The Illusory Transparency of Intention: Does June Understand What Mark Means Because He Means It?

Boaz Keysar

To cite this article: Boaz Keysar (2000) The Illusory Transparency of Intention: Does June Understand What Mark Means Because He Means It?, Discourse Processes, 29:2, 161-172, DOI: 10.1207/S15326950dp2902_4

To link to this article: http://dx.doi.org/10.1207/S15326950dp2902_4

Published online: 08 Jun 2010.

Submit your article to this journal

Article views: 71

View related articles

Citing articles: 8 View citing articles
The Illusory Transparency of Intention: Does June Understand What Mark Means Because He Means It?

Boaz Keysar

Department of Psychology
The University of Chicago

Keysar (1994) reported a perplexing phenomenon: Readers treated communicative intentions as transparent. When a protagonist experienced a negative event and left a sarcastic message, readers thought the addressee would perceive sarcasm more than when the event was positive and the same message sincere. This occurred even when the addressee did not know the protagonist’s attitude. Gerrig, Ohaeri, and Brennan (this issue) argue that this phenomenon is methodologically and theoretically flawed. Here I argue that their criticism is unfounded. I explain how the design firmly supports the phenomenon and the theoretical challenges that the phenomenon poses. As I argued before, the phenomenon can be explained by the standard view and by a simpler theory. The challenge is to discover which theory correctly accounts for the phenomenon and more adequately describes language use. I conclude that such nontrivial phenomena should not be shunned but welcomed as useful vehicles for theory evaluation.

I stumbled on a phenomenon by accident. Truly believing in the standard theory of language use, I constructed an experiment that assumed language users rely on mutual knowledge when they take the linguistic perspective of others (e.g., Clark & Marshall, 1981; Clark, Schreuder, & Buttrick, 1983; Greene, Gerrig, McKoon, & Ratcliff, 1994). A pilot study indicated that this assumption might not be warranted. The experiments that I reported (Keysar, 1994) investigated this further and revealed what seemed quite counterintuitive to me, who had always believed in the standard view.

Correspondence and requests for reprints should be sent to Boaz Keysar, Department of Psychology, The University of Chicago, 5848 South University Avenue, Chicago, IL 60637. E-mail: boaz@ccp.uchicago.edu
In these experiments, participants read that June recommends to Mark a certain restaurant. Mark has dinner there and he either loves it ("positive event") or hates it ("negative event"). The next day he leaves June a note saying, "About that restaurant, it was marvelous, just marvelous." Mark was being sincere in one case and sarcastic in the other. The interesting finding is that even though participants are aware that June does not know his actual attitude about the dinner, they still reported that she would perceive the precise intention that was actually intended. They tended to think that June would take him as sincere more when he intended his utterance sincerely than when he was being sarcastic. It was a very strange result. It is as if participants take Mark’s intention to be transparent, despite the ambiguity of his utterance, and despite June’s lack of access to critical information: hence the name of the phenomenon, the illusory transparency of intention. In my opinion, the standard theory that assumes that readers rely on common ground, should have been uncomfortable in the presence of these results.

Gerrig, Ohaeri, and Brennan (this issue) attempt to defend the standard view from the threat of the phenomenon: “It is exactly because Keysar’s phenomenon provides such a strong challenge to extant theories of language use that we opted to examine it more closely.” Their examination resulted in the conclusion that “Keysar’s (1994) demonstration of what he termed illusory transparency of intention was methodologically and theoretically flawed.” In this article, I argue that their criticism holds no water, but that the resulting debate can indeed be useful.

The methodological “flaw” that Gerrig et al. (this issue) identified is a “lack” of a baseline in the original experiments. I will show that the issue of baseline is a red herring and should be discarded. The alleged theoretical flaw is not identified in the Gerrig et al. article. Instead, the article attempts to defend the standard view by explaining how it can account for the phenomenon. This explanation does not add much information because it was already contained in my original article. Gerrig et al. also reported an experiment that supposedly undermines the phenomenon. I explain why the experiment does the exact opposite. I also discuss the paradigmatic implications of the approach I took in my 1994 article and explain the great advantage that the field of language use can gain from the exploration of phenomena such as the illusory transparency of intention.

METHODOLOGICAL ISSUE: ILLUSORY METHODOLOGICAL “FLAW”

The best demonstration for the validity of scientific results is that they are replicable by an independent laboratory. Gerrig et al.’s (this issue) first experiment provided just such support for the phenomenon. It is a perfect independent replication. Yet Gerrig et al. offered a strong methodological criticism: They suggested that the experiment lacks a crucial control and baseline condition. If they are correct, then the very existence of the phenomenon is in question. I argue that they are wrong and
that the criticism is based on a perplexing confusion. Therefore, I take the opportunity to clarify the methodological issue and draw more general conclusions about the role of baseline conditions in experimentation.

Is There a Need for Control?

The answer is no. The manipulation is very simple. The scenarios that participants read were identical except for the event information (positive vs. negative). The only thing that varies is the readers’ knowledge of the event, which suggests Mark’s attitude. So in this design, scenarios serve as their own control. One uses a control condition when there is a need to control for a particular confound (see a detailed discussion in Keysar, 1997). If there had been a potential confound in the Keysar (1994) experiments, then there would have been a need for control.

Is There a Need for a Baseline?

Though Gerrig et al. (this issue) called for control and baseline conditions, it seems that they actually mean that the experiments need a baseline, not a control. The second experiment that they reported is again a replication of my original experiments with an added baseline condition: a scenario that does not include the event information, so the readers do not know if Mark liked the restaurant or not. This would provide a measure of the inherent bias of the scenarios. Would participants think that June would perceive Mark’s attitude as sarcastic without their knowing what happened at the restaurant? As it turns out, on the average, participants are just as likely to attribute the perception of sarcasm to June when they do not know what happened as when they know he hated the restaurant.

Do we need this baseline? My argument again is unequivocally no. This is a perfect example of a baseline that provides no information. A baseline is needed in experiments for various reasons. An experimenter might want to evaluate the effect of different conditions with respect to a baseline. A baseline could be used to norm materials in order to preselect items for the main experiment. I demonstrate that these needs do not apply here. More important, I first offer a thought experiment that provides a principled argument that the phenomenon cannot benefit from a baseline.

What is the phenomenon that needs to be explained? The scenarios manipulated the event valence to induce two different intentions: sarcastic intention when he hated the restaurant and sincere intent when he liked it. The difference between the two conditions is the phenomenon: Participants were more likely to think that June would perceive sarcasm when Mark intended it than when he did not. A different way of putting it is that they were more likely to think that June would perceive sincerity when he intended it than not. It does not really matter in which direction one compares the conditions; it is the difference between the two conditions that makes the whole phenomenon. This is true independent of any inherent bias.
Scenarios vary in their bias. This is a fact of life in studying discourse. The elegance of the original experiments is that the design successfully avoids the issue of bias. Let me explain how. Consider Figure 1. This is the hypothetical result of experiments such as Gerrig et al.’s (this issue) second experiment. The y-axis represents the percentage of cases in which participants attribute to June the perception of sarcasm; the two original conditions are represented as sincere for the positive event information condition and sarcastic for the negative event information condition. The phenomenon is represented as the difference in percentage points between these two conditions. The condition labeled B7 represents Gerrig et al.’s findings: Without any event information, the results are identical to the sarcastic condition, the negative event information. Indeed, this reveals that the scenarios had a negative bias. With these particular scenarios you do not need to tell participants that Mark hated the restaurant; they think June will perceive sarcasm regardless of the event information.

Why might this be a problem for my original experiments? If participants can figure out that he was being sarcastic without their privileged event information, then they should also conclude that June has the same ability to detect the intention, and therefore they attribute the perception of sarcasm to her as well. Unfortunately, this logic doesn’t hold under scrutiny. Recall that participants’ task is to assess June’s perspective. Given that the bias is inherent in the scenario and is independ-
ent of the privileged event information, then from June’s perspective the answer should be the same even when readers learn he had a positive experience. So the problem does not change with the newly discovered baseline. The phenomenon still needs to be explained: Why is there a difference between the sincere condition and the B7 bias condition?

To see the “in-principle” irrelevance of the bias, consider the following thought experiment. Suppose that instead of discovering that the bias is strongly negative, as in B7, Gerrig et al. (this issue) had discovered that the bias is strongly positive, as in B1 (see Figure 1). In this hypothetical case, the baseline would have come out identical to the sincere condition. In that case, one would need to explain why the sarcastic condition is different from the baseline. The difference between the B1 baseline and the sarcastic condition is identical to the difference between the B7 baseline and the sincere condition. So the same difference would need explanation with this hypothetical baseline as with the actual baseline.

It does not matter if the bias is strongly negative or strongly positive, and it would not have mattered had it been anywhere in between. Consider Figure 1 again. Suppose that the bias had been equivalent to the B2 baseline; in this case, one would have to explain why participants attribute to June fewer sarcastic perceptions in the sincere condition and many more in the sarcastic condition. The same problem exists for the hypothetical B3 baseline. One could extrapolate the same for the B4, B5, and B6 baselines. The striking thing about Figure 1 is that as the hypothetical bias changes in our thought experiment, the phenomenon stays the same because the difference between the end points does not change.

The conclusion from our thought experiment is that Gerrig et al.’s (this issue) discovery of negative bias adds no information. In fact, it would not have mattered what the results of their experiment had been. Had their results come out like B1, B2, B3, B4, B5, B6, or B7, it would not have mattered. The difference between the two original conditions still needs to be explained.

Can the experiments benefit from a “neutral” baseline? Not really. One reason to have neutral baselines in experiments is to see if conditions differ from it. In the experiments that demonstrate the illusory transparency of intention, such a comparison would not be meaningful. Consider the hypothetical B5 and B4 baselines in Figure 1. B5 is one type of neutral baseline because participants are as likely to attribute the perception of sarcasm as sincerity. B4 could be considered neutral because it is equally distant from the two conditions. Now B5 is closer to the sarcastic condition than to the sincere one. Could we therefore conclude that positive privileged information has a stronger effect than negative information? Not really. Narratives vary on so many dimensions that one can write them to have any bias one wants. For example, one can write scenarios to have a 50–50 baseline (as in B5). A comparison to such a neutral condition, then, would not be informative, for the exact reason it would not be informative to compare the two conditions to a B4 type of neutral baseline.
Ceiling or floor baselines are not a problem here. Perhaps it would have been desirable to prenorm the items and select only those items that are not biased either way? One might think that with such items one could more convincingly demonstrate the existence of a phenomenon, perhaps because it would be more difficult to demonstrate the phenomenon with a truly unbiased set of materials. Gerrig et al. demonstrated that there is a “ceiling” effect in the sense that the baseline is identical to the negative information condition. They suggested that “By eliminating artificial floors and ceilings on choices of sarcastic readings, revised stories could provide a less ambiguous demonstration of the basic result.”

I argue that, in fact, the opposite is true. If anything, having a ceiling goes against the phenomenon. It might be difficult to undo a strong bias such as a ceiling, because to get an effect, the positive condition needs to do all the “work” of influencing the participants’ response. In contrast, if there is no inherent bias, it might be sufficient if each of the two conditions “pushes” the effect only slightly in its direction, so that their additive effect might “artificially” look like a large difference. So one need not worry about the existing ceiling effect of the baseline; it is an advantage, not a liability, given that the negative and positive conditions did differ from each other.

There is no correct answer, honestly. One reason to have norming could be to establish what is a “correct” answer. One way to conceive of a correct answer is to assume a situation when the reader has the exact information as does June, the addressee. So if the readers and June do not know what happened in the restaurant, the readers can simply use the information they have about the situation and the utterance and assume that June would perceive the same intention they perceive. This is the closest thing I can imagine to a correct answer, but it is unclear that estimating this answer buys us anything. A much better approach is to tell participants there is no correct answer and to mean it. The logic is, whatever correct answer there is, it should be the same in both conditions.

With this in mind, it is unconvincing that Gerrig et al. (this issue) concluded that to avoid the illusion of transparency, participants would always have had to answer that June would not perceive sarcasm: “To fail to demonstrate what Keysar would count as illusory transparency, a participant would have had to have responded ‘no’ to this question on each occasion it was asked, for every story.” This does not hold in two ways. First, to avoid the illusion participants simply have to provide the same answer in the two conditions: either “yes” or “no.” Second, there is no reason to assume that participants who truly take June’s perspective would answer that she would not perceive sarcasm. If anything, the opposite might be true. When readers do not know what happened in the restaurant, they are closer to June’s perspective, and they tend to think that she will perceive sarcasm. Gerrig et al.’s second experiment provides just such a baseline condition. When participants are as uninformed about the event outcome as June
is, they tend to attribute the perception of sarcasm to her. Thus, according to this baseline, the answer closest to June’s perspective should be more often “yes, she’ll perceive sarcasm” than “no.”

Methodological conclusion. It is very important in experimentation to separate the wheat from the chaff. The question of baseline is important. Some experiments need a baseline, but those reported in Keysar (1994) do not. Gerrig et al.’s (this issue) second experiment discovers the baseline and provides a thorough study of the variation among experimental items. Unfortunately, this does not add relevant, incisive information.

THEORETICAL ISSUES: WHAT DOES THE PHENOMENON MEAN?

A phenomenon is a useful vehicle for theory evaluation. In our efforts to understand what gives rise to intriguing phenomena, we could learn about the processes that underlie normal language use. I argue that the illusory transparency of intention is such a phenomenon.

Providing Explanation

If theories cannot explain an established phenomenon, they need to be revised or abandoned. The standard theory would be much more comfortable if the illusory transparency of intention did not exist. Suppose participants had provided the same answer in the two original conditions. Regardless of what they believed Mark’s intention to be, they would have attributed to June the perception of the same intention.\(^1\) This could be directly predicted from the assumption that language users rely on common ground. Had this prediction been correct, it would have fit right in with the standard theory’s basic assumptions.

Nonetheless, the standard theory can still explain the phenomenon. My 1994 article describes how the standard theory could account for the phenomenon and also provides a simpler alternative explanation for it. Here is a summary of these two possible explanations, starting with my preferred alternative and followed by the standard account.

---

\(^1\)One way to design this experiment might be the following: Imagine an experiment that has the same two conditions as the original Keysar (1994) experiments, with two additional conditions. In those conditions Mark either loves or hates the restaurant, but he also knows that June is informed about it. Suppose that when readers take June’s perspective, there is no difference between the two conditions when June is not informed, but there is a difference when she is informed. Such a result would be derived from the assumption of the standard model that readers rely on common ground among protagonists when they take their linguistic perspective.
**Construal.** We know that once people interpret ambiguous stimuli they tend to “construe” the stimuli as less ambiguous; consequently, they believe that others would interpret those stimuli just like they do (e.g., Griffin, Dunning, & Ross, 1990; Griffin & Ross, 1991; Keysar & Bly, 1995; Ross, 1990). Similarly, in the sarcasm experiments, once readers know what Mark intended, they might perceive his utterance as less ambiguous than it really is. Therefore, it is reasonable that they believe that June would be able to perceive the actual intention.

Participants’ reports are in line with this explanation. When asked to explain their responses, participants often made reference to the wording of Mark’s ambiguous utterance. For example, Mark described the restaurant as “marvelous, just marvelous.” Participants who thought that June would perceive sarcasm often explained that the phrase seems sarcastic, especially the repetition of *marvelous*. Participants who thought she would think he was being sincere explained it the same way: The phrase seems sincere, especially the repetition of *marvelous*. Once they interpreted Mark’s utterance, they found “objective” cues in the message; June should be able to perceive the same cues and thus detect his intention. In a later article I suggested that such construal plus insufficient adjustment to June’s perspective can account for the phenomenon (Keysar & Baldwin, 1999).

**Naive theories.** The second possibility I suggested was that perhaps people internalize Gricean principles and thus believe that speakers cooperate in conversation. I suggested that readers might believe that “If speakers cooperate then they should create utterances that are comprehensible by their specific addressees.” Therefore, readers can conclude that June would understand Mark, because he must have known that she had the tools to understand his particular utterance.

Gerrig et al. (this issue) argued in favor of this explanation and fleshed out the reasoning further. Consider this reasoning in the context of a scenario in which the standard theory’s prediction is easier to understand than in the restaurant story. David’s friend Nick asks if he should take Professor Jones’s class; David, who didn’t like the professor (or did like him in the sincere condition) tells his friend that professor Jones is a “real nice guy.” According to Gerrig et al., when the readers attempt to take the addressee’s perspective, they exhibit the following reasoning:

1. When David didn’t like the professor, his sarcastic utterance was not cooperative because it was ambiguous. The friend might be misled and take the class.
2. David seems to violate Grice’s (1975) cooperative principle.
3. Readers would go to great lengths to preserve the assumption that David is cooperative.
4. To do that, they imagine that David must have known that the friend would have the critical information that would enable him to identify his intention somehow.
5. If David believes the friend has the critical information, then the friend probably has that information.

6. Thus, readers attribute to the friend the perception of sarcasm.

This is a full-blown description of the possibility that I described above under “naive theories.” It is in accord with the standard theory. However, there is no direct evidence for this account. First, there is very little experimental evidence that language users actually follow such convoluted reasoning. Although it is assumed as correct by the standard theory, there is little direct and systematic evidence to support this assumption. Second, specifically for the phenomenon at issue, there is no reason to believe that readers go through these reasoning steps. They might simply be doing what I described above, construing the utterances and perceiving them as less ambiguous.

Consequently, the phenomenon has at least two possible explanations, but the data do not allow us to distinguish among them. To ameliorate the situation, Gerrig et al. (this issue) reported a third experiment, which is supposed to test the explanation of the standard theory. I explain next why this experiment does not do the job. It leaves us exactly where we were in 1994: with one phenomenon, two competing explanations, and no critical data leading us to prefer one over the other.

**Testing Theories**

I argue that readers exhibit the phenomenon not because they reason à la Grice, but because they simply construe the utterance with the actual intention in mind and then conclude that it is less ambiguous. But neither my personal belief nor consensus among many others is relevant here. The same requirement should hold for anyone who advocates the standard view. It is not sufficient that most people who work in the field assume Gricean-type explanations. The question is, what experiment can provide data to tease apart these two explanations?

Gerrig et al. (this issue) reported a very clever experiment that is supposed to test their preferred explanation. They added a condition in which Mark hated the restaurant but attempts to conceal this from June, perhaps to spare her feelings. As it turns out, when he tries to conceal his intention, participants no longer think that June would perceive sarcasm. The results are the same as in the sincere condition when he liked the restaurant.

I had liked this idea so much that several years ago I conducted the same experiment (reported in Keysar, 1998, pp. 186–188). The results were also the same as Gerrig et al.’s (this issue) results. Again, the good news is that independent laboratories get the same results with slightly different measures. The question is, what do these results mean?

Gerrig et al. (this issue) told us that the experiment tests the hypothesis that readers attempt to preserve the assumption of Mark’s cooperativeness. In the sarcastic condition, Mark is considered uncooperative because his ambiguous utterance
might mislead June. When readers understood why he was uttering a falsehood, when they understood why “the character wished not to impart negative opinions,” they no longer had to go through the Gricean chain of reasoning. Thus, they did not attribute the perception of sarcasm to June when Mark did not want her to know how much he had suffered. Gerrig et al. were correct in arguing that the concealment experiment can be explained by the standard theory.

The concealment experiment is a perfect example of a clever experiment that cannot distinguish between the two alternative explanations. The construal account naturally predicts the result: Readers construe the utterance in terms of Mark’s communicative intention or the intention he wants June to perceive. If his intention is sarcastic, they see the utterance as overly sarcastic. If his intention is sincere, or he had wants her to think that he is sincere, they see the utterance as conveying that sincerity. This prediction is perfectly realized in the Gerrig et al. (this issue) third experiment as well as the concealment experiment I reported earlier (Keysar, 1998). However, the experiment does not provide data that distinguish between the two alternatives, so there is no reason to favor one explanation over the other.

Indeed, the concealment experiments demonstrate the robustness of the illusory transparency of intention, and Gerrig et al. (this issue) are wrong to conclude that in this experiment “the illusory transparency of intention effect goes away.” The concealment experiments allow us to rule out a third, more boring, explanation for the phenomenon. Perhaps readers simply confused what they know with what June knows. Perhaps they did not encode the information about the restaurant as inaccessible to June. If June knows what happened in the restaurant, it is not surprising that she understood what Mark actually intended. Yet the concealment experiments tell us that readers do not take the factual knowledge about the event as transparent to June. If June knows what happened in the restaurant, it is not surprising that she understood what Mark actually intended. Yet the concealment experiments tell us that readers do not take the factual knowledge about the event as transparent to June. Instead, they take what Grice called m-intention as transparent; the intention that the speaker intended the addressee to perceive as intended (for further discussion, see Keysar, 1998, pp. 186–188). Regardless of what happened at the restaurant, readers think that June will perceive whatever intention Mark wants her to perceive. Indeed, they take the communicative intention to be transparent. Hence, the experiment demonstrates the existence of the illusory transparency of intention.

Paradigmatic Considerations: The Usefulness of Illusions and Other Phenomena

Gerrig et al. (this issue) expressed concern about the adequacy of the approach taken in the 1994 article as well as the appropriateness of calling the phenomenon an illusion. I do not share their concerns. In my opinion, the approach is quite fruitful. I explain how the search for similar phenomena could benefit, not hinder, the field of pragmatics.

Having illusions is not a problem for a field of research. The field of perception has benefited immensely from the discovery of a variety of visual and auditory illusions. The field of judgment and decision making has also benefited greatly from
the discovery of surprising phenomena. They pose a challenge to theories that must be able to explain how normal processing would give rise to these illusions. The best thing that can happen to the field of pragmatics is the discovery of analogous illusions and surprising phenomena that need to be explained. This would provide researchers with a powerful tool to evaluate theories that currently suffer from lack of nontrivial experimental support.2 Illusions need not be shunned, but should be welcomed.

Given substantive differences between the area of visual perception and language use, it is unlikely to expect the emergence of true analogs of perceptual illusions in pragmatics. One could debate whether the phenomenon I described is truly illusory, but that would be quite futile. The real challenge is to identify a function that phenomena such as this can have in the study of language use. In my opinion, such phenomena (illusory or not) can play a crucial role that is analogous to the one played by perceptual illusions. They can serve as tools to test competing theories.

Consider the illusory transparency of intention. It has already forced the standard theory to jump through hoops in order to explain it: The complex steps of reasoning offered as an explanation by the standard theory is quite a stunt. Once a theory provides an account for such a phenomenon, we can proceed and test this account further using the phenomenon as a vehicle. Gerrig et al. (this issue) attempted to do just that with their third experiment; they manipulated the phenomenon with the hope of testing the explanation the standard theory provides. Unfortunately, in this particular case their experiment does not have the capacity to rule out alternative explanations, but Gerrig et al.’s attempt is precisely in the spirit of what I am proposing. Clearly, then, the approach my 1994 article advocated has already proven useful.

CONCLUSIONS

When I wrote my 1994 article, I thought my methodological assumptions were obvious. I did not explain them at length originally. I avoided spelling out the irrelevance of a baseline because that seemed obvious, and saying the obvious would have been uncooperative because it violates Grice’s informativeness maxim. Clearly, I was proven wrong. I hope that I have corrected this here. More important, I hope my general suggestion is convincing: It would be most beneficial for experimenters to avoid irrelevant baseline conditions.

As for the theoretical and paradigmatic considerations, I believe that this debate has taken us one step further in some respects but left one foot standing exactly where we were in 1994. We learned a specific lesson about the illusory transparen-

---

2Much of the published experimental support for the standard theory of pragmatics can be easily explained in different, simpler terms, leaving the theory without much of an experimental basis (Keysar, 1997).
cy of intention from my original 1998 concealment experiment and its replication by Gerrig et al. The phenomenon is even more compelling than the original 1994 experiments have shown it to be. But we are not closer to identifying a theory that can explain it better than others can. On an optimistic note, I suggest that the phenomenon could be a useful vehicle. The more we have such nontrivial phenomena, the better we will be able to develop, test, and revise theories.

ACKNOWLEDGMENTS

This article was written with the support of Public Health Service Grant R29 MH49685 to Boaz Keysar. I thank Kate Baldwin, Dale Barr, Linda Ginzel, and Anne Henly for comments.

REFERENCES


