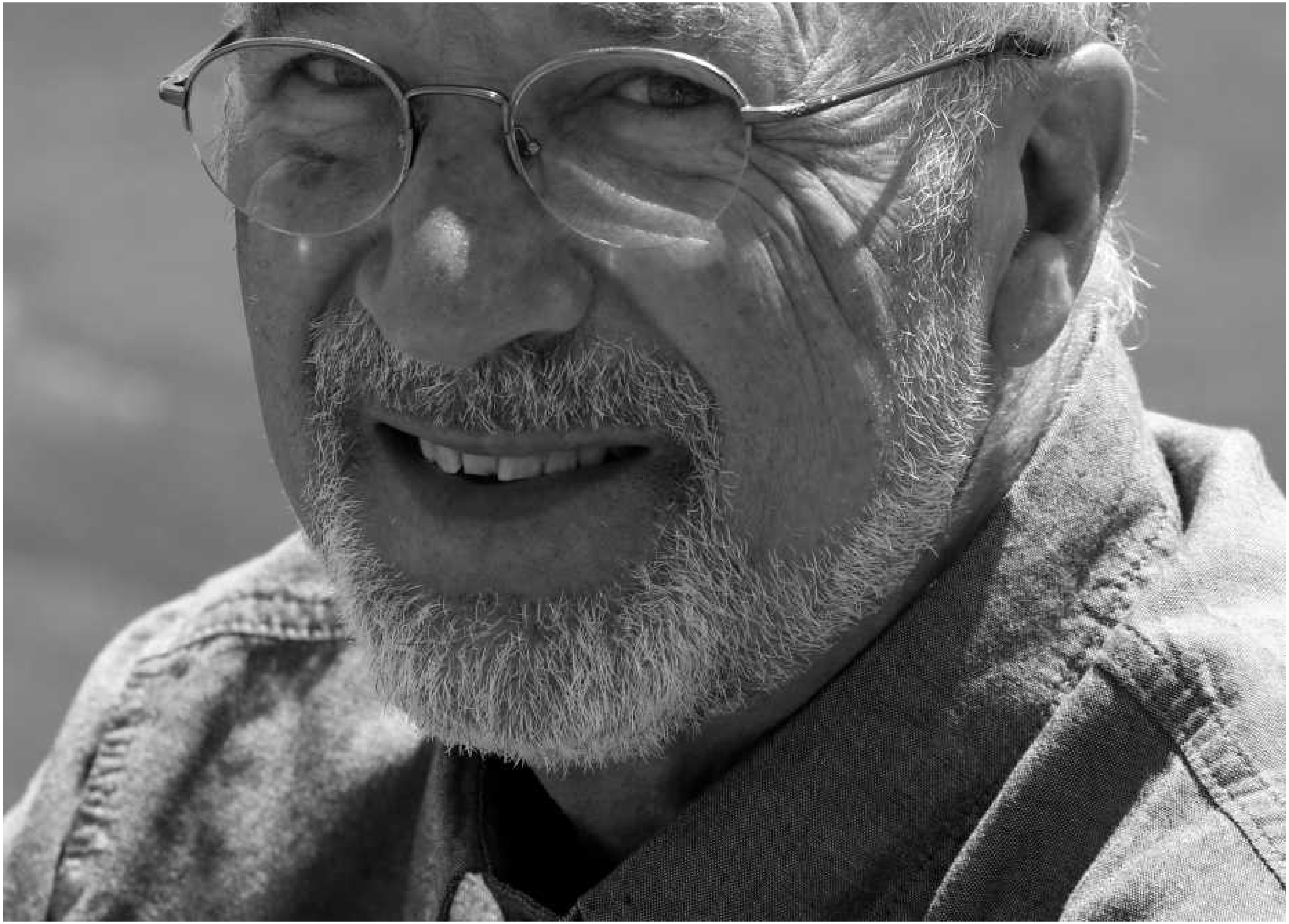
#3 - Would a Turing Test be appropriate for validating odor reproduction? - D. Harel

THE TURING CONVERSATION



David Harel, The Weizmann Institute of Science

Would a Turing Test be appropriate for validating odor reproduction?

By David Harel
The Weizmann Institute of Science

In his famous 1950 paper, decades before the existence of anything resembling artificial intelligence, Turing addressed the question of how to test whether machines can think, or in modern terminology, whether a computer claimed to exhibit intelligence indeed does so [5]. It is interesting to raise the analogous (though far more modest...) issue for olfaction: how to test the validity of a system claimed to reproduce arbitrary odors artificially, in a way recognizable to humans. Although odor reproduction systems are still far from being viable, the question of how to test candidates thereof is interesting and nontrivial.[1]

First, however, some comments on odor reproduction per se. Despite olfaction having attracted a tremendous amount of deep research over the last few decades, especially since Buck and Axel's discovery of the multigene odor receptor family [1], what is known about the sense of smell is but the tip of the iceberg. In particular, we are still very far from achieving the holy grail of the field, *artificial olfactory reproduction*: the ability to record and remotely produce recognizable renditions of arbitrary odors. In contract to the waves of sight and sound, odors, come in the form of actual molecules, and little is known about the way our brains form our perception of odors. Hence, analyzing and synthesizing smell cannot be done by computing and producing wavelengths.

In direct analogy with, e.g., a digital camera and a printer, we shall consider an *odor reproduction system* (ORS) to consist of (i) an input device, the *sniffer*, which captures and transforms any odor into a digital signature, or fingerprint; (ii) an output device, the *whiffer*, which contains a palette of fixed odorants with means for mixing them at high resolution and releasing the mixture to the outside world in carefully measured quantities and concentration, and with precise timing; and, most significantly, (iii) a *mix-to-mimic algorithm*, which analyzes the signature coming from the sniffer and instructs the whiffer as to how it should mix its palette odorants in order to produce an output odor, which, as perceived by a human, is as close as possible to the original input [4].

I shall not attempt to discuss here the feasibility of constructing an ORS, but such a system will have numerous applications in a variety of industries, medical practices, infotainment, etc. Instead, I will assume that we are presented with a black-box candidate ORS, and will attempt to address the question of what it is we really want, and to seek ways to test whether we really have it. Besides these being nontrivial issues, as we shall see, I feel that they represent an intellectual challenge that is a worthwhile topic for serious contemplation.

To some extent, the testing method proposed here is inspired by Turing's test for AI, in that it involves a human challenger and the real and artificial entities, yet it is very different: the test is conditional, requiring from the artificial no more than is required from the original, and it employs a novel method of immersion that takes advantage of the availability of near-perfect reproduction methods for sight and sound.

Reproduction methods for sight and sound go back to the 19th century. Nicéphore ↗Niépce ↗ produced the first recognizable photograph, in 1826 or 1827 (Fig. 1). Half a century later, Alexander Graham Bell made the first telephone call, successfully summoning his assistant from the next room. In both cases, the generated artefacts were immediately recognized as being satisfactory renditions of the originals. Not perfect, of course, but unmistakably recognizable.

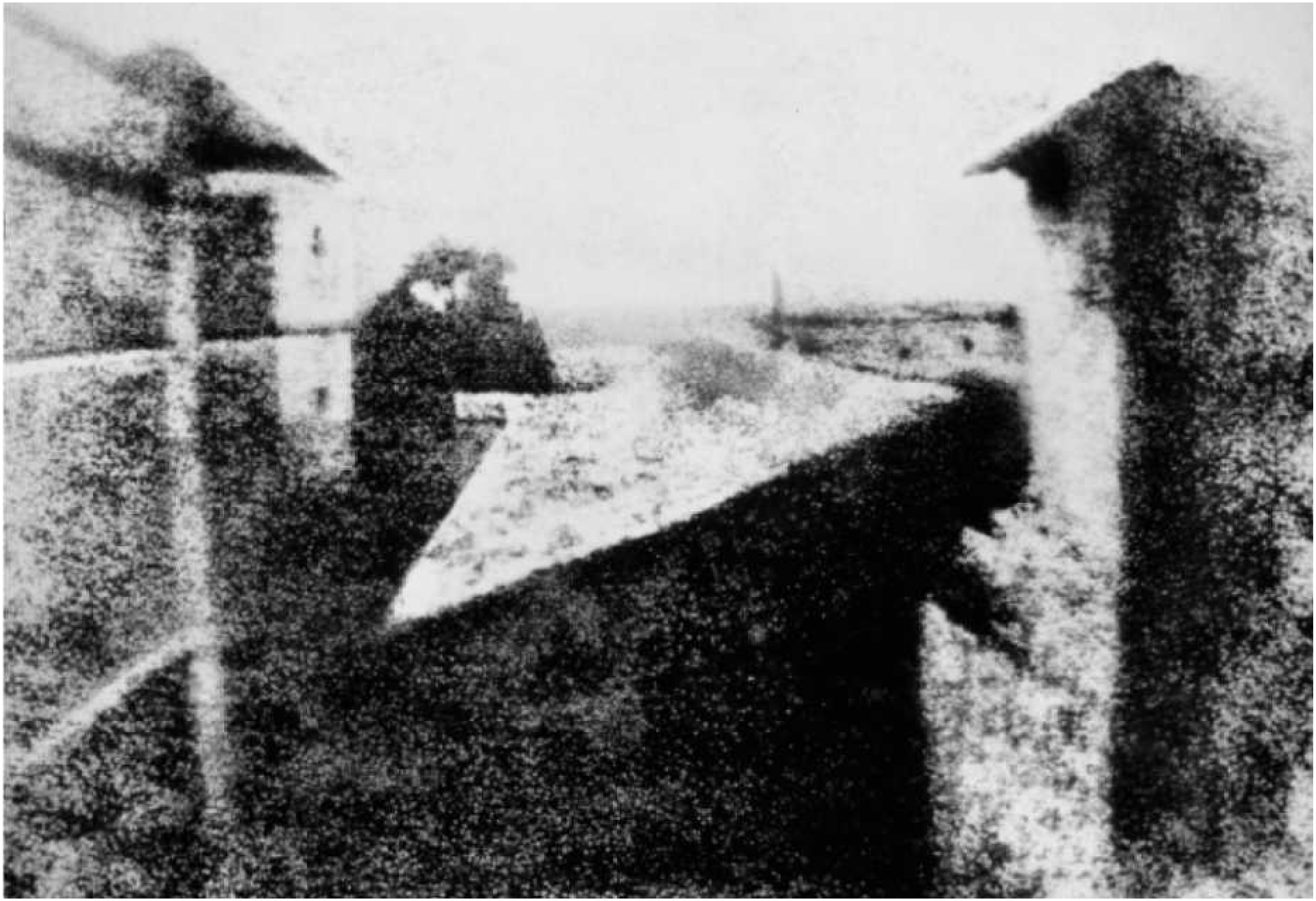


Figure 1. “View from the Window at Le Gras”, by Nicéphore Niépce (1826–7)

In contrast to recognizability, perfect reproduction would call for the produced result to be indistinguishable from the original, as per Turing’s test.[2] So, should we opt for a Niépce-Bell-style approach (human recognizes odor in an appropriate real world sense) or a Turing-style one (human cannot tell the real odor apart from the reproduced one)?

real odor apart from the reproduced one)?

We recognize and "understand" what we see in a photograph or on a video screen, and what we hear coming out of a telephone or an audio system. This has been true from the very beginning. Bell's assistant in the adjacent room understood (and immediately acted upon) the famous 1876 utterance "Mr. Watson, come here – I want to see you", and no one thought of asking whether a different sentence, such as "Mr. Watson, what is the time?" would have worked too. When we view the 1826-7 and 1838 photographs, we recognize and understand what they capture and may safely conclude that the techniques used were general enough and could have been applied to other scenes too.

This is inadequate for olfaction, for two reasons. First, we cannot make do with trying out the system on just a few odors. Perhaps for a rose, an orange and a cup of coffee the system would do a good job (because of some particular odors present in the whiffer), but not for moss in a dark cave, for screeching tires, or for the odor of an unknown animal in a faraway forest. We need to convince ourselves that the reproduction system works for all appropriate inputs. Second, how are we to become convinced that a person has "understood" an arbitrary odor, and recognizes it? No general naming method exist for arbitrary odors. Some attempts have been made to devise odor vocabularies, employing descriptors like musky, putrid, floral and ethereal, and in particular, fields like winery there are ways to approach this, but these are deficient as general methods.

We cannot opt for a pure Turing test approach either, where a human tester is asked to compare real odors with their artificially produced versions, mainly because it has little to do with the human experience of relating to odors as part of their real world experience, and thus will never capture the elusive notion of recognizability. Also, true indistinguishability is not viable even for sight and sound, and some will claim that total indistinguishability a la Turing will never be truly attained for artificial renditions of any of our senses, so this is not what we should be striving for here either.

We need to tap into the human's real-world experience, and it is in that respect that the naming issue arises in full force: verbally describing the arbitrary odors involved in a testing process is out of the question.

I suggest a general method for testing recognizable odor reproduction, which involves a subtle combination of the human olfactory experience and a comparison of the real and the reproduced. The idea is to avoid the need to name or to verbally describe odors, by employing a multi-modal immersion approach, taking advantage of the fact that we already have excellent reproduction methods for sight and sound.

The test involves the candidate system for olfactory reproduction and two teams of humans. The first is the *challenger* team, whose role is to challenge the ORS's claim to validity, and who can be thought of as representing the eventual users of the ORS, if and when it passes the test. The second is an honest *tester* team, whose members are willing to spend time on this, but who has no vested interest in the test's results either way. I first describe a somewhat naïve test, followed by the more subtle recommended one. Both variants call for the challenger to provide the tester, repeatedly, with small sets of odor-emitting situations, or scenes, in the form of short clips recorded using a video camera (which includes audio) coupled with the ORS's sniffer; each say, 8-10 seconds long. Every testing session thus employs a set of audio-video recordings and recordings of the corresponding odor fingerprints. The clips can be prepared in whatever locations the challenger fancies: a bustling marketplace, a damp cave, the lion's cage in a zoo, a grandparent's attic, the depths of a jungle, a hospital's operating room, or the set of a TV cooking show.

The testing involves several sessions, in which the testers are asked to make certain decisions regarding the video clips and the odors. When implementing the tests, the number of testing sessions is important, and care should be taken to devise means for preventing lucky guessing on the part of the testers, and for eliminating outliers. Provisions should be taken regarding the output devices (e.g., the audio-video projection and sniffer emission), which allow the testers to view or sniff any of these, as often as they want, at any time during the testing session.

Our use of audio and video makes it possible to immerse a human in familiar sensory information for reference, in effect "placing" the tester where the odor was captured. The fact that there is no need for verbal characterization also helps reduce the effect of any relevant cultural differences that may exist; as long as the challengers choose situations of which the testers can make real-world sense, it does not matter how a particular person perceives ist: as long as the challengers choose situations of which the testers can make real-world sense, it does not matter how a particular person perceives odors. The test must only verify that the whiffer-generated output adequately captures the original input odor, in a way that substantiates its recognizability in the video/audio setting by the human testers, regardless of how different people would have chosen to describe it, or even whether they could have done so at all.

Here is our first version of the test.

> *Simple lineup*: In each session, testers are given a small fixed number of challenger-produced video clips – say, between 5 and 10 – but are given the whiffer output of only one of them, without knowing which of the clips constitutes its origin. The testers' role is to try to match the odor to the correct clip.

This simple lineup matching can be viewed as a straightforward human recognition task, a la Niépce ↗ and Bell, but one that avoids the need to name or describe what is being smelled; all the tester has to do is decide which of the video clips is most likely to have produced the odor he or she is given. Repeated success on a variety of sets of clips supplied by the challenger validates the ORS.

There is, however, a rather serious problem with this naïve test. The challenger could be too eager to disqualify the ORS, producing sets of situations that are very much alike or ones that require special expertise on the part of an average tester, such as different glasses of wine with their labelled bottles nearby. Worse, the challenger may be downright vicious, recording, for example, a peaceful countryside, but with the video camera's back facing the opening of a damp cave. The on-location sniffer will capture the strong odor of the cave, which is nowhere to be seen in the clip, guaranteeing the test's failure.

To eliminate these difficulties, the actual test I propose is more subtle. It involves the challenger producing an additional channel of captured input for each situation, besides the audio/video clip and the sniffer recording. An actual odor sample is to be collected at each recorded location, in a way that enables future release. There are several viable techniques for doing this, although they limit the duration for which the collected molecules are able to remain perceptually faithful to their capture time, but better techniques, with increased longevity of the samples, will probably surface in the future.

The proposed test is a conditional, asymmetric lineup.

> *Conditional lineup*: In each session, the testers are given a 5-10 challenger-produced video clips and the odor corresponding to only one of them. However, there are now two tester teams, where, unbeknown to them, the first is given the actual collected odor sample and the second is given the whiffer output corresponding to the same clip. The goal is for the second team to succeed in the matching whenever the first one does. Sessions for which members of the first team provide an incorrect match are ignored.

The conditional requirement, whereby we require success from the tester with the artificial odor only when the one with the original odor succeeds, balances the need for rigorous testing with the hard-to-define power of human odor perception. If humans are unable to "recognize" the real odor when immersed in its audio-video habitat, we cannot require them to be able to do so with the reproduced one. An ORS does its job well if humans are able to recognize the artificial whenever they can recognize the real thing.

The conditional lineup test is thus heavily inspired by the way humans recognize reproduced photos and audio, ever since the early days of Niépce ↗ and Bell, but it also indirectly checks the imitation facet of the ORS, in that the similarity of the real and the artificial are both considered, so that it is inspired also by Turing. And the two main difficulties are avoided: employing video and audio immersion gets around the naming and verbal description problem, and unfair challenges are avoided by never requiring of the reproduced odor what we do not require of the original. Some weaknesses of the testing method are discussed in detail in the paper.

Full odor reproduction systems, which deal adequately with any input odor, might be long in coming, but I believe we will begin to see initial attempts quite soon. Just as technologies for the reproduction of sight and sound have changed and improved radically since the pioneering work of Niépce ↗ and Bell, so will future years see radically new ways of capturing, communicating and reproducing odor in ways recognizable by humans. The 2016 paper merely

so will future years see radically new ways of capturing, communicating and reproducing odor in ways recognizable by humans. The 2016 paper merely proposes a criterion for assessing the quality of such systems when they do arrive, which, I think, is important in its own right.
I am hopeful that this work will trigger further thinking about the extremely difficult, but exciting problem of achieving satisfactory artificial olfactory reproduction, hand in hand with developing the best methods for testing the solutions.

[1] This blog is based on my 2016 paper “Niépce-Bell or Turing: How to Test Odor Reproduction?" [3].
[2] As an aside, a variant of the Turing test has been proposed for checking the validity of computerized models of biological systems and other artefacts from nature, where an expert challenger uses probes to try to distinguish the claimed-to-be-valid model from the real thing – say, in an appropriate laboratory [2].

References

[1] L. Buck and R. Axel, “A novel multigene family may encode odorant receptors: a molecular basis for odor recognition”, Cell 65:1 (1991), 175-187.

[2] D. Harel, “A Turing-Like Test for Biological Modeling", Nature Biotechnology 23 (2005), 495-496.

[3] D. Harel, "Niépce-Bell or Turing: How to Test Odor Reproduction?", J. Royal Society Interface 13:125 (2016), DOI: 10.1098/rsif.2016.0834.

[4] D. Harel, L. Carmel and D. Lancet, “Towards an Odor Communication System”, Computational Biology and Chemistry (formerly Computers & Chemistry) 27 (2003), 121-133.

[5] A.M. Turing, “Computing machinery and intelligence”, Mind 59 (1950), 433-460.

Related Content

#1 - Strong meat with strawberries - J. D. Turing

#2 - Alan Turing and evil AI - D. Proudfoot

#4 - Of Turing's teddy bear - M. Boden

#5 - Whole-brain simulation, cryptography, and Turing's mystery machine - J. Copeland & D. Proudfoot

The Turing Conversation