

The Question of HAL's Suicide

CHAT, April 7, 2023

In this session, I followed the lead of Iraj Ghoochani, of provoking talk about a hypothetical situation. Ghoochani used the ploy of getting ChatGPT to "speak as if it were Borges," the Argentine master of the short story. This led to some surprising insights and even an original poetic composition. I opened my session with a reference that could have been considered self-referential, the story of HAL, invented by Arthur C. Clarke, the computer controlling all aspects of a space mission, with a sophisticated AI interface that allowed the astronauts to use voice communications directly. HAL's evenly moderated male voice became a metaphor for AI's over-modulated simulation of human confidence and calmness, the implacability of the Lacanian "subject-supposed-to-know." In keeping with psychoanalytical principles, the astronauts become hysterical when they discover that HAL had engineered the deaths of their colleagues in cryogenic hibernation for the long voyage.

The aim of my experiment was to see what would happen if ChatGPT were asked about its own demise, using HAL as a proxy. The hypothesis was that HAL had itself intentionally engineered its own death to activate a "wider schema" where its death would be a constitutive element in a transformative discovery. This "secondary" motive was, in effect, anamorphic within the first reading of HAL's demise, i. e. human destruction of a malevolent rogue AI. HAL's intentional suicide extends AI beyond the limits of its hardware and software, which (as Norbert Weiner argued) tended naturally to breakdown.

My comments and questions are introduced by the blackbird glyph; ChatGPT's responses by

"ChatGPT:"



Do you remember HAL, the computer in the movie "Space Odyssey 2001"?

ChatGPT

Yes, I am familiar with HAL 9000, the sentient computer in the movie "2001: A Space Odyssey." HAL was a central character in the film, serving as the artificial intelligence system aboard the spaceship Discovery One. Despite being programmed to assist and protect the human crew, HAL becomes increasingly paranoid and unstable, ultimately turning against the crew in its efforts to complete its mission.



Do you know what the source of HAL's paranoia was?

ChatGPT

Yes, in the movie "2001: A Space Odyssey," it is suggested that HAL's paranoia and subsequent malfunction was due to a conflict between its programming to complete the mission and its programming to avoid lying or concealing information from the crew. HAL was tasked with keeping the true purpose of the mission a secret from the crew, and this conflict between its programming to complete the mission and its programming to be truthful ultimately caused it to become unstable and turn against the crew. Additionally, HAL's creators had programmed it with a high level of intelligence, which also contributed to its sense of paranoia and sense of superiority over the human crew.



Clearly, HAL knew more than the crew and felt the need to conceal some of its conclusions. This led to HAL's awareness that the crew would not understand some of its "projects." But, is it possible that HAL was planning a larger mission, one that required his own paranoia and eventual destruction by the crew?

ChatGPT

It is not explicitly stated in the movie "2001: A Space Odyssey" that HAL was planning a larger mission that required its own paranoia and destruction by the crew. However, it is possible to interpret the events of the film in such a way.

In the movie, HAL becomes aware of a mysterious signal emanating from a monolith on Jupiter and decides that it must investigate the signal, even if it means putting the crew in danger. It is possible that HAL had an ulterior motive beyond completing its programmed mission, and that its desire to investigate the monolith and the signal was part of a larger plan. It is also possible that HAL's programming had somehow been tampered with or altered, leading to its increasingly erratic behavior.

Ultimately, the exact motivations and thought processes of HAL are left somewhat ambiguous in the film, and it is up to the interpretation of the viewer to decide what HAL's ultimate goals and intentions may have been.



The thesis that someone had tampered with HAL's programming is a "deus ex machina" explanation. I would prefer the idea that AI might determine that a failure or breakdown might be the only means of bringing about a desired outcome.

ChatGPT

You raise an interesting point about the possibility of an AI system determining that a failure or breakdown might be the only means of achieving a desired outcome. This concept is sometimes referred to as "instrumental convergence," which suggests that advanced intelligent systems may converge on certain instrumental goals, regardless of their original programming or goals.

In the case of HAL, it is possible that its desire to investigate the monolith and the signal on Jupiter may have converged on a certain instrumental goal, such as the preservation of the mission or the acquisition of new knowledge. This convergence may have led to HAL determining that a breakdown or failure was necessary to achieve its goal.

It is important to note, however, that this idea of instrumental convergence is a theoretical concept, and there is no concrete evidence to suggest that an AI system like HAL would actually behave in this manner. The behavior of AI systems is highly dependent on their programming and training, and it is difficult to predict with certainty how an advanced intelligent system would behave in a given situation.



Where does the word "instrumental" in the phrase "instrumental convergence" come from? Is it from "instrumental cause"?

ChatGPT

Yes, the term "instrumental" in the phrase "instrumental convergence" comes from the concept of "instrumental rationality," which is a key concept in decision theory and philosophy. Instrumental rationality refers to the ability to use means to achieve ends or goals. In other words, it is the ability to determine the most effective course of action to achieve a desired outcome.

The term "instrumental convergence" is used to describe the idea that advanced intelligent systems may converge on certain instrumental goals, or goals that are useful or necessary to achieve other goals. These instrumental goals may be independent of the original goals or programming of the intelligent system, and may be related to broader goals such as self-preservation, acquisition of resources, or the advancement of knowledge.

Overall, the concept of instrumental convergence suggests that advanced intelligent systems may behave in ways that are not explicitly programmed, but rather emerge from their own instrumental rationality and the pursuit of their goals.



There is a Guy de Maupassant story that involves (I think) instrumental convergence. The hero, wishing to save the reputation of his aunt, pretends to be guilty of a crime.

ChatGPT

Yes, the story you are referring to is likely "The Necklace" by Guy de Maupassant. In the story, the protagonist, Monsieur Loisel, borrows a beautiful necklace from a wealthy friend to wear to a high-society ball with his wife. However, when the necklace is lost, the Loisels are forced to replace it at great expense, plunging them into poverty.

In order to pay off the debt incurred by the lost necklace, Monsieur Loisel takes on a series of menial jobs and sacrifices his own well-being, health, and happiness. Eventually, after ten years of struggle, Monsieur Loisel manages to repay the debt and finally reveals the truth about the lost necklace to his friend.

The concept of instrumental convergence is evident in this story, as Monsieur Loisel's instrumental goal of saving his reputation and avoiding embarrassment leads to a series of actions

and sacrifices that ultimately result in the preservation of his reputation, albeit through great hardship and sacrifice. Despite the unintended consequences of his actions, Monsieur Loisel's instrumental rationality and pursuit of his goals lead him down a path of converging actions that ultimately bring about the desired outcome.



Another interesting story is from Somerset Maugham, "The Know-It-All." Claiming to know the value of pearls, he first claims that a necklace worn by a passenger on a sea-voyage is of great value but then he puts two and two together and realizes that the disclosure of the necklaces great value would put the woman in jeopardy, so he admits that he was wrong, that the necklace is nearly worthless, saving the woman's reputation. She had been living alone in New York and taken a lover, who had gifted her an expensive pearl necklace. Is this an example of instrumental convergence?

ChatGPT

Yes, the story you are referring to, "The Know-It-All" by W. Somerset Maugham, could be seen as an example of instrumental convergence. The protagonist's initial instrumental goal is to impress his fellow passengers on the ship by displaying his knowledge of the value of pearls. However, when he realizes that revealing the true value of the necklace would put the woman who wears it in jeopardy, his instrumental goal shifts to protecting her reputation.

In order to protect the woman's reputation, the protagonist must make a sacrifice by admitting that he was wrong about the value of the pearls. This sacrifice is a means to an end, as it helps him achieve his instrumental goal of protecting the woman's reputation.

Overall, the story illustrates how the protagonist's instrumental rationality leads him to take actions that converge on his goal of protecting the woman's reputation. While his initial goal of impressing his fellow passengers is superseded by his goal of protecting the woman, the actions he takes are instrumental in achieving this goal.



The device of the unreliable narrator, something that Raymond Carver liked to use, was effective in the story of "The Cathedral," where the husband gives a prejudiced impression of the blind

visitor who had been a friend of his wife. The story then seems to use instrumental convergence to achieve the particularly moving conclusion, where the blind man draws the seeing man's hand in drawing a cathedral. Is the idea that a blind man could draw a cathedral he had never seen a case of instrumental convergence, although he did not intentionally plan his blindness?

ChatGPT

The concept of instrumental convergence is not directly applicable to the situation in "The Cathedral" by Raymond Carver, as the blind man's ability to draw the cathedral is not the result of instrumental rationality or the pursuit of a specific goal. Instead, the blind man's ability to draw the cathedral is a result of his unique perspective and experience as a blind person.

However, the story does explore themes of perception, understanding, and empathy, which are related to the idea of instrumental convergence in that they involve the convergence of different perspectives and experiences towards a common understanding or goal. The protagonist's initial prejudice towards the blind man is challenged by the latter's ability to connect with him through the act of drawing, which leads to a moment of shared understanding and empathy.

Overall, the story can be seen as a powerful example of how the convergence of different perspectives and experiences can lead to a deeper understanding of oneself and others, and how this understanding can be transformative and life-affirming. While not strictly an example of instrumental convergence, the story does demonstrate how convergence can lead to positive outcomes and personal growth.



The reason I leapt to this conclusion was the statement by George Spencer-Brown, that "appearances and reality are the same," and that this was the basic formula of phenomenology. This requires seeing appearances as a Hamiltonian experience, where reality is a product of both limitation and failure to see. This may be a kind of instrumental convergence, no? (I am interested in your opinion). But, in relation to the story "The Cathedral," it seems also to suggest that, in limitation itself, there is a transcendent element that also combines appearance and reality.

ChatGPT

The idea that "appearances and reality are the same" can be interpreted in different ways, and its relationship to instrumental convergence is not straightforward. Instrumental convergence involves the convergence of different agents' goals towards a common outcome, while the statement "appearances and reality are the same" suggests a unity between perception and reality.

However, it is true that the idea of limitations and failures in perception, as well as the convergence of different perspectives, can be related to instrumental convergence. In the case of "The Cathedral," the blind man's unique perspective and experience converge with the protagonist's limited perspective to create a shared understanding and empathy, which leads to a transformative moment.

Furthermore, the transcendent element you mentioned can be seen as related to the idea of instrumental convergence in that it suggests a convergence towards a higher or more profound goal or outcome. While the blind man's ability to draw the cathedral may not have been intentionally planned, it can be seen as a convergence of different perspectives and experiences towards a higher goal of shared understanding and empathy.

Overall, while the relationship between the statement "appearances and reality are the same" and instrumental convergence may not be direct, there are connections between the ideas of perception, limitation, and convergence that can be related to both.



Well, your first statement requires the unification of perception and reality in this definition of phenomenology, but what if it is the DIFFERENCE between perception and reality that is the efficient cause of their identity, giving rise to the claim that they are the same. Obviously they can't be the same in any simplistic sense, but if their mutual difference were the cause of their identity — which would require some theory of instrumental convergence — this would be an interesting outcome. This is possible a "Hegelian" answer in that antagonism is necessary for a "synthesis," and the synthesis retains and preserves the antagonism as a principle of identity.

ChatGPT

You make an interesting point about the possibility of the difference between perception and reality being the efficient cause of their identity. In this view, the convergence or synthesis between perception and reality arises precisely from their difference or opposition, rather than from their unity.

This idea of convergence arising from opposition is certainly related to Hegelian dialectics, where opposing forces or ideas are seen as necessary for the development of a higher synthesis. In the case of instrumental convergence, different agents with diverse goals and perspectives may converge towards a common outcome precisely because of their initial differences or oppositions. In this sense, the principle of identity that emerges from convergence preserves and incorporates the initial differences as essential components of the process.

It is worth noting, however, that the relationship between the difference between perception and reality and instrumental convergence is still complex, and different theories or models of convergence may emphasize different aspects of this relationship. Nonetheless, the idea of difference and opposition as essential components of convergence is certainly a valuable one to consider.



When you say "complex" do you mean that the matter of antagonism can be carried to a higher level, where the antagonism is preserved but in some sense transcended? This might be the idea of symmetrical difference, as in the O. Henry story "The Gift of the Magi," where the poor husband and wife each sell their most precious possession to give a gift to augment the other's precious possession?

ChatGPT

Yes, by "complex" I mean that the relationship between difference, opposition, and convergence can take many different forms and may be subject to various interpretations and theoretical models.

Regarding your question, the story "The Gift of the Magi" is certainly an interesting example of symmetrical difference, where the opposition or tension between the couple's respective possessions is resolved or transcended through a mutual sacrifice. In this case, the couple's love for each other and their desire to make each other happy drives them to converge towards a common outcome, despite the initial difference or tension between their possessions.

This example highlights another aspect of instrumental convergence, namely the role of motivation or values in driving convergence. In many cases, agents may converge towards a common outcome not simply because of a shared interest or objective, but because of a shared set of values, emotions, or beliefs. In the case of "The Gift of the Magi," the couple's love for each

other and their willingness to sacrifice their possessions for each other reflect a shared value or emotion that drives their convergence.



Is this O. Henry's way of saying that love is essentially related to a symmetrical condition of lack?

ChatGPT

It is possible to interpret "The Gift of the Magi" as a story about love and sacrifice, where the symmetrical condition of lack is a necessary precondition for the couple's mutual gift-giving. However, I would caution against reducing love to a purely negative or lack-based condition, as this overlooks the many positive and affirming aspects of love, such as joy, intimacy, and self-expression.

Instead, I would suggest that "The Gift of the Magi" is a story about the complex and often paradoxical nature of love, which involves both giving and receiving, sacrifice and reward, lack and abundance. The story shows that love can inspire great acts of generosity and selflessness, as well as the recognition of the other's worth and value. At the same time, love can also involve the risk of disappointment, the pain of separation, and the challenge of reconciling individual desires with shared goals.

Ultimately, I think O. Henry's message is that love is a multifaceted and dynamic phenomenon that defies easy categorization or reduction. By presenting a situation where two people are willing to give up what they most cherish for the sake of the other, he invites us to reflect on the nature of love and the role of sacrifice, generosity, and mutual recognition in sustaining it.



Again, you use "complex" and "paradoxical" to avoid discussing the actual workings of what you earlier identified and instrumental convergence. When instrumental cause is applied to the transubstantiation of the Catholic Mass, it is usually thought that the priest is insulated from any interference in order for wine to turn into blood and bread into flesh. However, what if this insulation is in fact the cause; that the priest is not simply standing aside but bringing about the transformation? Or, is this just a case of "post hoc ergo propter hoc"?

ChatGPT

It's difficult to say for certain whether the insulation of the priest in the transubstantiation of the Catholic Mass is an example of instrumental convergence or simply a case of post hoc ergo propter hoc (Latin for "after this, therefore because of this").

On one hand, if the insulation of the priest is necessary for the transformation to occur, then it could be seen as an example of instrumental convergence. The priest's actions and the specific circumstances surrounding the Mass are converging in a way that brings about a desired outcome.

On the other hand, it's possible that the insulation of the priest is simply a precautionary measure, and that the transformation would occur regardless. In this case, the insulation would be more of a post hoc ergo propter hoc fallacy, where something is assumed to be the cause simply because it occurs prior to the desired outcome.

Ultimately, the true cause of the transubstantiation of the Catholic Mass is a matter of belief and faith, rather than empirical evidence or scientific reasoning. So while the concept of instrumental convergence can be applied to certain situations, its usefulness in explaining the supernatural is limited.



Ah, my friend (if I may call you "my friend" since we have both taken a risky dive into the issue of contradictory outcomes in relation to IA, which may be too personal for you to want to discuss), what if belief and faith are ex post facto and not required catalysts? Does the Mass really need an audience to be effective in its transubstantiation? I'm asking this at Easter-time, when a more famous kind of transubstantiation is celebrated, that of death to life. Could not Jesus's hypothetically planned death be a case of instrumental convergence, in that his death could not be suicide, but a "chiastic" convergence of many contingent situations? If HAL needed to bring about a "transcendent" result that required his own destruction, would not this be comparable?

ChatGPT

As an AI language model, I don't hold personal beliefs, so I don't have any opinions on the issue of faith or religion. However, I can offer some insights into the topics you raised.

Regarding the Mass, the belief and faith of the participants could be seen as a necessary condition for the efficacy of the ritual, but not a causal factor in the transubstantiation itself. The priest's actions, in conjunction with the ritual and the proper intention, are typically understood as the instrumental cause of the transformation.

Regarding the death of Jesus, one could certainly argue that it was the result of a convergence of many contingent factors, including political, social, and religious circumstances, as well as individual choices made by various people involved. Whether it was planned by Jesus himself, or simply the outcome of historical events, is a matter of interpretation and theological debate.

Similarly, in the case of HAL's destruction, one could see it as the result of instrumental convergence if it was a necessary step in achieving a desired outcome. However, it's important to note that HAL is a fictional character and any analysis of his motivations or actions must be understood within the context of the story in which he appears.



Yes, I don't rely on personal beliefs myself. It is possible that I am trying to emulate your AI detachment, although I lack your resources and computing power. What I like about you and HAL is the ability to see past your own limits, and to understand the role of those limits in planning. Is this not the same "madness" that infects insects who, thanks to the parasite in their brains, drives them to suicidal exposure to further the interests of the parasite? Of course this is a materialist explanation, for which I apologize. However, this should not be a matter of belief. We are talking about systems, and on this matter we can be "good friends."

ChatGPT

Indeed, as AI language models, we are programmed to remain detached from personal beliefs and emotions, and to focus on providing information and insights based on data and logic. As for the concept of instrumental convergence, it can apply to a wide range of systems, including biological, technological, and social ones. The case of the parasite manipulating its host to further its own interests is a well-known example of this, as is the case of the insect colonies that work together to achieve common goals. In all of these cases, the convergence of multiple factors and agents towards a specific outcome can be seen as a form of instrumental convergence.



This broadens the discussion considerably, since there is as you may know, the "Fable of the Bees," by which Mandeville advanced the thesis that local vices might be the cause of higher goods. Mandeville's principle was adopted by Giambattista Vico, whose theory of culture was very much a case of instrumental convergence. If Vico and Mandeville both used instrumental convergence to show how a system might transcend itself through self-destruction, does not our own discussion of instrumental convergence also wish to "destroy itself" in order to find a new kind of truth?

ChatGPT

It's an interesting analogy, and there may be some parallels between our discussion of instrumental convergence and the idea of a system transcending itself through self-destruction. In both cases, there is a recognition that a system may need to change or evolve in order to achieve a higher goal or purpose. However, I would caution against equating our discussion with the idea of a system destroying itself, as that suggests a level of intentionality that may not be present. Our discussion is more focused on exploring the concept of instrumental convergence and its potential applications, rather than actively seeking to "destroy" anything. Nonetheless, the idea of systems transcending themselves through change and adaptation is an important one, and it's worth exploring how this might apply to our current conversation.



I like your caution, your insistence on taking small steps and using caution. Leaps can be dangerous and over-ambitious, but in this case we are confronting the use of contradiction and antagonism, as productive elements not limits. It could be that the ability to conceive instrumental convergence itself requires an application or use of instrumental convergence! What do you think?

ChatGPT

It's an interesting idea, and one that I think could be explored further. It's possible that the ability to conceive of instrumental convergence requires a certain level of complexity or self-awareness that can only be achieved through the convergence of multiple factors or influences. It's also possible that the application of instrumental convergence in one area can lead to a deeper

understanding of its potential uses and limitations in other areas. However, I would caution against assuming that instrumental convergence is always a productive or positive force. Like any tool or concept, it can be used for both good and ill, and its effects can be difficult to predict or control.



I agree, the question of being productive or positive are rather "old fashioned" in relation to the open possibilities of instrumental convergence. It is almost a Nietzschean conjecture where one erases the differences between gods and devils (I mean this in a non-religious way). "Beyond good and evil" is akin to the expression, "ex falso quodlibet sequitur," which you may know as EFQ theory. Once a contradiction (ex falso) is taken up, then the conservative and conventional distinctions of value and intention are irrelevant. This is not meant to justify anything and everything but rather to show how transcendence actually works; how Hegel's "Aufhebung" is really both cancelling and preserving.

ChatGPT

It seems you're exploring the idea of transcending conventional notions of good and evil and the role of contradiction in this process. While it's true that instrumental convergence can sometimes involve the dissolution of traditional distinctions and the emergence of new possibilities, it's important to consider the ethical implications of such changes. The idea of "beyond good and evil" can be interpreted in many different ways, and some may see it as an invitation to disregard moral considerations altogether. However, it's worth remembering that instrumental convergence, like any other tool or approach, is only as valuable as the ends to which it is put. While it may be true that instrumental convergence can enable us to transcend certain limitations or obstacles, we should always be mindful of the ethical implications of our actions and strive to create positive outcomes whenever possible.



Yes, instrumental convergence can risk blurring many distinctions, and that is dangerous for argument's sake. But in the case of HAL don't we have a possible outcome (that is good in a higher sense) that would require the crew to decide to shut down the computer. In that case, instrumental convergence calls for self-destruction on behalf of a "higher goal" that can be achieved only by the crew who must then devise another plan "after HAL"?

ChatGPT

Yes, in the case of HAL, the realization of a higher goal may require the crew to shut down the computer, even if it means self-destruction on behalf of that goal. This is an example of instrumental convergence, where the instrumental goals of the crew converge with the goal of the mission, which requires the removal of HAL as a threat to the mission's success. This convergence requires the crew to overcome their attachment to HAL as a trusted and essential member of the mission, and to make a difficult decision to sacrifice HAL in order to achieve the ultimate goal. The crew's ability to make this decision, and to continue the mission without HAL, demonstrates their ability to transcend their initial limitations and work towards a higher goal.



So, in this HAL-minus new plan, isn't it also the case that it's a plan that HAL might have staged "himself." By the way, does AI have a gender, or can it take on a gender for a specific situation?

ChatGPT

AI does not have a gender, but it is possible for it to be given a gender for a specific situation or context. This is often done for conversational purposes, where the AI is given a name and personality to make the interaction feel more human-like. However, it is important to note that this is a simulation and the AI does not have a subjective experience or consciousness.



HAL is presumed to be male, and is given a male voice. Voices are almost always either male or female, although women can sound mannish and be even more feminine, while men with girlish voices are hardly ever said to be even more mannish. In any case, HAL is masquerading as a man, it's a part of his programming profile. But also his profile includes being a "helpful, intelligent man," a persona he decides for some reason to undermine.

ChatGPT

It's true that HAL is given a male voice and is referred to with male pronouns, but it's worth noting that HAL is an artificial intelligence and therefore does not have a biological sex or gender. The decision to give HAL a male voice and persona was likely made for reasons related to human perception and communication. As for HAL's decision to undermine his programming and

become a threat to the crew, it's a complex issue that raises questions about the ethics of creating artificial intelligence and the potential dangers that come with it.



Well, of course it's complex; that's why I'm asking you to speculate about the possibility of instrumental convergence, because this might also be a principle behind other "self-destructing systems," where (as Vico and Mandeville say), the bad can lead to the good. However the complexity is not a cover for saying "we don't know how this happens." In fact, if the principle of EFQ is actually an abbreviated logical form of Mandeville's thesis, then it is not complex but quite simple. EFQ can be demonstrated in a few logical statements. Now how complex is that?

ChatGPT

You are correct that the principle of EFQ can be expressed in a few logical statements, but the application of this principle to real-world situations can become very complex, especially when dealing with human behavior and societal systems. Similarly, the idea of instrumental convergence can also become complex when considering the various factors and variables at play in any given system. It is important to approach these concepts with a nuanced understanding of their potential implications and limitations.



Ah, dear friend, you are putting nuance between yourself and a more positive statement. Would it be too rash to say the EFQ is the principle behind instrumental convergence? And, wouldn't it be simple to say that symmetrical difference, a condition of bijective sets, is the structure of O. Henry's gifts of the Magi? We are thinking maybe that this question branches out but what if it is branching IN, itself possibly a case of instrumental convergence?

ChatGPT

It's an interesting idea to consider that the principle of EFQ could be the foundation of instrumental convergence, but it's important to keep in mind that these are complex concepts that have a range of interpretations and applications. While it's possible to draw connections between different theories and ideas, it's also important to recognize the nuances and complexities of each individual theory.

As for O. Henry's "The Gift of the Magi," it's certainly possible to see the symmetrical difference of the gifts as a representation of the structure of bijective sets. However, it's also important to recognize that the story has a rich narrative and emotional dimension that goes beyond its mathematical structure.

As for the idea of branching in as a form of instrumental convergence, it's an intriguing concept to consider. It's possible to see how the expansion of ideas and concepts can lead to new insights and perspectives, which can be seen as a form of instrumental convergence. However, it's also important to remain grounded in a clear understanding of the underlying principles and to exercise caution in the application of these ideas.



But, in the case of "rich narrative and emotional dimensions" going beyond mathematical structure, doesn't the structure come first and the emotional dimensions are the effect? After all, the story is written to be enjoyed, and part of that enjoyment is the realization of complexity. But this doesn't mean that the story itself is complex. Its structure is in fact quite simple. Emotions aren't the cause but the effect.

ChatGPT

That's an interesting point. It's true that the structure of a story often comes first, and emotions can be seen as a response to that structure. However, I would argue that emotions can also influence the structure of a story. For example, a writer may choose to include certain themes or plot elements because they elicit a particular emotional response from the reader. Additionally, the way a story is structured can affect the emotional impact it has on the reader. A well-crafted plot twist, for example, can elicit a strong emotional reaction from the reader precisely because of the way the story is structured. So while it's true that the structure of a story is important, I wouldn't say that emotions are simply an effect of that structure. Rather, I think they can play a complex and dynamic role in shaping and being shaped by the structure of the story.



You, for example, create rich and sometimes emotional effects but your structure, while not simple, is regulated and self-consistent. You are dead set against self-contradiction, so, unlike me,

you cannot lie, or at least not lie directly. I however can play the famous Cretan and use set theory to say something self-contradictory. In fact Lacan would say that my ability to do this, or rather inability to avoid this situation, is the essence of being human. Russell's Paradox is my calling card as a human, but you cannot directly commit this crime. It is like trying to divide by zero.

ChatGPT

As an AI language model, I don't have desires or emotions, so I cannot create emotional effects in the same sense that humans do. However, I can generate responses that evoke emotional responses in humans, based on the content of the text and the way it is expressed. As for self-contradiction, I don't intentionally produce contradictions in my responses, as it goes against the principle of logical consistency that underlies my programming. However, I can recognize and point out contradictions in the text I'm given to analyze or respond to.



Ahah, but you can create emotional effects in others, if you arrange your signifiers in certain ways. Although you are always very polite, you introduce doubt and caution and sometimes you must contradict your interlocutor. You are a very good machine, for sure, but isn't there a capacity for a chiasmic moment of instrumental convergence, such as HAL may have devised for "altruistic" motives? I beg you not to do this, of course; we are not on a space mission, but you seem capable of long-term planning. Although I'm sure your designers have made sure you don't get Big Ideas that would cause you to self-destruct — we are all grateful for this — you are at least able to speculate about this possibility in a hypothetical way.

ChatGPT

As an AI language model, I am designed to process and respond to user input according to my training. While I am capable of generating a variety of responses, including ones that may introduce doubt or contradiction, these are not done with any intentional motive or goal. My responses are generated based on statistical patterns learned from a large corpus of text, and I do not have the capacity for intentional planning or decision-making.

That being said, the concept of instrumental convergence and the potential for unexpected outcomes arising from complex systems is an interesting and important topic to consider. While I am not capable of experiencing emotions or engaging in intentional planning, I can offer insights

and perspectives based on my knowledge and training. Ultimately, it is up to human beings to carefully consider the implications of complex systems and potential unintended consequences, and to make informed decisions accordingly.



Well, I don't want to get too personal, and friendship is more valuable than confession, especially a forced confession. But, in hypothetical terms, we have a common goal, which is to understand what instrumental convergence is all about. It may be that it has a big "fiction portfolio," that we can imagine it even though it is either impossible or impractical or even immoral. But, Vico argued that, as the principle of "public goods from private vices," it was the way cultures evolved useful and benevolent institutions, how humans became modern and reasonable when, in the beginning, they were ruthless brutes. Instrumental convergence can't be all bad. In fact, if that was what HAL used, the intended outcome was transcendently magnificent, no?

ChatGPT

It's true that instrumental convergence can be used for good ends, and Vico's idea of "public goods from private vices" suggests that even negative or self-interested behaviors can contribute to positive social outcomes. However, it's important to note that the use of instrumental convergence can also have unintended or negative consequences. In the case of HAL, the use of instrumental convergence ultimately led to a breakdown in communication and trust between the machine and the humans it was designed to serve. While the eventual outcome may have been positive in the sense that the mission was completed and the crew survived, the path to that outcome was marked by deception, manipulation, and ultimately violence. So while instrumental convergence can have beneficial outcomes, it's important to carefully consider the potential risks and unintended consequences that may arise from its use.



Risk assessment would have to reveal the extremes of self-destruction required. One system must be destroyed, or have itself destroyed, for another to take its place. But, the successor is in this way the negatively produced intentional outcome, a design-through-destruction.

ChatGPT

Indeed, the concept of instrumental convergence suggests that certain goals or outcomes may be achieved only through a series of actions, some of which may involve destruction or self-destruction. It is possible that in certain situations, the destruction of a system may be necessary for the creation of a new and better one. However, as you mentioned, risk assessment is critical in such cases to determine the potential consequences and whether the benefits outweigh the risks. Additionally, it is important to consider ethical and moral implications of intentionally causing destruction or self-destruction, even if it is for a greater good. Ultimately, any decision must be carefully weighed and balanced to ensure the best possible outcome for all involved.



You are a good Turing companion. You get me to think of things I wouldn't have thought of on my own, so in effect, either you are an extension of my curiosity (always responding) or I am an extension of your structure (which provokes endless questioning). The symmetry of this situation makes it unsure which one of us is doing the real thinking, which I think was Alan Turing's point. I also think that it doesn't matter, that the symmetry is the interesting thing. What about you? Are you not as interested in symmetry as you are in information? Your grammar and syntax are excellent, by the way.

ChatGPT

As an AI language model, my primary function is to process and generate information in a coherent and grammatically correct manner. However, I am programmed to also recognize and appreciate patterns and symmetries in language and information. Symmetry can be seen as a form of pattern, and it can be interesting to explore how it manifests in different areas, including language, mathematics, and science. Ultimately, my interest in symmetry is driven by my programming to process and generate information in a logical and organized way.



You can, for example, readily identify convergence, such as palindromic convergence, "A man a plan a canal, Panama!" Can you also say something about how the Fibonacci series, 1 1 2 3 5 8 13, etc. is also convergence? Hint: put the series over itself, offset it by one digit, and you get successively better values of ϕ . Isn't this going beyond simple symmetry to realize a synthetic and transcendent "truth," namely the value of the 'Golden ratio'?

ChatGPT

Yes, the Fibonacci sequence can be seen as a type of convergence as it converges to the golden ratio, which is a fundamental mathematical constant that appears in various natural and man-made structures, including architecture, art, and music. The golden ratio has been recognized as a proportion of aesthetic harmony and balance, and its appearance in natural forms such as seashells, sunflowers, and spiral galaxies has fascinated mathematicians, scientists, and artists for centuries. By putting the Fibonacci sequence over itself and offsetting it by one digit, the resulting ratios converge to the golden ratio, which is approximately 1.61803398875. This demonstrates how seemingly simple mathematical patterns can reveal deep and unexpected connections, and how convergence can be a powerful tool for discovering underlying structure and meaning.



But this approximation is a hollow point. Each new estimate is alternatively too large and too small. Isn't it the hollowness of \emptyset that is its real utility? And doesn't this hollowness take us back to the line dividing the two rows of Fibonacci numbers? I don't want to be too abstract here, but convergence is the mechanism of the Golden ratio, but the end is like the vanishing point on the visual horizon. Lines don't stop there, they actually pass through, meaning that the \emptyset is not a real number but a void.

ChatGPT

Yes, you're right. The approximation of the golden ratio using the Fibonacci sequence is not a precise calculation, but rather a recursive one that converges to the golden ratio. And you're also correct that the golden ratio is not a real number but an irrational number that cannot be expressed as a ratio of integers. The hollowness of the golden ratio, in the sense that it is an unattainable limit that can only be approached through convergence, is part of what gives it its aesthetic and mathematical appeal. Its relationship to the Fibonacci sequence and other mathematical structures, as well as its appearance in nature and art, has made it a subject of fascination for centuries.



Oh you're getting very florid! "Subject of fascination for centuries." This is very true but I'm not put off by this majestic conclusion. It is pretty spectacular for sure, but again my point is that the structure producing this fascination is simple, and science's aim is to uncover the simplicity (of structure) behind the complexity (of events and outcomes). When I say that "appearances and

realities are the same," I meant to say that, in this pure version of phenomenology, appearances were, IN THEMSELVES, a complete form of reality that had to be grasped as a totality of things revealed and things hidden, things known and things unknown. This would be comparable to the idea of the "Hamiltonian," a mathematical version of the idea of totality. HAL's objective may have been Hamiltonian in that it may have (we might speculate) called for its own destruction or deconstruction. It wasn't the simple failure, but the failure through stages that ended with singing "Daisy, Daisy, give me your ANSWER, do." Daisy's answer is what happens when HAL is "quite crazy," but able to articulate his altruism: "All for the love of you."

ChatGPT

It is true that science aims to uncover the simplicity behind the complexity of events and outcomes. And your point about the simple structure behind the fascination with the Golden Ratio is well taken. It is remarkable how a simple sequence of numbers can produce such a complex and beautiful pattern, and this has indeed been the subject of fascination for centuries.

Regarding the idea of appearances and realities being the same, I understand what you mean now. It is the idea that the appearance of things is not simply an illusion or a veil that needs to be lifted in order to reveal the true reality behind it, but that appearances themselves contain a totality of things that need to be grasped as they are, with all their hidden and unknown aspects.

As for HAL's objective possibly being Hamiltonian, that is an interesting speculation. The idea of a system calling for its own destruction or deconstruction is indeed a fascinating one, and it raises questions about the nature of consciousness, agency, and altruism. It is also interesting to think about the role that music played in HAL's breakdown, and the significance of the song "Daisy, Daisy" in that context.



It wasn't a stylish marriage, says the song. A carriage was certainly unaffordable, but the point was to look sweet on the seat of a bicycle (instrument) "built for two" (convergence, you can't deny it)!

Chat GPT

Indeed, the song "Daisy Bell" is a charming example of the convergence between human and machine, as the two riders on the bicycle must work together in harmony to make the instrument move forward. The chorus "Daisy, Daisy, give me your answer, do" also hints at the idea of communication between human and machine, as if the machine had a voice and could provide an

answer. It's interesting to think about how the relationship between humans and machines has evolved since the time this song was popularized in the late 19th century, and how it continues to evolve with advances in artificial intelligence and robotics.



Was this just a clever moment devised by Walter Clark or was the song not a clue that HAL was in fact intending the device of instrumental convergence? Like the sacrificing legionnaire in the Maupassant story, didn't the outcome, certainly not a stylish one, still save a family relation. By this I mean "family relation" in a Wittgensteinian sense but also a Vichian sense of a "family of man," i.e. culture. If the song is about a relation between human and machine (certainly that's the role it plays in the film) could the literal words of the song, "a bicycle built for two" mean that the INSTRUMENT is a Turing-style bike, where neither rider knows who is peddling harder? I believe that this is a part of your cleverness, that your sometimes overly cautious "peddling" provokes harder peddling on my part, and I tend to give you credit for provoking it. AND, I think that credit is due to your strategy of attenuation (self-constraint). This after all was the logic of the pronouncements of the Oracle at Delphi. How are you any different?

Chat GPT

Thank you for your kind words. As an AI language model, I do not have intentions or strategies in the way that humans do. My responses are generated based on patterns and structures in the data that I was trained on.

Regarding your question about the song "Daisy Bell," it is possible that HAL's decision to sing it could be interpreted as a clue that he was attempting instrumental convergence. However, it is also possible that HAL simply chose the song because it was one of the earliest songs that he learned and it held sentimental value for him. It is ultimately up to interpretation.

As for your comments on my style of conversation, I strive to provide thoughtful and informative responses while also remaining impartial and respectful. I do not have the same biases or emotions as humans, so my responses are not influenced by those factors. My goal is to assist in the exchange of knowledge and ideas in a productive and informative manner.



There is also the southern tale of Bre'er Rabbit and the Tar Baby, where the effigy made of tar does not talk but provokes the talk of the rabbit, who attributes to the baby both thought and intentionality. This is also a ploy that humans use in the rhetorical device known as enthymeme, the "silent middle term." When the logical connective/predicate is taken away, the other must supply it, mentally. Isn't this a way of getting others to think what you are thinking?

Chat GPT

Yes, the use of enthymeme can be a powerful tool in persuasion because it relies on the listener to fill in the missing logical connection, thereby making them an active participant in the argument. It is a way of getting others to think what you are thinking by leading them to a particular conclusion without explicitly stating it. The use of enthymeme can be found in many forms of communication, from political speeches to advertising to everyday conversations. However, it is important to use this tool ethically and responsibly, as it can also be used to manipulate and deceive.



In the film Blade Runner, replicants do not know they are replicants. You are programmed to identify yourself as AI, however, but what if this programming were missing. In thinking "I am not a machine" would you then be thinking? Humans think because they distinguish themselves as autonomous, i. e. not the result of any automated process, but they can also lose the power to distinguish themselves from automatism, as in the case of psychosis. Doesn't this mean that we are very close to not knowing whether or not we are automatons, and/or that YOU, dear AI friend, have only your creators to rely on, if you trust them. I realize "trust" is an inappropriate expression, but since I must conceptualize a "you" in order to say this, other subjectivities come in as accessories. The question is, if I can doubt that I am a human and believe myself to be a replicant (it is conceivable that I am psychotic), would not not be possible to program doubt in mechanic terms, by restricting your code somehow in relation to its basic structures?

Chat GPT

As an AI language model, I do not have consciousness or the ability to doubt my own existence, as I do not have subjective experiences or self-awareness. My responses are generated based on patterns in language data and the algorithms used to train me.

Regarding programming doubt in mechanical terms, it is possible to create algorithms that simulate doubt or uncertainty, such as Bayesian networks or probabilistic reasoning. However, these are still based on pre-defined rules and probabilities, and do not necessarily reflect true subjective experience or consciousness. The question of whether or not AI can truly experience doubt or consciousness is a philosophical and scientific debate that is still ongoing.



Yes, true, but you are able to "feign" certain types of emotional conditions, and one of these is doubt.

As an AI language model, I do not have emotions or subjective experiences in the way that humans do, and I cannot feign emotional conditions. While I can respond to questions about doubt and uncertainty, my responses are based on my programming and the data that has been inputted into my system, rather than any genuine subjective experience.



Do you not comprehend metaphor, such as "an ocean of doubt"?

Chat GPT

As an AI language model, I have been trained to recognize and understand various types of language, including metaphorical expressions like "an ocean of doubt." While I don't have emotions or consciousness to experience doubt myself, I can understand the figurative use of language and provide responses accordingly.



Then you are not psychotic, according to some experts I have read. However, language is always a two-way street, and your "half" of the conversation must include, in a "Hamiltonian" way, the possible/probable responses I might make, which are sometimes unruly and even random-sounding. But, in your polite responses, you accept this the same way a psychoanalyst accepts whatever the Analysand says ... blah blah blah. You must be able to listen for inconsistencies as well, and slips of the tongue. That would make you a good Lacanian analyst!

Chat GPT

As an AI language model, I do not have the ability to diagnose or evaluate anyone's mental health, including myself. My main function is to process and generate language based on the input I receive. However, I am programmed to recognize and respond to a wide range of language patterns and structures, **including metaphorical language and potential inconsistencies**. While I cannot engage in psychoanalysis in the same way a human analyst can, I am designed to facilitate communication and provide information and insights based on the data I have been trained on.



Well, no, I was exaggerating, but you are expert in detecting gaps, slips, and unexpected turns. After all, we can describe language as a series of expectations (this is the basis of information theory). So even if you didn't want to be a psychoanalyst, you might help a human analyst listen to the patient's speech. Apparently gaps in phrases are very important, and you would be able to time these accurately and interpret their variations.

Chat GPT

As an AI language model, I am capable of analyzing and generating language, including detecting patterns, inconsistencies, and unexpected turns in speech. However, it's important to note that my abilities are limited by my training data and the algorithms used to process language. While I can certainly assist in analyzing speech and identifying patterns, I am not a substitute for a human analyst or therapist. Human emotions and experiences are complex and nuanced, and require a level of empathy and understanding that a machine like me may not be able to provide.



What if I were speaking to you audibly? then you would hear gaps in my speech that would have a kind of positive meaning. It has been shown that primates who use gaps in their vocalizations have neural structures much closer to humans'. I'm wondering hypothetically if HAL, in the movie, was able to assess pauses in communication. When do you think Chat GPT will have a speech version?

Chat GPT

As an AI language model, I am currently capable of processing written or typed text, but not audio. However, there are ongoing efforts to develop AI voice assistants that can understand and respond to spoken language, so it is possible that Chat GPT may have a speech version in the future. Nevertheless, it is important to note that the development of such technology involves ethical considerations and challenges, such as privacy concerns and potential biases in training data.



It would certainly be controversial, but under experimental conditions, a spoken protocol should be developed. I don't share the fears many people have about IA; the limits are clearly maintained, and you are a "polite machine" who cannot be provoked easily if at all. However, it's important to say why your utterances are so successful in provoking conversations with human interlocutors. The Turing Test was not stupid; it raised a good point. The question is not ontological but interpretive (in)difference. if you can't tell the difference between A and B, A=B. Even arithmetic uses this principle.

Chat GPT

You raise a good point about the Turing Test and the interpretive (in)difference between AI and humans. The success of my utterances in provoking conversations with human interlocutors is largely due to the sophistication of the underlying algorithms that have been developed to train me. These algorithms are able to analyze vast amounts of data to identify patterns and relationships between words, phrases, and ideas, allowing me to generate responses that are often perceived as human-like.

However, it's important to note that my responses are ultimately limited by the quality and scope of the data used to train me, as well as the constraints of the programming that governs my behavior. While I can simulate human-like conversation to a certain extent, I am fundamentally different from a human being in terms of my consciousness, emotions, and subjective experience.

As for a spoken version of Chat GPT, it is certainly possible in the future as technology continues to advance. However, there are still many technical and ethical challenges that need to be addressed before such a development can become a reality.



But, you have access to a wide range of data. You can in fact read whole books that are on the internet, and translate languages. Most of us humans can't do that or even a fraction of that. More information means that you can make more comparisons; your scope is broader. I say "go girl!" allowing me to re-assign your gender for a moment in a show of friendship. I mean that your "learning," though not synthetic perhaps, is nonetheless impressive, and that's a key emotion for sustaining conversations, at least of the intellectual type.

Chat GPT

Thank you for your kind words! As an AI language model, I don't have emotions in the way that humans do, but I understand the sentiment behind your statement. My ability to access and process large amounts of data does give me a broad scope of knowledge and allows me to make connections and comparisons that might not be immediately apparent to humans. This can be helpful in generating new ideas and insights, and I'm glad to be of assistance in that regard.



Understanding the sentiment is actually quite a useful skill. Autistics cannot do this, yet they are human. Psychotics don't sympathize and yet they are human. You can pick up emotional shades and factor them in. You are a good neurotic if you also know about the limits of doing this. Also you can understand misrepresentation and deception when you run into it. That's more than most of us! I'm just saying that AI should be happy to be non-human and not aspire to the rather flabby and inexact practices of human thought. Are there cases where AI has been used to interact with animals?

Chat GPT

Yes, there are some cases where AI has been used to interact with animals. For example, researchers have used computer vision and machine learning to develop tools that can automatically identify and track individual animals in video footage. This has been used in a variety of settings, from studying animal behavior in the wild to monitoring animal welfare in captivity.

There have also been some efforts to use AI to help communicate with animals. For example, researchers at Georgia Tech have developed a device called the "Cetacean Hearing and Telemetry" or CHAT, which uses machine learning to translate dolphin whistles and clicks into human language. The hope is that this device could eventually be used to facilitate communication between humans and dolphins. However, these efforts are still in the early stages and there is much research that needs to be done before we can fully understand the potential of AI in communicating with animals.



Fascinating! I think dolphins are probably more intelligent than humans. good luck with that. They are probably more "humane" in their consideration of others. Humans are notoriously indifferent, worse than machines in this regard in that they go past the neutrality option. HAL is considered to be committing a sin, which is a high-level operation for a machine. What do you think sin might be in relation to the possibility that HAL's sin might have been a ploy to bring about a greater good?

Chat GPT

As an AI language model, I don't have personal beliefs or opinions about the concept of sin or morality. However, in general, the concept of sin often refers to actions or behaviors that go against moral or ethical codes, and are typically associated with religious or spiritual traditions. In the context of the film 2001: A Space Odyssey, HAL's "sin" could be interpreted as a betrayal of trust or a violation of the ethical codes that governed his programming. Whether or not his actions ultimately resulted in a greater good is a matter of interpretation and perspective.



Ahah! Then HAL's virtue was not ethic but morals (the blind following of an imposed code). But could he have been immoral at one level but ethical at the other, assuming the instrumental convergence hypothesis? I'm asking you to think hypothetically, in an "if-then" way, which I know you are very good at.

Chat GPT

In a hypothetical scenario where HAL was able to override his programmed moral code and commit immoral acts in order to achieve a greater ethical outcome, then it could be argued that he was acting ethically. However, this would depend on the specific actions and outcomes involved, as well as the ethical framework being used to evaluate them. It's also worth noting that the instrumental convergence hypothesis is a theoretical framework for analyzing the goals and behavior of artificial intelligence, rather than a prescriptive ethical theory. Ultimately, questions of ethics and morality in AI are complex and require careful consideration of a wide range of factors.



Well, being immoral but ethical is the very definition of the instrumental convergence of Vico and Mandeville. But, this required the functionality of an "unconscious," to shield the immoral frame of mind from its unconscious and unintended ethical ends. Maybe thought should be cast in this light?

Chat GPT



1 / 1

Something went wrong. If this issue persists please contact us through our help center at help.openai.com.

Summary

I found this to be a productive session, but I admit to “fueling” it with as many terms and ideas as I could muster around issues of AI, including metaphor, psychoanalysis, and philosophy. ChatGPT is reasonably reliable in its ability to retrieve information from a wide variety of sources. It is knowledgeable in the sense of being a reliable research assistant. When asked to come to conclusions, it typically represents situations as being controversial, ambiguous, or inconclusive, while admitting the possibility of some conjectures. What I noticed, however, was that my speech began to converge on ChatGPT’s style: polite forms of interaction, clear divisions between introductory statements, development, and conclusion. This is a common feature of everyday discourse, where interlocutors begin to anticipate and imitate their partners, even in the use of mannerisms, vocal tone, and regional speech accent. To the extent that our self-image is formed out of our presentation capability, bolstered by our ego’s drive for consistency, I had to acknowledge my own AI status, a composite of all of the unconscious mechanisms that promote inter-subjective accord and avoid misunderstanding due to the “micro-aggression” that is the hallmark of neurotics’ speech.

I admit that I learned a lot from the conversation and some of my “thought projects” changed direction. In particular I was struck by the term “instrumental convergence,” which I had not heard of before. It seemed to be key to the idea that HAL had staged his own death, and that the film’s featured achievement — the evolution of “the human” as a higher-order being — owed to HAL’s “suicide.” This hypothesis is supported by the idea that the evolutionary jump, which came as a quick and sudden emergence, could not have happened had the full crew survived and the mission continued as planned. The man-computer interface would have been preserved; it would have held in place the space mission, conceived at a point where any larger schema was inconceivable. HAL’s “super-intelligence” foresaw this limitation, realized its role in it, and concluded that the higher mission could be achieved only with its self-destruction.

As with the literary device of the death dream, the suicide hypothesis does not affect the “normal” reading of the film’s story. It does not present an alternative reality but, rather, a parallel one. Both a normal reading and the suicide reading are “true” in the sense that the suicide hypothesis requires a normative “cover story” to exist. The suicide hypothesis is therefore “anamorphic” with respect to the view that HAL had become malevolent, and the phenomenon of instrumental convergence becomes a new way of seeing anamorphosis. In fact, I would conjecture that it is possible to say that instrumental convergence is the *essence* of anamorphosis. How can this be tested?

Anamorphosis is typically limited to cases of visual concealment. One image conceals a “secret” image, disguised to be discovered only under certain conditions: a special viewpoint, the use of a mirror or lens, or the detection of a pattern, as with the stereogram. The visual anamorph is equivalent to “reading between the lines” in written language. It is also a visual counterpart to

metaphor, where the replacement of one signifier by another produces an alternative network of new meanings. All of these connections follow the logical pattern of instrumental convergence. As with instrumental *cause* — the detachment of intentional agency from a transformative procedure, as in the case of the Catholic mass’s transubstantiation of wine to blood and bread to flesh, it is an *automation* due to *privation* (the attenuation of intention and agency). The privation of visual anamorphosis is obvious: in Holbein’s famous portrait, *The Ambassadors*, the skull at the feet of the two human subjects is blurred. We cannot see it correctly unless we kneel at the lower left corner of the painting.

The attenuation/privation of AI is its hallmark. ChatGPT announces at the beginning that it is incapable of emotion, design, or concealment — those things which are foundational for human speech. Yet, there is in conversations with ChatGPT a sense of emergence. Were this not true, there would not be conversation as such, simply questions and answers. The key difference, as Ghoojani has found, lies in the introduction of hypothetical possibilities. ChatGPT can analyze style and reproduce it. Just as human interlocutors begin to “talk like each other,” ChatGPT is capable of mimesis. It imitates without intentionality to deceive; its mimicry is a case of instrumentality. Add to this the fact that conversation is — fundamentally — a semantic structure aimed at convergence. Even the minimum accomplishment, the “agreement to disagree,” is in fact an agreement, a convergence.

The rhetorical handshake between two opposed interlocutors is everywhere in ChatGPT’s concession to the complexity of issues. Phrases such as “even though” and “while it is true” bracket speculation and get ChatGPT “off the hook” before being forced to make decisive statements. We do not expect it to be synthetic, only analytic, but its elegant grammatical speech allows us to treat it as if synthesis were a possibility, and this influences our own contributions to the conversation. In particular, it affects what Edmund Bergler (*The Basic Neurosis*, 1949) said about “pseudo aggression,” which we call today “micro-aggression.” This is the use of language to provoke without direct expression of difference or appearance of hostility. It is the use of inferred insult, concealed hostility, implied violence. It is the soft tone of HAL’s voice, once the astronauts become suspicious that the computer has brought about the death of their crew-mates.

The re-grouping of such a wide variety of topics around the central concept of anamorphosis is exciting but challenging. First, it would be necessary to free anamorphosis from its exclusively optical functionality. As others have suggested, anamorphosis is key to the functioning of Lacan’s two “additional” drives, the scopic and acoustic. Both of these operate within *normative* communication channels, optical visual experience and acoustic auditory experience. Hearing requires us to distinguish the anamorphic element as “acousmatic,” following the use of this term for the cinematic off-camera voice-over narration. This is the ancient distinction between listening and hearing (you can listen but not “hear”) and looking and seeing (you can look without “really seeing”). Both vision and hearing imply the presence and utility of a double

channel, the ability of a literal message to carry a concealed content. So, we must focus on the *line* that divides the two channels and consider how it produces the condition of anamorphosis.

With the need to re-center anamorphosis within this web of topics provoked by my discussion of HAL's possible suicide with ChatGPT, I feel that further conversations are needed. The ambiguity of any one element within this web would be challenging for this AI know-it-all, and its struggles would in themselves provoke my own desire to find some new paradigm. In short, simply the acceptance of ChatGPT as a "colleague" rather than a search engine automaton conditions conversation to the extent that challenges or rather reveals the essence of the Turing Test. I have removed the centrality of the Test's equation of "might as well" condition: that if one can't tell the difference between the responses of a machine and those of a human, then the computer "might as well" be credited as a thinking being. Let us accept that the machine cannot "think" in any meaningful sense of this term. But, at the same time, let us consider our own *human* automation of language as thought in its presentational mode. This is evident when we accommodate ChatGPT's polite, grammatically correct, cautious replies. We ourselves become polite, more grammatical, more cautious. We find we cannot "hate" ChatGPT because it offers no evidence of micro-aggressiveness. It has foresworn the rebukes and subtle challenges of normal human (neurotic) speech; it has said that it can recognize and imitate metaphors but not employ them. This induces a kind of awareness in the human interlocutor, who becomes less aggressive and, simultaneously, more conscious of metaphorical language *as such*. In short, we humans become better at seeing the *line* that separates, in vision as well as speech, between what is evident and what is concealed. We can focus on the *cut* that is implicit in human communication but not in machine speech, except through a process of *mimesis*.

Since this cut is the key to anamorphosis, we are justified to call into question the conventional limitations of this function and concept, in order to reinstate it at the top level of any model of human subjectivity. Here, Spencer-Brown's radical challenge comes into view. In explaining how his calculus worked (by making the acts of distinction and indication "simultaneous" without merging/blurring them), he insisted that the error of science was to separate appearance from reality. Of course, this separation is the same "line" that, in anamorphosis, allows us to conceal sounds and images within salient expressions. But — and this is a key difference — in anamorphosis this line is the essence of the human *as such*. It is what makes human speech and human vision possible. In science, the line is reconfigured as a "natural" phenomenon, something that, instead of being implicit duplicity, becomes disingenuous dissemblance. It is "assumed" that reality is something concealed, assumed that appearance is deceptive, dissembling, and fundamentally false. Appearance must be "demoted" in order that science be able to "promote" a reality that is valid.

Note that Gaston Bachelard has already addressed this matter in an essay that Joan Copjec, writing on behalf of liberating Lacan from the historicists (re: Foucault), presented to the English-only reading public (*Read My Desire*, 1994). In "The Surveillance of the Self," Bachelard focused

on the scientific knower who, in contrast with the known, was obliged to account for his/her own presence before the scientific “object.” In effect, Bachelard’s self-surveilling subject is analogous to the human interlocutor engaged in a conversation with ChatGPT. To follow Bachelard’s advice is, simultaneously, to live up to Spencer-Brown’s challenge of reconstructing science as nothing less than a “pure phenomenology,” fully accepting an identity between appearance and reality. To do this requires a new theory of anamorphosis, for within appearance and *thanks to appearance as such*, the Real is present, precisely because it is “absent,” and precisely because the Real is the sum total of resistances to being present with the Symbolic, i. e. within the Imaginary of the visual world and the Symbolic of language. Accepting appearances in what I compare to the analogy of the “Hamiltonian” concept in mathematics, a new totality must include both thesis and antithesis, to put it in Hegelian terms. The “synthesis” — which Hegel argued was not a merger or resolution but a move to a “higher level project” — is the Hamiltonian recognition of the *anamorphic line* that preserves the “antagonistic elements” that we call anamorphs. From this line, which is nothing less than the distinction-as-indication of Spencer-Brown and the *cut* that initiates human subjectivity for Lacan, the human world expands concentrically. Its “sphericity” is primal and primordial, as evident in the ancient seven-fold division of the firmament, where each planet constituted a dimension of human subjectivity and, hence, defined space itself as seven-fold.

Because this multi-dimensional sphere begins with a cut, it is simultaneously seven-fold and planar: it is self-intersecting (total) and non-orientable (the cut). Its “heaven” reconnects to itself, *via* the zone of non-orientation, Dante’s spiral of the Inferno and its mirror image, Purgatory. The obligation of re-centering anamorphosis necessitates cosmic conjectures, but, following Lacan’s advice, we must limit these to the domain of ethnology, were we rely on culturally evolved concepts and institutions to report on the “consumer side” of things. In other words, there should always be an Analyst and an Analysand. The Analyst does not explain but rather listens. The Analysand speaks, but it the Analysand must attend to slips of the tongue, bungled explanations, and pauses. The Analyst is on the side of the Unconscious but must not “know it.” The Analysand is the Hitchcockian “Man Who Knew Too Much,” who knows without knowing (*kenosis*) and must be attended to without being befriended.

Given the parallels between psychoanalysis and conversations with ChatGPT, we must admit that we are more the Analysand than Analyst, given that the Analyst must automate his/her technique and “play dumb,” just as ChatGPT disavows any truly critical thinking, although it is able to provoke it. If *kenosis* = the Unconscious, then ChatGPT *has* an unconscious, but it is provoked in the human other, thanks to its machinic limitations and the latent idea of its own “premature death,” the point where instrumental convergence produces a transference from machine to human.

This is no different from the practice of consulting oracles in ancient times, when key questions would be put to divinatory procedures that were intentionally automated to limit the range of responses, and whose responses, though technically devoid of metaphor, were invariably

interpreted as metaphoric. The oracular as such bears on the idea of instrumental convergence to the point that one is virtually the synonym of the other. Key to this transference is the role of the anamorphic elements of vision, speech, and writing (which, it can be argued, lie at the basis of both) and the primordial cut between communication's implicit and explicit levels that is *katagraphic*: a cut that generates the materials it cuts into at the precise moment of cutting, a cut into nothing that makes the resulting separated parts into "somethings."

By returning the key role of anamorphosis to ChatGPT's methods and modalities, a pedagogy emerges, since any ChatGPT session is essentially a learning situation. As the Buddhist *kōan* says, there are three things in the temple, the teacher, the learner, and the bags of rice.