one's having thoughts that are different, even if only slightly, from the thoughts one had earlier.

How does an unconscious thought process work to resolve such difficulties? The unconscious thought process plainly must take into consideration the main thrust of one's thinking. One's general approach to the matter at hand does not change; it is simply that difficulties with affected that approach get resolved. So the unconscious thought process must be sensitive to most of the things one holds true in that connection, and most of the relevant inferences. And taking those things into consideration in effect constitutes operating charitably with the relevant thoughts.

Such charitable assessment of one's own earlier thinking is necessary if the unconscious thought processes are to resolve a difficulty with that earlier thinking, as against simply coming up with new ways of thinking about the matter at hand. To come up with a salutary resolution of whatever the difficulty was, the unconscious thought process must charitably assess how best to make those thoughts fit well together.

Such unconscious charitable construal of one's own thoughts would operate much as one charitably construes remarks made by somebody else. The charitable construal of one's own thinking operates on most of the relevant thoughts, but charitably modifies one or more beliefs or inferences in the light of the overall body of beliefs and inferences that one is committed to about the matter under consideration. Such charitable self-construal can in that way dispel confusion, resolve difficulties, and solve problems. And it results in one's in effect finding out, as one might put it, what one had really thought all along about the matter at hand.

Charitably construing others is unavoidably anchored in what assertions one sees as true and what inferences as acceptable. Similarly, imposing rationality on our own thinking, by dispelling confusion and resolving difficulties, must also rely on how we ourselves think about the issues under consideration. New thoughts can dispel confusion or resolve difficulties only if they fit comfortably with one's overall approach to those issues.

Charitable assessment of one's own thinking can also occur consciously. Sometimes that will simply be to get one's conscious thoughts in order and make their inferential connections more transparent. Other times conscious self-assessment may lead one to modify some of one's thoughts, to eliminate some belief or inference as mistaken, or to come up with new thoughts altogether.

Thinking always involves adjusting one's thoughts and coming up with new thoughts. So the charitable assessment of one's own thoughts figures not only in dispelling confusion and resolving difficulties, but more generally in guiding and regulating our thought processes. And our ability to regulate our own thinking by such charitable self-assessment likely develops hand in hand with our ability to interpret others. This is an important way in which our thinking is responsive to the interactions we have with others.

Self-construal of one's thoughts also occurs in connection with consciousness. For a thought to occur in one's stream of consciousness is for it to appear subjectively that one has that thought. So one must subjectively represent

oneself as having that thought, which is in effect to construe oneself as having it. But the charitable self-assessment of one's thoughts that occurs in resolving difficulties need not be conscious. For a thought to be conscious, one must explicitly represent oneself as having that thought, even if that representing does not itself occur consciously (Rosenthal 2005, chs 2-4).

But the self-interpretation that does occur when a thought is conscious introduces a second type of confusion that can occur in our thinking. That self-interpretation can itself go amiss, so that one is confused about what it is that one actually thinks. In such cases charitable self-construal can again help induce clarity about what one's thoughts actually are.

We take mutual understanding in conversation so much for granted that we seldom reflect on what such understanding consists in and what it requires. Without a solid grasp of how mutual understanding does operate, we can often be at a loss when misunderstanding occurs. The various mechanisms of charitable construal provide the needed theoretical framework. And those charitable operations are also fundamental to guiding the way we ourselves think.

City University of New York, USA

## References

- Austin, J. L., *Philosophical Papers*, Oxford University Press, 1961/1979.
- Bloom, Harold (1997), *The Anxiety of Influence: A Theory of Poetry*, 2<sup>nd</sup> edn., Oxford: Oxford University Press.
- Brittain, Charles, and Peter Osorio (2022), "The Ciceronian Dialogue," in *The Cambridge Companion to Cicero's Philosophy*, ed. Jed W. Atkins and Thomas Bénatouïl, Cambridge: Cambridge University Press, 25-42.
- Davidson, Donald (1973-1974), "On the Very Idea of a Conceptual Scheme," *Proceedings and Addresses of the American Philosophical Association*, 47: 5-20.
- Metz, Cade (2024), "OpenAI Unveils New ChatGPT That Can Reason through Math and Science," *New York Times* (September 12), <https://www.nytimes.com/2024/09/12/technology/openai-chatgpt-math.html>
- Mill, John Stuart (1859/1991), *On Liberty*, in *On Liberty and Other Essays*, ed. John Gray, Oxford: Oxford University Press.
- Mitchell, Melanie (2024), "The Turing Test and Our Shifting Conceptions of Intelligence," *Science*, 385, 6710 (August 15): DOI: 10.1126/science.adq9356
- Plato (1992), *Republic*, tr. G. M. A. Grube, rev. C. D. C. Reeve, 2<sup>nd</sup> edn., Indianapolis: Hackett Publishing Company.
- Quine, W. V. (1951), "Two Dogmas of Empiricism," *The Philosophical Review*, 60, 1 (January): 20-43.
- Quine, W. V. (1960), *Word and Object*, Cambridge, Massachusetts: MIT Press.

- Quine, W. V. (1970), "On the Reasons for the Indeterminacy of Translation," *The Journal of Philosophy*, 67, 6 (March 26): 178-183.
- Remer, Garry (2022), "Philosophy, Rhetoric, and Politics," in *The Cambridge Companion to Cicero's Philosophy*, ed. Jed W. Atkins and Thomas Bénatouil, Cambridge: Cambridge University Press, 200-214.
- Rorty, Richard M. (1979), *Philosophy and the Mirror of Nature*, Princeton: Princeton University Press.
- Rosenthal, David (1989), "Philosophy and Its History," in *The Institution of Philosophy*, ed. Avner Cohen and Marcelo Dascal, Peru, Illinois: Open Court, pp. 141-176.
- Rosenthal, David (2005), *Consciousness and Mind*, Oxford: Clarendon Press.
- Rosenthal, David (2022), "Mental Appearance and Mental Reality," in *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*, ed. Josh Weisberg, Cambridge: Cambridge University Press, pp. 243-271.
- Turing, A. M. (1950), "Computing Machinery and Intelligence," *Mind*, 59, 236 (October): 433-460.
- Wilson, N. L. (1959), "Substances without Substrata," *The Review of Metaphysics*, 12, 4 (June): 521-539.

Reports from the Department of Philosophy  
Vol. 52

**TRUE COLORS, TIME AFTER TIME**  
Essays Honoring Valteri Arstila

Edited by  
Alexander D. Carruth,  
Heidi Haanila,  
Paavo Pylkkänen,  
Pii Telakivi

University of Turku  
Finland

Copyright © 2024 Authors

Cover Image: Polaris Koi

**SERIES EDITORS:**

Olli Koistinen  
Juha Räikkä

Department of Philosophy  
University of Turku  
FI-20014 Turku  
Finland

ISSN 1457-9332  
ISBN 978-951-29-9959-0 (PRINT)  
ISBN 978-951-29-9960-6 (PDF/Online)

Painosalama Oy, Turku 2024