The Delphic Room

An Artistically Derived Metaphor

Jalhed, Hedvig; Rylander, Mattias; Åberg, Kristoffer

Conference paper

AR@K24 Symposium:
The Artist, The Ghost and The Machine

Kristiania University College, Oslo
18–19 March 2024

The Delphic Room

An Artistically Derived Metaphor

Hedvig Jalhed
Mattias Rylander
Kristoffer Åberg

Abstract

In his well-known thought experiment regarding artificial intelligence (AI), John Searle sketched out the philosophic idea of "The Chinese room" – a room in which comprehensible rules (a program) allow a person to perfectly correlate one set of unknown linguistic symbols (a question) with another (an answer) of the same unfamiliar kind. In our creation of an AI-based micro-opera for humans and machines, we have come to reflect upon our concept as an artistic response to Searle's arguments and a mirroring complement to his debated figure. Our immersive and interactive opera was conceived as a modular series of musically paced meetings between individual visitors and a singing seeress in contact with the digital realm. As an analogy to the Delphic oracle, the seeress delivered AI-prompted answers to the visitors' questions in real time, framed by poetical, musical, and theatrical structures. In Searle's Chinese room, goal-oriented computational mechanisms remain detached from understanding during the linguistic operation. In our Delphic room, understanding is key for carrying out the aesthetic operations intended to artistically stimulate a coupling of intellectual and visceral information processing in open-ended and personal ways.

Keywords: artistic research, philosophy, experiment, opera, artificial intelligence, interaction, information, Delphic oracle, Searle, Chinese room

But you would have lots of fun with me For instance, I am the greatest ventriloquist that ever lived, I am the first ventriloquist in the world!

from *The Phantom of the Opera* by Gaston Leroux

Introduction

In this paper, we expand on how artificial intelligence (AI) can be integrated into opera in an artwork that is presented as both experiment and metaphor. The paradoxical dream of displaying human nature by way of mechanically performing beings permeates operatic technique and technology. In opera, human performers can be disciplined into machine-like creatures – opera singers have been described as singing and acting machines (Frigau Manning, 2013). Opera is also said to have technology as such as its "true object," taking the shape of human attempts to transcend the conditions of earthly existence by making and using tools and machinery (Ridout, 2012). While opera today might be contrasted with more explicitly media-rich genres, "the history of opera is fundamentally intertwined with media history, as well as with the broader history of technology" (Sterne, 2016, p. 159). What we find throughout the continuous development of spectacular technologies for the operatic arts, is the relationship between human and machine. Drawing on such a view, integrating AI as a creative technology into opera comes not as a breach of operatic tradition, but a continuation of it.

Previous research using speech shadowing in face-to-face interaction shows how so-called echoborgs – humans whose actions and outputs are determined by a computer program – increase the program's chances of passing as an autonomous human mind compared to text interfaces (Corti & Gillespie, 2015). Actors and opera singers have always been echoborgs – or, rather, "cyranoids" (Milgram, 2010) – as they deliver lines and actions from instruction rather than their own spontaneous intention. The novelty is how we can now use computer programs as instructors in real time.

The question of intelligence and creativity in science and art

Our artistic exploration stems from the situation depicted in philosopher John Searle's thought experiment The Chinese Room (1980). In short, the argument concerns the distinction of mind from brain, the equation of mind with intelligence, the reduction of intelligence to calculation, the possibility for disembodied minds to outperform embodied minds, and the role of intentionality in mentalization. Without aiming to cover the whole debate about the Frankensteinian hope that AI can display properties of a humanoid mind, a brief note on the topic will frame our artistic position.

In refutation of a certain strong AI hypothesis, defined as an appropriately programmed computer being a mind that has cognitive states and the ability to understand, Searle (ibid.) presented his counterargument in the form of the now renowned thought experiment, The Chinese Room. A man who knows English but no Chinese is locked in a room, and is given two batches of Chinese symbols together with a set of rules in English for correlating the second batch with the first batch only by identifying the shapes of the symbols. Given a third batch of Chinese symbols together with further instructions in English, the man is able to correlate elements of the third batch with the first two batches, again in response to certain sorts of shapes of the symbols. Unbeknown to the man, the first set of symbols is a script, the second is a story, the third contains questions, the man's responses to the questions are answers, and the instructions in English are a program. Further complicating the situation, the man is also given stories in English and is asked questions about the stories which he can answer even though some information is not explicitly stated in the stories. This is due to the man's understanding in the form of scripts – human beings' representations of the sort of information related to concepts like stories. After a while, the man improves at following the instructions for manipulating the Chinese symbols, and the programmers improve at writing the programs to the extent that, from the point of view of somebody outside the room, the answers to the questions are absolutely indistinguishable from those of native Chinese speakers. However, the man only understands the manipulation of the English symbols, and does not understand the manipulation of the Chinese symbols at all, although ostensibly analogous computational operations are performed. Thus, the specific strong AI argument is presumably refuted.

With recent technological advancements, AI has gained the capacity to computationally produce visual, audial, and verbal stimuli from existing human data with increasing variety and velocity. However, it has also become clearer that artificial processes which generate such output differ radically from evolved ones. Large language models like ChatGPT use extreme amounts of data to find correlations and probabilities, but the human mind operates with small amounts of data to generate descriptions, predictions, and explanations of what is also not the case and what could or could not be the case (Chomsky, 2023, n.p.). Accordingly, creative imagination is necessary if we hold *explanations* to be what philosopher of science Karl Popper (1963) described as powerful and highly improbable theories. Such theories are counter-intuitive solutions with a capacity to change both minds and actions. Creativity as a condition for what we call intelligence was proposed by mathematician Ada Lovelace in the 19th century. Moreover, it has been pointed out that creativity should be regarded as a particular *process* rather than a kind of *product* (Green et al., 2023), and while AI can be both original and efficient, it lacks other constituents for creative processes to unfold, such as intrinsic motivation, intentionality, authenticity, and problem finding (Kharkhurin, 2014; Runco, 2023).

Whether we vouch for Lovelace's test of machine intelligence and require the program to originate ideas to earn the epithet "intelligent" or accede to Alan Turing's predictions of machines altering their own instructions to rival an intellectual human class (1950), how we understand the concept of "intelligence" is key. Furthermore, Turing's own famous test of computer intelligence – known as The Imitation Game (ibid.) – is far from unproblematic: Trying to fool a human that your output represents a human mind is simply a test of human gullibility.

Indirect communication and the artistic production of metaphors

Art provides opportunities for indirect communication, enabling the transference of the morally and politically transgressive through fictional characters and situations, as well as the rendering of historical and unfashionable expressions of former generations delivered by performers and media that preserve and deliver authored output posthumously. Pseudonyms also convey indirect communication for the sake of enabling "a passionate subjectivity" (Garrett, 2012), strikingly analogous to authenticity understood as unfiltered and honest self-expression. Art often works by analogy (highlighting similarities between two things) and metaphor (leading thoughts to one thing by using another), an insight applicable to all kinds of mentalizations (visualization, auralization, and conceptualization) that the arts can stimulate by physical means (cf. Arnheim, 1974; Larson, 2012). Art as a form of cognitive play through supernormal stimuli targeting our propensity for pattern recognition enforces ideas and actions (Boyd, 2009). Analogy and metaphor work by "bisociation" (Koestler, 1964) through the mapping of patterns. And we tend to use ourselves as constituents of metaphor with our own corporeal experiences as sources when conceptualizing new target domains (Lakoff & Johnson, 1999). This is recognized as one of the key elements of art, "a duplicitous logic of representation: there is what it is or presents, and there is what it conveys only in some figurative form" (Deacon, 2006, p. 22). The cognitive mirrors that artistic patterns hold up are sensory triggers for self-reflection, we believe, which was the starting point of our project.

The artistic concept

Our artistic concept was manifested as an interactive and immersive AI-based one-to-one micro-opera for humans and machines called *The Prophecies*, performed publicly in Halmstad and Skövde in the fall of 2023. The opera was structured as modules in a series of musically paced meetings between interacting visitors and artists. Two questions instigated the work: How can artificial intelligence be used as a supplement rather than a substitute in opera? And what fictional situation can be construed to reflect metaphorically on the subject of AI by way of opera?

Training an AI agent as an "artificial lyricist" (Rylander et al., 2023), prompting a live-performing opera singer as an echoborg singing in karaoke style, was the obvious answer to the first question. The pre-composed music allowed for different arias compatible with a programmed libretto for stochastically AI-generated text to be performed by the singer. In the interactive system, the AI agent produced stylized and personalized lyrics in real time. Features to increase the probability of visceral information processing were promoted in the text generation, music composition, and theatrical setting. We leverage how humans "understand" and process verbal information not only through intellectual calculation, but also through the affective impact it has on sensorimotor activation (cf. Vergallito et al., 2019; Zwaan, 2003). Generating personally relevant answers via the AI agent supposedly increased the likelihood of experienced relevance. Moreover, some general design features were incorporated into the concept, such as raising self-awareness through personal address (Carmody & Lewis, 2006), evoking heightened resonance with ideas through the use of the word "you" (Orvell et al., 2020), and facilitating listeners' tracking of information through song by greater rhythmic regularity, fundamental frequency stability, discrete pitch intervals, and a more discernible metrical structure compared to speech (Vanden Bosch der Nederlanden et al., 2020).

Our response to the second question connects people's tendency to draw parallels between AI and divine spirits aand AI agents, conceptualized as non-natural entities with great power over human life (Spatola & Urbanska, 2020), with opera's historical preoccupation with human dependency on external forces. Operatic plots generally tend to center on individuals struggling for self-determination and taking a stand against superior powers of human, societal, or divine kinds, and until the 19th century, operatic heroes and heroines were normally rescued by intervening gods instead of relying on themselves to survive (Fend, 2020). Therefrom, tragic endings became the rule; God "died," and man had to invent another agent *in loco parentis*. Consequently, if opera says something about human belief in ourselves (self-confidence) and higher powers (faith), and if AI is viewed as a kind of divinity, this should logically be reflected in contemporary opera (Jalhed, 2024).

We are not philosophers discussing art and design, but artists and designers inspired by philosophy. As artistic researchers, we avoided explicit explanations, synopses, and commentaries to the artwork at its execution, but instead aimed at maximizing interpretational openness. Such an approach is in line with the view that art-making has co-evolved with humans' cognitive capability of theory of mind (ToM), enabling us – through controlled sensory activation – to stimulate, aggravate, and thereby train, test, and play with our ability to imagine each other's psychical conditions (Tague, 2017). Art is not a matter of direct instrumental communication per se (cf. Deacon, 2006), and while artists set the stage aiming at an end-state in the mind of the perceiver, the way the aesthetic object is experienced by the subjective individual is out of the artists' control (Donald, 2006).

Mirroring problem formulation

Searle asks: "What psychological and philosophical significance should we attach to recent efforts at *computer* simulations of *human* cognitive capacities?" (Searle, 1980, p. 417, our italics). The nature of the question itself (more than the endless responses to it) struck us as topical for artistic processing. Not primarily aiming at making a contribution to the philosophical debate, but instead creating prerequisites for a potentially affective and mind-altering experience that brings attention to the issue of "technoanimism" (cf. Aupers, 2002), we reformulated Searle's problem into a mirroring one that could inspire artistic arrangements: What psychological and philosophical significance should we attach to efforts at *human* simulations of *computer* cognitive capacities? That is, simulations of almost godlike abilities of intentionality-free calculation and vast data access that obscure any "metamagical" operations – defined as going one level beyond magic to the non-magical lying behind the apparent magic (Hofstadter, 1985).

Outline of the artwork

Putting together how opera centers on human–god relations with contemporary tendencies to treat AI as an all-knowing deity, and adding the twist of humans simulating computer-like capacities, led our thoughts to the Delphic oracle. As opera once arose from a wish to revive the features of Hellenistic drama, this also seemed fitting from a cultural heritage view, and we decided to use the operatic medium to construct a system and a setting in which visitors could meet a seeress connected to the presumed divinity of our time – an oracle in contact with the digital realm.

We framed the initial idea for the opera as a thought experiment, here put side-by-side with Searle's wordings (1980, pp. 417–418):

The Chinese Room	The Delphic Room
Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles.	Suppose that you go into a room. In the room is another person, isolated from the outside world but supposedly in contact with supernatural forces. Suppose that you present yourself with your name and ask a personal question to the person.
Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that "formal" means here is that I can identify the symbols entirely by their shapes.	Now suppose that the person in the room begins to speak. Some utterings are mere gibberish, but some distinguishable references can be picked up. Depending on how much relevant information you can pick up, you consider the answer to be credible and useful.
Suppose I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch.	Suppose you approach the room again. Now, there is another person outside the room, asking you to write down your name and the question. After submitting the note to this person, a musical tune begins. You then enter the room with its isolated inhabitant.
Now just to complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view - that is, from the point of view of somebody outside the room in which I am locked - my answers to the questions are absolutely indistinguishable from those of native Chinese speakers.	Now suppose that the isolated person greets you by your name and repeats your question as if in telepathic contact with the first person you met. When your question is answered, the sentences are syntactically coherent and no gibberish occurs. Suppose that the semantic references are very precise. Suppose that general propositions are delivered with personal address – as if the person knows about you. Moreover, suppose that the verbal answer fits metrically with the tune played, so it gives the impression of not being spontaneous, but being written and rehearsed before you even had formulated the question – as if predicting your visit.

Table 1: The Chinese Room compared to The Delphic Room.

The Chinese Room is a critical metaphor for the problem of computer simulations of human intelligence neglecting the role of intentionality. The Delphic Room became our metaphor for the problem of human simulations of computer intelligence neglecting the role of authenticity. The oracle in the ultimate, operatic stage delivered computational results devoid of authenticity, as the digital deity informing her utterances had no personal aptitudes and appetites constituting a complex we can call a "self." Partly unburdened by the efficacy, speed, and lexical richness of the computer program, she was relieved of cognitive load and even responsibility. The technoanimistic messages were personalized and dynamic, and all non-verbal signs of human character enabled the artificially produced information to pass as intentional while the operatic character could be associated with the supernatural.

Reflective comments

One difference in relation to Searle's thought experiment was that the AI agent became just one "room" in the line of many in a row of agentive interactions with possibilities for both verbal and non-verbal communication. Another difference was that the visitor did not only get a written answer. Instead, the answer was delivered by the operatic echoborg who read the prompt, and misread, filled in blanks, and corrected mistakes made by the AI agent. During this process, she could occasionally add intelligence, including creativity, in the process of extending AI with human activity. It is easy to get stuck in the discussion about AI as if it constituted the main artistic contribution in *The Prophecies*, while it was just one cog in a greater agentive complex. The AI agent was but "a man in a Chinese Room," a machine learning model with no understanding built into its (albeit capable) correlation between inputs and outputs. Initially, it was not very good at following instructions and returning responses, but with improvement, its output became less dependent on human adjustment.

The setup that we created can still be used as a version of the Turing test, since it tests humans, not the AI agent. The challenge is about human capability to separate humans from machines, despite our anthropomorphic biases (cf. Eagleman, 2023). In sum, the artwork reflects opera's capacity to develop in accordance with new technology and contemporary operatic displays of human longing for self-transcendence.

Concluding remarks

Opera as an *artform* has offered symbolic representations of humans-as-machines as well as humans-among-machines. This "technoromantic" (Barron, 1997) figure frames the corporeal human body as intertwined with technological development. Operas as *artworks* have furthermore been preoccupied with the human inability to solve existential problems and the need for supernatural aid. Our opera arose from the idea of reconnecting with this tradition and creating a blend with contemporary relevance, with AI in an animistic cameo role.

If The Chinese Room is a metaphor for calculation without intention, The Delphic Room is a metaphor for interpretation of intention equivalent to ToM – parsing all information as if coming from or concerning an embodied self. Intentionality in terms of ToM has supposedly produced animistic superstition as an evolutionary by-product; an early and basic mechanism for religion to build from (cf. Dunbar, 2003; Peoples et al., 2016). Artists have always exploited this, but art has also provided opportunities for reflections upon these very impulses. It has been argued that AI development must include aspects of so-called "hot cognition," for instance ToM capabilities, and that designers and researchers within the field need to collaborate with psychiatrists, psychologists, and neuroscientists (Cuzzolin et al., 2020). We suggest that artists could also contribute to such undertakings with skill and knowledge about some principles of intentionality.

The god *from* the machine interfering in action is an operatic trope. The god *in* the machine instructing action is an opera ghost. What makes AI eerie is not its supreme industrial efficacy, but our own evolutionary programmed expectations of intentionality where there is none. We see patterns and jump to conclusions because of our adaptations, but the fact that we seem to imagine the same thing over and over again – that is, humanoid superiority – says more about our need to expand the creative range of our imagination than our need for technical assistance.

Acknowledgements

We would like to thank our colleagues and collaborators who have contributed to and facilitated the production and presentations of *The Prophecies*: artists and performers David Hornwall, Agnes Wästfelt, Katariina Poikela, and Hadi Ali, as well as the team at PlayLab hosting the LudoKonst-event in Skövde represented by Lars Kristensen, Rebecca Rouse, and Therese Ramstedt.

References

Arnheim, R. (1974). *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press.

Aupers, S. (2002). The Revenge of the Machines: On Modernity, Digital Technology and Animism. *Asian Journal of Social Science*, 30(2), 199–220.

Barron, S. (1997). *Art planétaire et romantisme techno-écologique*. Université Paris 8.

Boyd, B. (2009). *On the Origin of Stories: Evolution, Cognition, and Fiction*. Harvard University Press.

Carmody, D. P., & Lewis, M. (2006). Brain Activation When Hearing One's Own and Others' Names. *Brain Research*, 1116, 153–158.

Chomsky, N. (2023, March 8). The False Promise of ChatGPT. *The New York Times*.

Cuzzolin, F., Morelli, A., Cîstea, B., & Sahakian, B. J. (2020). Knowing me, knowing you: Theory of mind in AI. *Psychological Medicine*, 50, 1057–1061.

Deacon, T. (2006). The Aesthetic Faculty. In M. Turner (Ed.), *The Artful Mind: Cognitive Science and the Riddle of Human Creativity* (pp. 21–53). Oxford University Press.

Donald, M. (2006). *Art and Cognitive Evolution. In M. Turner (Ed.), The Artful Mind: Cognitive Science and the Riddle of Human Creativity* (pp. 3–20). Oxford University Press.

Dunbar, R. I. M. (2003). The Social Brain: Mind, Language, and Society in Evolutionary Perspective. *Annual Review of Anthropology*, 32, 163–181.

Eagleman, D. (2023). *A Proposed test for Human-Level Intelligence in AI*.

Fend, M. (2020). Opera. In T. McAuley, N. Nielsen, J. Levinson, & A. Phillips-Hutton (Eds.), *The Oxford Handbook of Western Music and Philosophy* (pp. 600–627). Oxford University Press.

Frigau Manning, C. (2013). Singer-Machines: Describing Italian Singers, 1800–1850. *The Opera Quarterly*, 28(3–4), 230–258.

Garrett, E. (2012). The Essential Secret of Indirect Communication. *The Review of Communication*, 12(4), 331–345.

Green, A. E., Beaty, R. E., Kenett, Y. N., & Kaufman, J. C. (2023). The Process Definition of Creativity. *Creativity Research Journal*. https://doi.org/10.1080/10400419.2023.2254573

Hofstadter, D. R. (1985). *Metamagical Themas: Questing for the Essence of Mind and Pattern*. Basic Books, Inc.

Jalhed, H. (2024, February 27). *Deus in machina: Integrating artificial intelligence into new music drama*. LTTA2 Learning Teaching Training Activity: MUSic higher EducatioN meetS the cyber dimension, Malmö. https://musense.eu/en/activities/18/

Kharkhurin, A. V. (2014). Creativity.4in1: Four-Criterion Construct of Creativity. *Creativity Research Journal*, 26(3), 338–352.

Koestler, A. (1964). The Act of Creation. Hutchinson & Co.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.

Larson, S. (2012). *Musical Forces: Motion, Metaphor, and Meaning in Music*. Indiana University Press.

Orvell, A., Kross, E., & Gelman, S. A. (2020). "You"speaks to me: Effects of generic-you in creating resonance between people and ideas. *PNAS*, 117(49).

Peoples, H. C., Duda, P., & Marlowe, F. W. (2016). Hunter-Gatherers and the Origins of Religion. *Hum Nat*., 27.

Popper, K. (1963). *Conjectures and Refutations*. Routledge.

Ridout, N. (2012). Opera and the technologies of theatrical production. In N. Till (Ed.), *The Cambridge Companion to Opera Studies* (pp. 159–176).

Runco, M. A. (2023). AI can only produce artificial creativity. *Journal of Creativity*, 33.

Rylander, M., Jalhed, H., & Åberg, K. (2023). The Artificial Lyricist: Prototyping an Interactive Opera for Humans and Machines. In S. Ferraris, V. Rognoli, & N. Nimkulrat (Eds.), *From Abstractness to Concreteness: Experiential Knowledge and the Role of Prototypes in Design Research* (pp. 831–847).

Searle, J. R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3, 417–457.

Spatola, N., & Urbanska, K. (2020). God-like robots: The semantic overlap between representation of divine and artificial entities. *AI & Society*, 35, 329–314.

Sterne, J. (2016). Afterword: Opera, Media, Technicity. In K. Henson (Ed.), *Technology and the Diva: Sopranos, Opera, and media from Romanticism to the Digital Age* (pp. 159–164). Cambridge University Press.

Tague, G. F. (2017). *Art and Adaptability: Consciousness and Cognitive Culture*. Brill.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460.

Vanden Bosch der Nederlanden, C., Joanisse, M. F., & Grahn, J. A. (2020). Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. *NeuroImage*, 214.

Vergallito, A., Petilli, M. P., Cattaneo, L., & Marelli, M. (2019). Somatic and visceral effects of word valence, arousal and concreteness in a continuum lexical space. *Scientific Reports*, 9.

Zwaan, R. A. (2003). The Immersed Experiencer: Toward an Embodied Theory of Language Comprehension. *Psychology of Learning and Motivation*, 44, 35–62.