

SHORT AND SWEET**Space-time disarray and visual awareness****Jan Koenderink**

University of Leuven (K.U. Leuven), Laboratory of Experimental Psychology, Tiensestraat 102-box 3711, BE-3000 Leuven, Belgium; and Delft University of Technology, EEMCS, MMI, Mekelweg 4, NL-2628 CD Delft, The Netherlands; e-mail: jan.koenderink@telfort.nl;

Whitman Richards

MIT Computer Science and Artificial Intelligence Laboratory, The Stata Center, Building 32, 32 Vassar Street, Cambridge, MA 02139, USA; e-mail: wrichards@mit.edu;

Andrea J van Doorn

Delft University of Technology, Industrial Design, Landbergstraat 15, NL-2628 CE Delft, The Netherlands; e-mail: a.j.vandoom@tudelft.nl;

Received 10 November 2011, in revised form 4 February 2011; published online 14 March 2012.

Abstract. Local space-time scrambling of optical data leads to violent jerks and dislocations. On masking these, visual awareness of the scene becomes cohesive, with dislocations discounted as amodally occluding foreground. Such cohesive space-time of awareness is technically illusory because ground truth is jumbled whereas awareness is coherent. Apparently the visual field is a construction rather than a (veridical) perception.

Keywords: visual awareness, space-time, local sign, causality, amodal occlusion, specious moment.

Is the space-time of awareness a pre-established container, waiting to be filled with visual experiences? Or is it created along with such experiences, as a structural aspect of them? This reminds one of the famous Leibniz-Newton controversy in physics (Clarke 1717). The issue was whether space is an empty “container” (Newton’s absolute space-time: “Absolute space, in its own nature, without regard to anything external, remains always similar and immovable” (Newton 1687, Scholium II).), or whether “space” is nothing beyond a relation between objects (Leibniz in Clarke 1717: “... all we need in order to have an idea of place (and consequently of space) is to consider these relations amongst things and the rules of their changes; we do not need to imagine any absolute reality beyond the things whose location we are considering.”). In the latter case it would make no sense to speak of “empty space”. Geometry would be about relations between actual objects. The outcome (after various surprising changes of perspective) is still debated.

We consider an analogous problem in awareness. Visual space-time is commonly understood as (close to) veridical (Helmholtz 1867) representation of Newtonian space-time, requiring little explanation. This is perhaps the reason why Lotze’s (1852) concept of “local sign” or Michotte’s (1962) concept of apparent causality have been largely disregarded. Lotze required a physiological explanation for visual location; he considered a mere reference to the container concept unsatisfactory. Michotte showed that causality may be perceived where none exists in the physical scene. He thus showed that causality is a construction of the mind on the basis of spatio-temporally structured optical patterns. Thus the notion that perceptual space-time is a mere representation of physical space-time is perhaps suspect.

Our empirical approach to the question is to scramble physical space-time. Then “veridical perception” is a scrambled mess. We show that the space-time of visual awareness is often coherent in such cases. Thus, mental space-time is not a veridical representation of physical space-time at all, but a Leibnizian, relational structure of strands of awareness.

In [Figure 1](#) the strips have been sloppily assembled, yet the reassembled image looks reasonably cohesive (try screening off the upper and lower ragged boundaries). This has struck many authors of books on visual arts or photography (eg, Clifton 1973). In terms of experimental phenomenology (Metzger 1930) the presentation is cohesive, whereas scrutiny reveals dislocations of edges. In

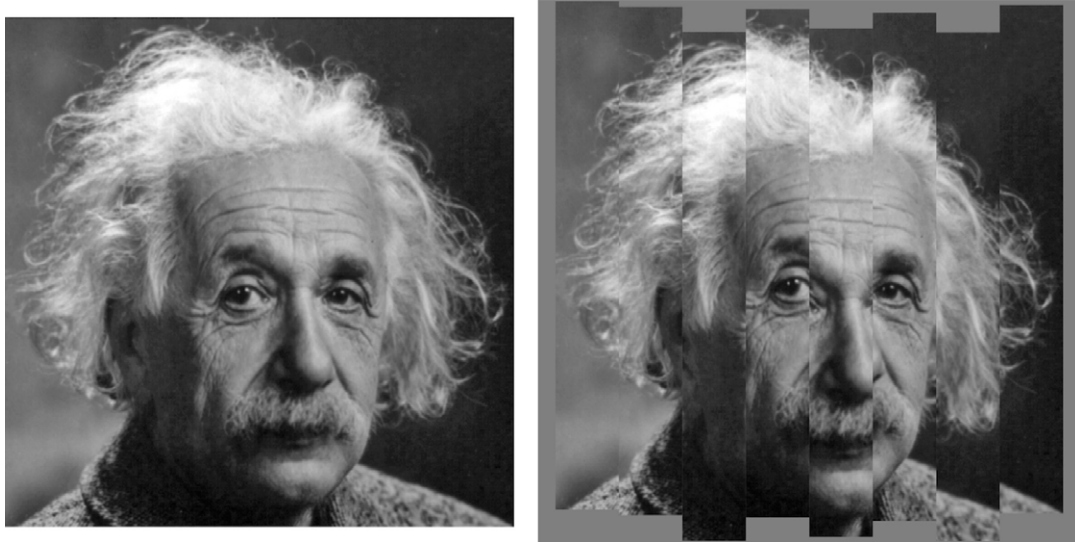


Figure 1. The original image (left) was cut into vertical strips. These strips are sloppily assembled at right.

the laboratory one forces “immediate visual awareness” through limiting viewing time, eccentric presentation, diverting reflective thought, and so forth (Ihde [1986](#)).

We extend such observations to local disarray in space, time, and space-time. Purely spatial cases are illustrated with figures, whereas spatio-temporal cases require movie clips.

Consider local spatial disarray. Use a rectangular array of apertures as windows on randomly displaced independent copies of the image. Lacking data is filled with white. The displacements are about a quarter of aperture size. Local dislocations are “hidden” at the edges through cracks between the apertures ([Figure 2](#)). Even large dislocations (up to half the tile size) are “visually acceptable”.

Disarray is apparent under scrutiny (in [Figure 2](#) notice the dislocation of the nose); it disappears under mild eccentric fixation. Even serious disarray is not salient in immediate awareness.

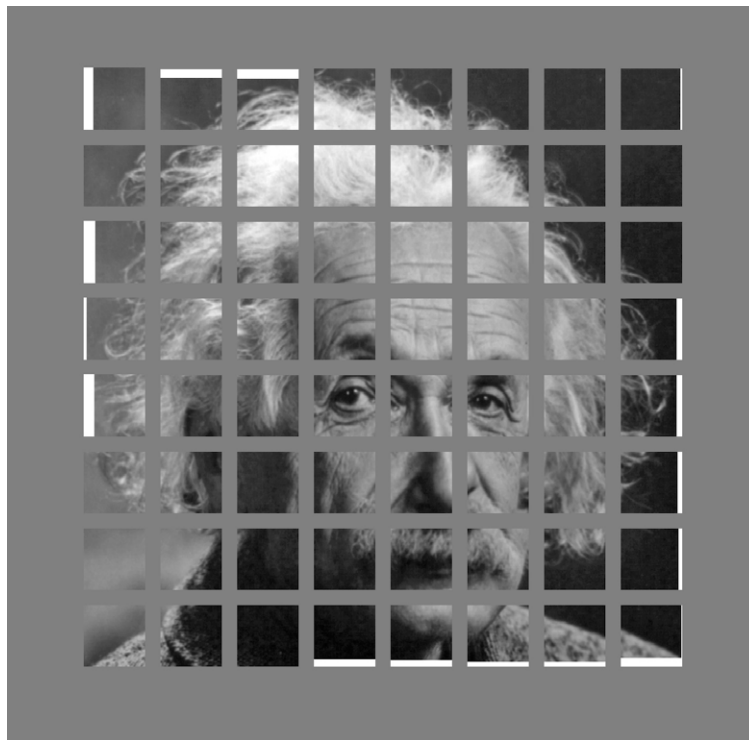
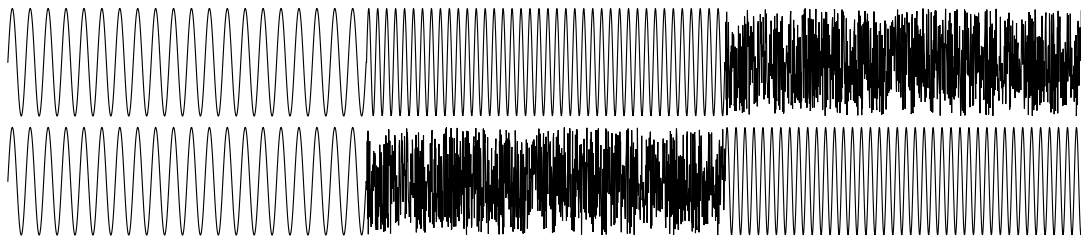


Figure 2. A tiled image with rather large random displacements within the tiles (used as apertures). The cracks appear as a grid, amodally occluding a single image instead of being part of it.

All instances of disarray are different, noticeable when you present them in quick succession. “Temporal cracks”, short flashes of a uniform gray image between two presentations, kill the apparent motion (Rensink et al [1997](#)). Then vision relies on purely spatial structure ([Movie 1](#)).



Movie 1. The demo is based on a painting by van Gogh “Wheat field with cypresses”. There are five parts. 1: The painting is shown without any intervention. 2: The painting is tiled. Each tile is filled with a randomly shifted copy of what “should be there”. Random shifts are drawn anew for each frame. Notice the turmoil which looks much like a continuous deformation. Occasionally one spots the edge of a tile, though this requires some scrutiny. 3: Like 2, but here we introduced “cracks” between the tiles. The impression is not that different from that in 2, although the movements seem confined to the tiles. Occasionally one believes to see the tiles themselves moving (which they don’t). 4: As 2, but here we introduced flashes between the frames. Notice that the impression of a continuous turmoil is gone. One notices occasional dislocations between the tiles. 5: As 4, but here we have both the cracks and the flashes. The major impression is that of a coherent image. Some scrutiny reveals occasional dislocations. The contrast with the impression of turmoil as in 2 is very striking. Please click to play. (A higher quality clip is available for download on the *i-Perception* website.)



Sound clip 1. The clip presents two sounds with a longish pause in between. In the first presentation you get two pure tones, a low one followed by a higher one (“dah-di”), followed by noise (a buzzing sound like “bzz”). The graph at top shows the sound amplitude as a function of time. (Period 300ms, sampling frequency 10kHz.) What you hear is the expected “dah-di-bzz”. In the next presentation (after a 1000ms pause) the presentation is “dah-bzz-di”. The sound amplitude as a function of time is shown in the graph at bottom. What you will hear is more like “dah-di-bzz” though, apparently the temporal order was rearranged in your awareness. For best effect you should listen to the pair various times, carefully comparing the second to the first. Please click to play.

Without flashes, one sees a turmoil of smooth random movements, like a flood bed seen through the rippling water surface of a shallow stream. With flashes, one enjoys a steady presentation. Scrutiny reveals occasional dislocations, but rather large disarray easily goes unnoticed. The effect is quite striking.

Benussi’s demonstrations in acoustics (Albertazzi [1999](#)) suggest similar effects for the temporal domain. We illustrate this with [Sound Clip 1](#). In the first presentation you hear a sequence of a low tone, a high tone, and a noise burst (“dah-di-bzz”). After a period of silence you are presented with the low tone, the noise burst and the high tone in that sequence (“dah-bzz-di”). However, what you hear is “dah-di-bzz”. The sequence is reordered in your acoustic awareness, “dah-di” being a sensible Gestalt.

In the visual domain, consider a video sequence free of “scene cuts”, and shift “apertures” of image frames randomly towards future or past. Such a movie looks jerky, due to the sudden dislocations. Use



Movie 2. This clip is based on a short sequence from Sam Peckinpah's movie "The Wild Bunch". In this scene ("LET'S GO!") the men check their guns, and start walking towards the final shoot-out. They understand it will mean their end. Notice that the clip is free of scene cuts. Although the camera (and the players) move, one has a continuous view. There is plenty of movement, except for a short break in the middle, where the men line up just before walking off towards the left. There are five parts. 1: The scene straight from the movie. 2: The same scene in locally temporal disarray. Notice the obvious "jerks" as the clip suddenly shifts towards past or future. 3: Same as 2, except for flashes between the temporal shifts. The flashes mask the apparent movements. The clip apparently runs smoothly, although the periodic flashes are objectionable. 4: Here we introduce spatiotemporal disarray. The image is tiled, and in the tiles we randomly shift the image both in space and in time. Apart from the "jerks", we introduce random shifts. This looks really bad! 5: Like 4, but we add both cracks and flashes ("temporal cracks"). This part should be compared with 4 and 1. Does it look more like 1 than like 4 (except for the cracks and flashes)? Most people we tried certainly think so. This is surprising! Because we used rather extreme disarray, scrutiny reveals a certain degree of incoherence (mostly dislocations). One experiences some turmoil and occasional dislocations. Yet judge: does it look more like 1 (disregard cracks and flashes) or like 4? Applying a lesser amount disarray would most likely yield examples that would not really look different from 1. In this short paper we can't provide a full parametric study though. Please click to play. (A higher quality clip is available for download on the *i-Perception* website.)

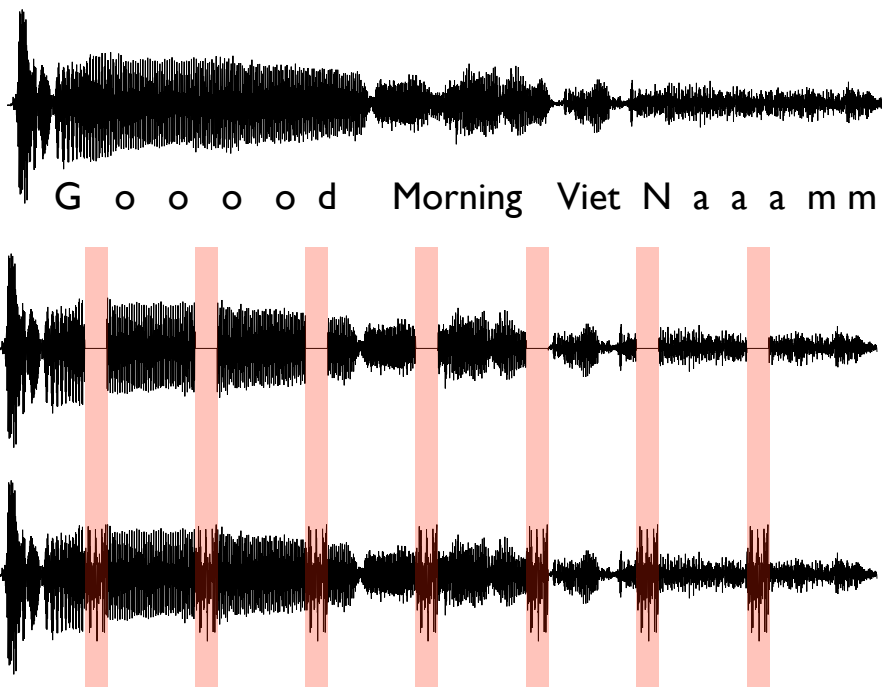
temporal cracks to hide these, and the movie appears smooth. The movie progresses steadily; jerks are gone ([Movie 2](#)). Immediate visual awareness deals gracefully with disarray in physical space and time alike.

Next consider disarray in space *and* time. Take a video sequence and tile all frames in the same way. Also "tile" in the temporal domain. (As before, this involves grouping sets of consecutive frames.) Apply both spatial and temporal disarray to each tile separately. The spatio-temporally disarrayed movie looks horrible, with violent local dislocations and strong jerks. Cracks and flashes ("space-time cracks"; [Movie 2](#)) yield an acceptable movie without obvious dislocations or jerks. Scrutiny slowly reveals many and major inconsistencies.

The cracks spoil the pleasure in viewing the movie. One sees a grid occluding the movie and a series of flashes added to it, much like lightning in a landscape. Visual awareness "does not blame" the movie for these pesky elements: it blames them on some unknown external cause. The movie appears as an integral entity seen behind, or through, the perturbations. The effect is stunning.

Neither does this experience stand alone; it works as well in the acoustic (time) domain ([Sound Clip 2](#)) and is similar to Bregman's (1994) well known "occluded BB...s" ([Movie 3](#)). Spatio-temporal cohesion is a construction of microgenesis (Brown 2002), just as the content of awareness is. Cohesion in spite of jumbled optical structure implies that it is "illusory", in the sense of mis-representing the physical (optical) data. Reality is a construction such that awareness makes better sense than the ground truth!

Microgenesis *imposes* coherent space-time and causality, rather than "represents" physical space-time and reality. The space-time of awareness is evidently Leibnizian, rather than Newtonian. It is



Sound Clip 2. The basic sound is Robert Williams famous “Gooood Morning Vietnaamm!!”, from Barry Levinson’s movie *Good Morning, Vietnam* (1987). It is repeated thrice, with one second pauses in between. You will hear: the straight sound, the sound with periodic interruptions, the sound with the interruptions filled with noise. The interruptions interfere with the intelligibility of the speech, but when filled with noise the speech flows as if not interrupted, “behind” the noise cracks so to speak. The track at top show the original sound. The reddish bars show the occurrences of pauses (center track) or noise bursts (bottom track). Please click to play.



Movie 3. The movie shows four versions of the same picture. 1: The original picture. It is composed of a number of easily legible words, written in “hollow” type. 2: In the second picture the letters are masked by strips. In this rendering we introduced additional contours by “closing” the fractional letters. This is the kind of rendering that one often finds in the literature. 3: In the third picture the letters are again masked, but the outlines of the parts are not closed. It is easier to read the text than in the case of the second picture. 4: in the fourth picture the maskers are revealed as gray bars. These are quite easily “discounted” in visual awareness, it is as if one sees the words run on behind them. Please click to play. (A higher quality clip is available for download on the *i-Perception* website.)

nothing beyond the meaningful relations between threads of awareness. It is not a “representation” of space-time as immediately given by the (meaningless) optical structure.

Visual awareness is experience of one’s optical user interface (Hoffman [2009](#); Koenderink [2010](#)), rather than of some physical scene. This fits in seamlessly with current notions from biology (ethology, eg, Koenderink [2010](#); Lorenz [1973](#); Tinbergen [1951](#)): evolution optimizes fitness rather than veridicality.

Acknowledgments. Jan Koenderink was supported by the Methusalem program by the Flemish Government (METH/08/02), awarded to Johan Wagemans (KUL). We gratefully acknowledge the administrative support of Stephanie Poot.

References

- Albertazzi L, 1999 “The Time of Presentness. A Chapter in Positivistic and Descriptive Psychology” *Axiomathes* **10** 49–74 ◀
- Bregman A S, 1994 *Auditory Scene Analysis* (Cambridge, MA: MIT Press) ◀
- Brown J W, 2002 *Self-embodying Mind: Process, Brain Dynamics and the Conscious Present* (Barrytown, NY: Barrytown/Station Hill Press) ◀
- Clarke D D, 1717 *A Collection of Papers, which passed between the late Learned Mr. Leibnitz, and Dr. Clarke, In the Years 1715 and 1716* (London: James Knapton) ◀
- Clifton J, 1973 *The Eye of the Artist* (Westport, CT: North Light Publishers) ◀
- Helmholtz H von, 1867 *Handbuch der physiologischen Optik* (Leipzig: Voss) ◀
- Hoffman D D, 2009 “The interface theory of perception: Natural selection drives true perception to swift extinction” in *Object Categorization: Computer and Human Vision Perspectives* (Cambridge, UK: Cambridge University Press) pp 148–165 ◀
- Ihde D, 1986 *Experimental Phenomenology, An Introduction State* (New York: University of New York Press) ◀
- Koenderink J J, 2010 “Vision & Information” in *Perception beyond Inferences* (Cambridge, MA: MIT Press) pp 27–57 ◀
- Lorenz K, 1973 *Die Rückseite des Spiegels* (München: Piper Verlag) ◀
- Lotze R H, 1852 *Medicinische Psychologie oder Physiologie der Seele* (Leipzig: Weidmann’sche Buchhandlung) ◀
- Metzger W, 1930 “Optische Untersuchungen am Ganzfeld: II. Zur Phänomenologie des homogenen Ganzfelds” *Psychologische Forschung* **13** 6–29 [doi:10.1007/BF00406757](https://doi.org/10.1007/BF00406757) ◀
- Michotte A, 1962 *The Perception of Causality* (Andover, MA: Methuen) ◀
- Newton I, 1687 *Philosophiae Naturalis Principia Mathematica* (London: Josephi Streater) ◀
- Rensink R A, O’Regan J K, Clark J J, 1997 “To see or not to see: the need for attention to perceive changes in scenes” *Psychological Science* **8** 368–373 [doi:10.1111/j.1467-9280.1997.tb00427.x](https://doi.org/10.1111/j.1467-9280.1997.tb00427.x) ◀
- Tinbergen N, 1951 *The Study of Instinct* (London: Oxford Clarendon Press) ◀



Jan Koenderink (1943) studied physics, mathematics, and astronomy at Utrecht University, where he graduated in 1972. From the late 1970's he held a chair "The Physics of Man" at Utrecht University till his retirement in 2008. He presently is Research Fellow at Delft University of Technology and guest professor at the University of Leuven. He is a member of the Dutch Royal Society of Arts and Sciences and received a honorific doctorate in medicine from Leuven University. Current interests include the mathematics and psychophysics of space and form in vision, including applications in art and design.



Whitman Richards is professor at CSAIL (MIT). His main research focus has been visual perception: mechanisms and models. Beginning first with studies of early visual processing, current work is now at a very high cognitive level, with emphasis on perception as a complex system of semi-autonomous modules—roughly akin to Minsky's "Society of Mind." In the mid-seventies, his research activity was redirected after meeting David Marr. Rather than concentrating on mechanisms of vision, the emphasis changed to understanding the minimal conditions that should be satisfied for a vision system "to work." Computational studies that met Marr's criteria turned out to be major advances in vision understanding. His contributions appear in a book called "Natural Computation", which covers work in vision, hearing, and motor control.



Andrea van Doorn (1948) studied physics, mathematics, and chemistry at Utrecht University, where she did her master's in 1971. She did her PhD (at Utrecht) in 1984. She is presently at Delft University of Technology, department of Industrial Design. Current research interests are various topics in vision, communication by gestures, and soundscapes.