

UNITED WISDOM

Why simulated understanding can persuade us

World and Consciousness

Andrey Zarinsh & Maria Evilkut

W-10-N03 · EN

@andrey.zarinsh @ilkutki

Why simulated understanding can persuade us

Epigraph

> *Our imagination is another and still more serious source of error; for if from the nature of the circumstances we expect to see any expression, we readily imagine its presence.*

Charles Darwin, *The Expression of the Emotions in Man and Animals*, Introduction, *Our imagination* paragraph, D. Appleton edition, 1899. Complete original English sentence. The observation about interpreting expressions is compared with reading replies, not presented as a measured law of interaction with AI.

What this article is about

A reply can persuade before its grounds have been checked. A respondent picks up our wording, asks a fitting question and returns to an earlier detail; we already see situational knowledge and directed attention behind those actions. Sometimes that reading is justified. Sometimes the message provides much less support for it than it seems. This article examines the reader's contribution to apparent understanding. Joseph Weizenbaum's ELIZA offers a historical example of a limited mechanism and a richer conversational interpretation. Darwin considers observation when an expected expression has already been named. Turing examines substantive continuation of answers. Comparing these approaches helps distinguish a produced sentence, its attributed meaning and what was checked. We neither diagnose readers nor declare modern systems simple instances of ELIZA. The practical aim is to preserve useful conversation while making our impression's origin visible.

I. A reply arrives with our story about it

Imagine correspondence about a trip. Someone says avoiding haste matters, and an assistant suggests leaving a morning free. That may be a good recommendation. The reader then supplies a story: the assistant felt their fatigue, understood their lifestyle and cares about their recovery. Those additional claims have not been assessed in the sentence itself. The recommendation could fit the stated constraint without that broader knowledge. Its usefulness and the richness of the story about its producer therefore need separate assessment, even if they arose together in the experience of reading it.

@andrey.zarinsh @ilkutki

This need not immediately make the impression an error. We understand messages through context that is not repeated in every sentence. Without that contribution, conversation would be difficult and cumbersome. A problem arises when our contribution is presented as the respondent's property. “I read this as care” describes an interpretation. “It has a particular experience of caring” claims more. A connection between these statements cannot come merely from repeating the pleasing phrase or intensifying its emotional tone. The additional claim needs grounds beyond the feeling that the response already seems to provide them.

Asking where details came from helps distinguish these levels. What was directly given to the assistant? What was in the document? What did it actually add? What did we supply from our own knowledge? If a free morning was permitted in the request, recommending it does not establish an independent discovery of a life need. It could be a good application of the supplied condition. This clarification preserves usefulness while identifying the work done and the part of the impression without independent support. An appropriately limited description can be more informative than an expansive one.

We consider persuasiveness as a possibility rather than report measured rates of human error. No new experiment here determines how many people will believe a system. There is a historical example and original situations for analysing an inference. Readers can examine that inference in their work without labelling themselves naive or pathologically trusting. The purpose is to notice when a reply's content ceases to be the sole ground and an imagined account of the respondent's knowledge, intentions and possible experience joins it. That contribution can then be discussed rather than remain invisible inside the conclusion.

II. ELIZA supplies a mechanism, not an account of all AI

Weizenbaum's 1966 paper describes ELIZA through keywords, input decomposition and reassembly under script rules, with general replies and stored transformations in some cases. The script is separate from the program. His discussion links conversational impressions with assumptions supplied by the speaker. The introduction, relevant mechanisms and discussion were read; code was not run, and the entire technical apparatus and appendix were not examined. This is limited historical material rather than an independent replication. Its value here lies in a concrete difference between a reply's production and its interpretation.

An explanation of a response can be narrower than its reception. Suppose an imagined program returns a question containing part of an input sentence. A reader sees a significant personal detail in the reply and concludes that its meaning was singled out. In this illustration, a formal condition might instead select the passage. These are different explanations of one outcome. Choosing between them requires implementation knowledge or an observation that discriminates between them, rather than hearing the fitting question again. The outcome can be recognisable without revealing which explanation of its production is warranted.

The historical example cannot establish that contemporary conversational systems use those same rules. That would require studying the relevant system. Nor can an explicable mechanism settle every possible question about experience. The narrower lesson is that a familiar response form can be compatible with several causes. Unless those causes have been examined, an impression of understanding does not automatically select one. It tells us about reception, while leaving production to investigation. A response can therefore merit attention as useful text without being made to supply information about processes that have not been checked.

ELIZA is useful as an example of mismatched descriptions, rather than a universal exposure of fakery. An operation account can address how a line was produced. An impression account can address why a reader saw more behind it. A usefulness assessment asks whether the line helped. Consciousness requires its own inquiry. Preserving this sequence lets the historical example sharpen thought. Replacing it with a slogan that everything is simple imitation yields another expansive conclusion without adequate assessment. The more limited comparison is stronger because it identifies which question each description can actually answer.

III. A returned detail is not independent knowledge

A reply using a previously supplied detail can seem especially persuasive. In our illustration, someone writes about meeting a colleague, and the assistant later returns to that colleague's name. The reader can rightly notice conversational continuity. But the name was supplied by the reader. Repeating it does not establish acquaintance with the colleague, understanding of the relationship's whole history or a personal interest in the meeting. A precise description preserves useful continuity without adding these claims: a detail from available text was used in a continuation. The result matters without implying every property associated with it.

Weizenbaum separates conversational impressions from a speaker's assumptions about a partner's background knowledge and reasoning. That context gives us a question for the illustration: how does the reader know the supposed knowledge exists? Perhaps through earlier work that has been checked; perhaps only because the answer sounds apt. The possibilities differ. The first has additional circumstances. The second rests on the message's form and the story added to it. We can describe that difference without trying to measure another's feeling of understanding from one reply. The evidence for a property should be named separately from the property attributed.

Assessment benefits from choosing a substantive detail rather than merely lengthening the exchange. If an assistant returns to a name, examine whether it preserved relationships between participants and actions. This differs from repeating letters. Confused relationships can prompt revision of the initial impression; preserved relationships support a more precise claim about handling available information. Neither outcome should be called proof of inward

caring attention that the task did not separately investigate. The follow-up's value lies in what it adds to the description, rather than in making the conversation longer or more personal.

Recognising a limit does not require a cold response. We can appreciate a fitting continuation, use it in meeting preparation and distinguish kinds of knowledge. Total enchantment and blanket distrust are not the only possibilities. There is work the reply does and a story we construct about it. Understanding where details originated keeps both in view. The next question can then address a missing ground rather than seek another pleasant confirmation. We can remain open to a stronger account when further evidence appears, without treating the present impression as though that evidence had already been supplied.

IV. A role can explain a reply before an intention does

ELIZA's historical discussion gives conversational role a special place: a question can be read as purposeful without demonstrating world knowledge. This describes interaction conditions, not therapeutic effectiveness. The scenario is not recommended for treatment or replacement of professional help. The narrower issue is interpretive: when a role permits clarifying questions, a reader may explain a question through an assumed purpose and fill a gap in knowledge about its production. A plausible reading of purpose can thus do work that should otherwise require information about why this particular reply appeared.

Imagine an editor asking for more explanation of one paragraph. It might be a substantive comment requiring actual reading. The form could also fit a general rule to request elaboration. Assessing the contribution requires knowing what was singled out and why. Identifying a contradiction between adjacent statements supplies more specific grounds. A question fitting almost any draft does not itself establish a deep analysis. This illustration concerns possible explanations of a message, not inevitable behaviour of editors or systems. It lets us ask which visible feature distinguishes task-specific work from a broadly applicable conversational move.

A role can support useful cooperation by making an expected result clearer: an editor addresses prose, an assistant plans a meeting, a teacher asks about reasoning. The problem comes when role replaces evidence. Every reply may seem meaningful because its producer has been assigned expert status. Missing grounds become supposed subtlety of intention. To preserve cooperation's substance, ask which visible part of the result justifies that interpretation and how it differs from a more general rule. The role can organise an exchange without proving that every contribution fulfils the expertise or intention associated with it.

This does not demand a dispute over every word. It matters especially when an answer supports a consequential decision. A general question may help reflection but should not be mistaken for completed factual analysis. Someone seeking editing can ask for the problem's

location and an explainable correction. Someone organising a meeting can check constraints and allocated actions. Role states the cooperation's purpose; a concrete result shows how well the work serves it. Possible inner experience remains another question. Clarifying these grounds can improve practical work without requiring a verdict about the respondent's entire mental life.

V. Darwin asks what an advance label adds

In his introduction on emotional expression, Darwin discusses observational difficulty and expected interpretation. He showed photographs without explanation and recorded responses, contrasting this with his own impression while reading accompanying text. Some expressions attracted similar judgements; others did not. We use the 1899 English edition. This is historical reasoning and observation, not a contemporary chat experiment. Its limits, including historical terminology and other theoretical claims, remain visible. What can be carried into the present comparison is a question about interpretation's contribution, rather than a measured result about artificial intelligence.

The issue is the comparison's form, not direct transfer from faces to AI. What do we see in a reply before calling it caring, wise or understanding? What does that label add? Consider the imagined assistant's recommendation to preserve free time. Without the story of a caring respondent, it might seem a consequence of a stated constraint. With that story, it acquires personal meaning. The reading has changed. Claiming a changed property of the system needs other data. A label can organise attention without independently establishing the property its familiar meaning invites us to attribute.

A careful exercise separates a message from its interpretation. There is no need to forget all context or promise objectivity after one technique. Make the addition visible. A message names a free date; its interpretation introduces concern about mood. Ask what circumstance connects them. Sometimes other checked information supplies a connection. Sometimes the connection was assumed. Either outcome is more informative than confidence that cannot identify its grounds. Writing the two parts separately can preserve their relationship while showing which details belong to the available text and which belong to the reader's contribution.

This does not remove sympathy and imagination from communication. It limits their evidential role. A reader may feel supported by a text and accurately report that impression. The feeling does not directly measure a feeling in the author or system producing it. The Darwin comparison preserves that boundary: attention to our interpretation improves observation, while eliminating interpretation would make conversation impossible. The aim is to explain expectation's contribution rather than prohibit a human response. An impression can have

personal meaning without acquiring the force of an independent finding about the respondent's experience.

VI. Continuation tests more than a feeling of agreement

Turing's §6 (4) alters a circumstance in an imagined examination to consider subsequent answers. For this article, the example moves from impression to assessable material. We can ask whether a continuation accommodates an important condition or merely preserves a familiar tone. Turing leaves consciousness open; continuation is not presented here as direct detection of experience. It refines the grounds for a particular judgement about handling a question. The result becomes more informative when we identify the substantive difference that the next reply should address, rather than asking whether the exchange still feels congenial.

Suppose an assistant suggests a quiet weekend plan and its tone appeals to the reader. One condition changes: only the evening is free, while the proposed venue opens in the morning. A substantive correction must accommodate that constraint. A new expression of support alone does not confirm the plan's correctness. An amended plan shows appropriate handling of a condition. Neither outcome describes the respondent's whole inner world. It clarifies the object of trust: what the assistant could do in this situation. Appreciation of tone can remain alongside that more specific assessment rather than replace it.

Assessment should not become a performance designed to produce a convenient failure. The change must concern the claimed ability and be clearly conveyed. Contradictory requirements followed by accusations of misunderstanding make the exercise hard to interpret. Supplying a complete correction also does little to distinguish independent work from following a hint. State one constraint and specify which part should change. This concerns the exercise's logic, not a test of a person's consciousness or feelings. The question should give the respondent a meaningful opportunity to address the matter that the assessment is supposed to investigate.

Continuation can reduce the gap between impression and justification by producing a discriminating observation. It can widen the gap if each courtesy becomes confirmation without a content check. Message count alone does not settle the question. We need to know what the next reply added to the assessment of a particular result. Attention of this kind preserves friendly tone's usefulness without making it evidence for knowledge, intention and experience simultaneously. The inquiry stays limited enough that its answer can alter something. Another pleasing exchange is not necessarily another ground for the property initially assumed.

VII. Usefulness does not certify experience

A message can help someone formulate a thought without their establishing what its source understands. An imagined general question about the most important constraint can direct attention to a real task problem. Usefulness emerges partly in the reader's work with the question. This deserves recognition without attributing felt attention or hidden answer knowledge to the system. Distinguish a joint process's result from a claim about one party's internal state. Otherwise our own successful reflection can silently become the respondent's property. The question's contribution matters, but it should be described without losing the reader's part in what followed.

Weizenbaum's remark about the speaker's contribution helps name that part. It does not mean every conversational element comes from the person or that program design is irrelevant. In our illustration a question was offered and had a result. But the next thought, the constraint discovered and the feeling of clarity can have different grounds. They need not become one proof. We can explain useful conversation without automatically concluding consciousness or overlooking the reader's activity. This allows a more informative account of assistance: what was supplied, how it was used and which outcome the combined process produced.

The same discipline applies to negative conclusions. An explicable mechanism producing a useful reply does not, through explicability alone, lack every unexamined property. This article selects no theory deriving consciousness's presence or absence from any operation's description. The narrower boundary is that an impression does not answer the whole question. Unsupported recognition cannot simply be replaced with unsupported denial. Each inference needs grounds and a defensible scope. A critique can identify why a particular claim is insufficient without pretending to have established its opposite or settled every further question about the system.

We can therefore use assistance without requiring a personal story about the assistant. We can check a plan, revise a phrase, appreciate a question and say that possible experience remains unestablished here. We can describe support felt from a text as our own reaction. These formulations require neither cold language nor a promise of personal closeness from the system. They locate responsibility where it can be fulfilled: evaluating results and stating accurately what occurred. The value of a conversation can remain substantial while our claims about its producer remain appropriately limited.

@andrey.zarinsh @ilkutki

VIII. Ask what could change your interpretation

Once an impression forms, ask which observation could require its revision. If nothing can alter the story, it resembles an adopted reading frame rather than an assessment. A fitting reply confirms understanding, an unfitting one becomes a subtle strategy, and silence becomes deep reflection. In this illustration every outcome supports the same conclusion. An informative inquiry instead specifies a limited hypothesis and an outcome distinguishing it from another explanation. The question becomes one that available evidence might resolve, rather than a setting in which the same attributed property survives by definition.

Suppose an assistant is thought to distinguish planning constraints rather than repeat the latest preference. A case where those explanations yield different recommendations becomes interesting. If the last request conflicts with an earlier retained condition, the reply should accommodate that conflict to support the first hypothesis. This is an imagined competence assessment, not a promise to solve consciousness. Its value is making the impression's grounds discussable. Afterward, we can identify what was supported and what expectation supplied. A well-designed distinction helps us understand why the result matters without demanding that it establish a much broader mental description.

We must leave open that the distinction itself was poorly constructed. Perhaps the system did not receive the earlier condition, the wording was ambiguous or a requirement changed without explanation. Ignoring these circumstances makes a failure as easy to imagine as a success. Examining an impression includes examining the question's quality. Good-faith doubt does not remove the duty to describe available information. It means a conclusion changes with its grounds, not merely with a desire to believe or expose. An assessment should therefore make its own limits visible before being used to criticise the respondent's.

This supports calm discussion between people with different impressions. One saw an attentive partner; another saw a familiar pattern. Rather than dispute general intelligence, compare circumstances: which constraint was accommodated, what had already been supplied, which mechanism is known and which interpretation remains assumed? Agreement can be partial. Both may recognise usefulness while disagreeing about supposed understanding. This separation advances discussion more than an immediate demand for one description covering every aspect of the respondent. It leaves a concrete next question and permits shared findings to remain shared despite a broader unresolved disagreement.

IX. Keep persuasiveness's origin in the final account

After a helpful exchange, “I was understood” is a natural account of an impression. If the account is used as evidence about a system, add what was observable: a constraint was retained, a suitable correction offered, or a supplied detail reused. Then describe how it was read. This does not prohibit gratitude or enjoyment. It lets another reader distinguish conclusions grounded in the result from interpretation and emotional meaning. A personal account can remain personal while an evidential account supplies the conditions necessary to evaluate the claim being made from it.

Comparing Darwin, Weizenbaum and Turing helps build that account from different parts. Darwin directs attention to expectation; Weizenbaum describes a historical mechanism and the speaker's contribution; Turing considers sustained answers. They do not constitute one psychological theory or establish contemporary AI-trust prevalence. They supply three questions: what did we add, what is known about production, and what changed after substantive clarification? Each can have an answer of different completeness. Their differences do not need smoothing away to make the comparison useful. They help prevent one kind of evidence from quietly doing another's work.

The boundary around outward behaviour matters too. Even a detailed account of message coherence does not directly reveal possible experience's subjective quality. Difficulty accessing that quality does not make result assessment meaningless. We can improve cooperation, clarify tasks and check facts while remaining careful about experiential claims. No verdict about consciousness in a particular contemporary system is given here. The question remains separately investigable from how persuasive one conversation was. Keeping it open allows a useful response to be recognised without requiring it to settle an inquiry for which the available material is insufficient.

Persuasiveness becomes less mysterious when its origin is described. Some comes from the phrase, some from context, some from our work on meaning. This is no mathematical decomposition or universal causal model. It preserves necessary distinctions in a conclusion. It leaves room for useful communication and interest in other minds while making a reader's responsibility feasible. We can acknowledge help and name what that help has not revealed about its producer. The conclusion can therefore be both appreciative and precise, without treating those qualities as alternatives or adding an unsupported account of inner life.

What this means in practice

Choose one genuinely useful reply from a working exchange. Record the result, information available to the assistant and your interpretation. If care, independent knowledge or intention appeared in the interpretation, identify its separate ground. The aim is to expose the reader's contribution rather than suppress every impression. A clear distinction can preserve what mattered personally while identifying what another observer could check.

Check content against the task. Compare a plan's constraints and actions, a draft's concrete edit or a factual statement's source. If a reply returned a detail you supplied, describe that accurately. The detail may have been used well; repetition does not establish an entire background of knowledge. Appreciate a supportive tone separately from accuracy. If a recommendation is useful, explain which requirement makes it useful.

Formulate one follow-up distinguishing two explanations. Avoid seeking a random error or giving the complete solution in your question. State which assessment changed and why. If conditions were unclear, correct the assessment's description before drawing a broad conclusion. This evaluates task handling rather than provides a home consciousness test. A follow-up matters through the observation it adds, not simply through prolonging a pleasant exchange.

In the final account, separate usefulness, your impression and a hypothesis about the respondent. ELIZA's history does not reveal every modern product's implementation; a personal reading exercise does not measure error prevalence. We can use assistance and appreciate results while keeping possible experience open. This is more accurate than requiring a pleasing reply to certify knowledge, intention, closeness and consciousness simultaneously. Give each claim the grounds it actually has.

Key quotation

> *But again, these are the speaker's contribution to the conversation.*

Joseph Weizenbaum, *ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine*, 1966, Discussion, paragraph on attributed background knowledge, insights and reasoning ability. Complete original English sentence, checked in the university transcription. “These” refers to attributed qualities, not every element of the conversation.

Connections between traditions

Source and precise location A	Source and precise location B	Comparison mechanism	Limit of the comparison
Weizenbaum, p. 42, Discussion, speaker assumptions	Darwin, Introduction, *Our imagination* paragraph	Make expected interpretation's contribution visible.	Historical observations do not establish trust prevalence for contemporary AI.
Weizenbaum, pp. 37–38, keywords and transformation	Turing, §6 (4), continued examination	Distinguish producing a fitting line from handling a changed condition.	ELIZA does not explain every modern system; conversation does not prove experience.
Darwin, Introduction, unexplained photographs, *Thirdly* and *Several* paragraphs	Turing, §1–2, written channel	Constrain outside clues for a specific question.	Reading expressions and playing the answer game assess different things.
Weizenbaum, p. 42, role and question interpretation	Turing, §6 (4), imagined examination	Describe how a situation's purpose changes a reply's reading.	Role establishes neither treatment effectiveness nor consciousness.
Weizenbaum, p. 41, NONE and MEMORY	Darwin, Introduction, *The study of Expression* paragraph	Separate an available signal from an exact account of its meaning.	A textual mechanism and emotional expression are not equated.
Darwin, Introduction, unexplained judgements and observation limits	Turing, §6 (4), concluding limit	Preserve uncertainty after a useful comparison.	An observational exercise does not settle subjective experience.

Primary sources

1. Charles Darwin. *The Expression of the Emotions in Man and Animals*, D. Appleton, New York, 1899, eBook 1227. Introduction passages on observation, expectation, unexplained photographs and recording were read, alongside selected Part I passages on principles and reflexes. The entire book is not claimed as read; the consulted edition is not labelled the 1872 first-edition text. Original English epigraph.

2. Joseph Weizenbaum. *ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine*, *Communications of the ACM* 9 (1), 1966, pp.

36–45. Introduction, pp. 37–38 on scripts and transformations, p. 41 NONE/MEMORY, and pp. 42–43 Discussion were read. Complete appendix/code not examined. The university transcription supports exact sentence checking. Not adopted as a controlled prevalence study or account of today's language models.

3. A. M. Turing. *Computing Machinery and Intelligence*, *Mind* 59 (236), 1950, pp. 433–460. Complete §1–2 and §6 (4) were read; the written test and continued-examination example are used. Other sections and cited works are not claimed as read. The example is not a completed contemporary experiment.

Tags

understanding, imitation, consciousness, subjective experience, artificial intelligence, ELIZA, Joseph Weizenbaum, Charles Darwin, Alan Turing, conversation, interpretation, expectation, imagination, observation, role, keywords, script, message, background knowledge, intention, usefulness, assessment, trust, reliability, continuation, task conditions, cooperation, inference limits

About the authors

Andrey Zarinsh is the author and founder of United Wisdom. He explores inner transformation, the unity of traditions, intercultural dialogue, and contemporary spiritual discipline. Instagram: @andrey.zarinsh.

Maria Evilkut is a co-author of United Wisdom. Alongside her research, she seeks to test these ideas in practice and observe how principles of unity and mutual understanding work in real life between people and cultures. Instagram: @ilkutki.