Seeking Empathy in Conscious Cities

Claire McAndrew¹ and Itai Palti²

¹UCL Institute for Digital Innovation in the Built Environment
²The Centric Lab, London

Abstract
The vision of the conscious city has entered the radar. It takes as its heartland, the idea of a ‘conversation’ between inhabitants, digitally imbued objects and responsive architectural fabrics at the city scale. Can advances in the internet of everything, neuroscience, AI and big data enable social opportunities in a more sentient city? This chapter considers the ethics of an architectural dialogic — bringing questions of computational neutrality and democratic participation to the fore in the design and curation of ‘intelligent architecture’.

Seeking Empathy
Seek (1969–70) also known as Blocksworld, was an installation produced by Nicholas Negroponte with the Architecture Machine Group at MIT, originally shown in the exhibition Software / Information technology: its new meaning for art¹. This 5 x 8 foot superstructure contained metal cubes that when displaced by the small population of gerbils, were straightened and/or spatially recalibrated by a robotic arm, opening up the opportunity for a dialogic and self-reconfigurable architectural form.

Almost fifty years on, the vision of the conscious city has entered the radar. It takes as its heartland the idea of a ‘conversation’ between inhabitants, digitally-imbued objects and responsive architectural fabrics at the city scale. Made possible by advances in the internet of everything, neuroscience, artificial intelligence and big data, it explores the opportunities that might come from a more sentient city. Palti and Bar (2015) suggest: ‘The conscious city considers new parameters for successful planning. It presents an opportunity to raise the intelligence of our surroundings and improve our well-being.’ They speculate on how rapid developments in data technology and behavioural science could offer the prospect of our streets alleviating ailments such as stress, anxiety and boredom by being sensitive to the pervading moods of people in different parts of the city. This chapter suggests that the conscious city when conceived as an environment of care, calls not only for empathetic sensing, but intelligence in response.

¹ Software / Information technology: its new meaning for art was curated by artist and critic Jack Burnham for the Jewish Museum in Brooklyn, New York City from 16 September to 8 November 1970 and the Smithsonian Institution in Washington, DC from 16 December 1970 to 14 February 1971.
The idea of a ‘conversation’ between cities and its citizens is not altogether new. The architect behind the Smithsonian National Museum of African American History and Culture, David Adjaye (2015), has said: ‘I believe that for architecture to be emotionally relevant to people, that there has to be a connection, there has to be a relationship, that architecture cannot be autonomous. If it’s not connected to the lives of people, the histories of people, I think there’s a problem’. Juhani Pallasmaa et al (2015: 7) suggest that ‘architectural spaces are not just lifeless stages for our activities. They guide, choreograph, and stimulate actions, interests and moods, or in the negative case, stifle and prohibit them.’

For architecture, to ‘care’, might suggest more than emotional relevance or the inciting of mood. The word empathy stems from the Greek empatheia (from em- in + pathos feeling), and today means ‘the ability to understand and share the feelings of another’ (Oxford 2016). For the conscious city, it is not enough to sense the underlying pathos. Nor is it sufficient to react without empathy. Seek, a system entrenched in responsive capabilities, was ultimately unable to ‘sense’ the aggression and sickness that eventually became rife as the gerbils wrought havoc within the supercube. For cities to be empathetic they need to sense the emotional states of the citizens.

Despite its imperfections, Seek has been described as one of the earliest examples of ‘intelligent architecture’ (Shanken 1998). The programming behind it followed a simple production rule format (If… Then…) that instructed the robot to leave the blocks alone unless one of them has been knocked or fallen out of alignment with the grid. Blocks would be straightened to the nearest cell if nudged. If they had been knocked to the floor blocks would be repositioned on top of the nearest stack, in what has been described as ‘an inspiring image of a machine that paid attention to the preferences expressed by the gerbils and then completed and formalized them into new, pleasing structures’ (Wardrip-Fruin and Montfort 2003: 247). The exhibition catalogue notes,

If computers are to be our friends they must understand our metaphors. If they are to be responsive to changing, unpredictable, context-dependent human needs, they will need an artificial intelligence that can cope with complex contingencies in a sophisticated manner (drawing upon these metaphors) much as Seek deals with elementary uncertainties in a simple-minded fashion. (Architecture Machine Group 1970: 23).

The ability for the robot to ‘learn’ and develop what we might today conceive to be artificial intelligence (AI), was bound by the simplicity of the production rule format. For
Negroponte, an architecture machine in the truest sense would not simply serve human needs in a cause-and-effect fashion, but would sense and respond to sensory inputs. The process of design would be ‘a dialogue between two intelligent systems—the man and the machine—which are capable of producing an evolutionary system’ (Negroponte 1969: 9).

The concept of conscious cities as spaces of dialogue opens a conversation between societal care, AI and governance about the design of cultures of care. The sections that follow unpick three elements of this dialogue. 01 Voice of AI — questions the neutrality of ‘the programmer’ and posits a future of singularity and transparency. 02 Flaws and Faults considers how human biases and errors might be subsumed collectively within a democratic, conscious city. 03 Care for whom? raises questions of inclusive design and the ability of the conscious city to respond to and reconcile individual and societal needs.

**Dialogue 01. Voice of AI**

Seek’s attempts to handle unexpected events failed, and the relationship between the programmed environment and its inhabitants surfaced as a source of friction rather than cooperation. The resulting motif—a struggle between biological life and an AI—sat neatly within the era’s growing artistic interpretation of an imminent threat from ever smarter technology.

In 1968 (two years before Seek), *2001: A Space Odyssey* depicted the computer brain of a space shuttle becoming aggressive towards its captive human. The mysterious nature of computer thinking added another layer to existing folktales of man-made beings turning on their makers. From the Golem of Prague to Frankenstein, the fear of creating sentient beings as a form of self-destruction is deeply embedded in our collective memory. In George Orwell’s *Nineteen Eighty-Four* the figure of an eye in the sky entrenched our suspicion that technology as an enabler of tyranny is a real danger in the not so distant future.

In the field of AI the term ‘singularity’ refers to machine intelligence that surpasses the capabilities of human intelligence. Our fears of self-inflicted doomsday resurface in the debate surrounding singularity with bright minds such as Stephen Hawking warning it could signal the end of the human race: ‘It would take off on its own and redesign itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn’t compete, and would be superseded’ (Luckerson 2014).

We tend to imagine any being more intelligent than us as a direct threat to our existence because in biological evolution, superior intelligence secured human dominance at the cost of others. The scientific community is now debating methods to assure that AI will not become a threat to humans, suggesting a set of restrictions on self-modification, or built-in functions such
as ruling that ‘cooperation [with humans] is always preferred to conflict’ (Shulman 2010: 3). No existing proposal confidently lays out a solution, perhaps in recognition that AI might inherit human logic that allows us to break even our own laws.

The clear advantage of AI over human capabilities is the faster and more consistent methods of analysing large amounts of data. A collaboration between Harvard and Vermont universities applied machine learning tools to identify markers of depression in 43,950 Instagram photos from 166 individuals, performing better at diagnosis than general practitioners examining patients in-person (Reece and Danforth 2016).

This superhuman ability is coupled with the unfortunate and sometimes disturbing transference of the very human quality of prejudice. Researchers at the University of Bath and Princeton University have demonstrated this transference, stating ‘that if AI is to exploit via our language the vast knowledge that culture has compiled, it will inevitably inherit human-like prejudice’ (Flaherty 2016). In early 2016, Microsoft introduced an AI chat robot named Tay, to Twitter. In less than 24 hours, Tay went from tweeting ‘Can I just say that I’m stoked to meet u? Humans are super cool’ to ‘Hitler was right I hate the jews’ (Horton 2016). Microsoft quickly pulled the plug on Tay, explaining that her responses are derived from her interactions with humans: ‘a coordinated attack by a subset of people exploited a vulnerability in Tay’ (Lee 2016).

Our fears of a super-intelligence are based on two factors: a suspicion that there is a correlation between high intelligence and aggression (although the opposite is true, see Huesmann, Eron and Yarmel 1987), and that a powerful non-human would almost certainly pose a threat to our existence. We might instead look at the ‘flaws and faults’ in our own behaviour and conclude that a greater intelligence would know better what is good for us than we know ourselves (that is, unless it learns too much from us). Would a greater intelligence augment these qualities or overcome them?

There is as yet no set protocol for transparency in AI development, a mechanism that might introduce a level of accountability and confidence. In December 2016, Apple published its first AI paper as a response to growing criticism from the research community about its secretive operations: ‘The move towards more openness with the community is important for Apple as the push for more advanced AI software spreads across the tech industry’ (Tilley 2016). Might growing transparency allow us to spot and understand why and how thought processes at odd with our ideals of care, equality and democracy are seeping into technology?

Dialogue 02. Flaws and faults
Negroponte (1970) expressed a desire for technological humanism and embodied a belief in ‘collective ownership of information and information processing and the participation of users in decision-making processes’ (Vardouli 2011). This view was informed by social and political movements in Europe and made visible through Negroponte’s reference to Martin Shubik’s ‘Information, rationality, and free choice in a future democratic society’:

Modern decision theory, economics, psychology and game theory recognize, as a basic case, clearly motivated individual choice under conditions of complete information. It is also recognized that two unfortunate facts of life remove us from the relative simplicity of this basic case. The first concerns man as an information processor and the second the conflict of individual and group preferences. (Shubik 1967: 772)

Fifty years on, we still like to believe we are rational creatures, logical in thought and optimal in all our choices. In the quietness of the laboratory, with choices described and the luxury of time, we can be the optimising machine we desire. Allow the choice to be experienced rather than described (Camilleri and Newell 2009) or even take the experience of choice out of the laboratory and into the wilderness of the city (McAndrew and Gore 2013) and the patterns of human decisions change. This is not to say we become irrational, only that there are differences in choice mechanics².

It has even been suggested that humans are ‘boundedly rational’ in outlook (Simon 1957). Limited by the bounds of time, the tractability of the problem and cognitive power we interweave heuristics or shortcuts into our thought processes that exploit regularity of the environment (Gigerenzer and Selten 2002). Realising these constraints we search through available alternatives that will satisfy (rather than optimise) our minimum requirements. The result is bias.

These flaws and faults are important as we contemplate conscious cities as spaces of dialogue. We have already suggested that a greater intelligence—which eliminates cognitive shortcuts and biases—might know better what is good for us than we know ourselves. But what if these cognitive heuristics form part of the dialogue to which cities sense and respond? Could we end up creating sub-optimal cultures of care? Might we produce responsive environments

---

² Barron and Yechiam (2009) have shown that behaviour can differ from judgement, illustrating that after a suicide bombing people can both believe the risk to decrease and behave cautiously. This holds a close relation to George Orwell’s Nineteen Eighty-Four ‘doublethink’, in which it is possible to accept two contradictory beliefs as correct. People learnt doublethink and newspeak (a linguistic design to limit freedom of thought) to ‘fit in’. The result was a deterioration in mental efficiency, reality testing and moral judgement.
with a set of design conditions that simply satisfice—adequately, but not perfectly meeting our needs?

To think of conscious cities as spaces of dialogue also necessitates thinking about the collective, for ‘The aggregation of individual wants and powers into social wants and powers is one of the central problems of political science, economics and sociology’ (Shubik 1967: 774). We can see this in studies of human cognition, such as the phenomenon of groupthink, a coined by William H. Whyte, an American commentator on cities, people and open spaces. ‘We are not talking about mere instinctive conformity—it is, after all, a perennial failing of mankind. What we are talking about is a rationalized conformity—an open, articulate philosophy which holds that group values are not only expedient but right and good as well’ (Whyte 1952). It denotes a mode of thinking in which concurrence-seeking within a group becomes so dominant that it overrides a realistic appraisal of alternatives (Janis 1982). This desire for harmony results in unwittingly irrational and even dehumanising outcomes for the Other.

Writer, editor and broadcaster Ian Hislop (2016) opens up the issue of citizens becoming trapped in their own social media ‘echo chambers’, hearing only the opinions of those in their own circles and dismissing the facts and views of those that exist outside of it. Mary Cross (2011: 62) applies this theory to Twitter, which ‘submerges independent thinking in favor of conformity to the group, the collective’. She notes that New York Times columnist David Carr refers to the digital social sphere as a ‘throbbing networked intelligence’. We are reminded by Sarah Robinson (2015: 47) of Dewey’s definition of empathy as ‘entering by imagination into the situations of others’. The question is, how can the consciousness of a city be ‘read’ democratically without being locked down in its own ‘filter bubble’?

Such concerns — albeit termed ‘loopthink’ — are also appearing in machine intelligence circles: ‘Machine intelligence may also come to mimic human foibles, including the psychosocial phenomenon of “groupthink,” in which excessive conformity to the group dynamic inhibits appropriate critical reassessment of a group’s policies and actions’ (Cheshire 2017: 7). Cheshire envisioned a future where some components of care are provided by AIs operating as interactive collections of cognitive entities.

This loopthink would be a type of implicit bias, similar in some respects to the human bias in groupthink, that resists appropriate reappraisal of information or revision of an ongoing plan of action. Instead, digital processing of morally relevant data gets stuck in a loop of uncritical, rationalized, repetitious uniformity. Lines of code click along quietly,
despite signals that things might be headed in the wrong direction, signals ignored or sidelined by the AI. (Cheshire 2017: 8)

It is not clear how individual wants and collective desires can be assembled in a democratic society. With empathy contingent on understanding and sharing in the feelings of the Other, how can the conscious city care equally and act equally and with fairness in its decisions?

**Dialogue 03. Care for whom?**

Arguably, conscious cities have existed since the early stages of urbanisation. The city’s consciousness, or rather awareness, of its populace is reflected in the observations, predictions and reactions of decision-makers. The dialogue between decision-maker and citizen has varied with the balance of power—between plebeians and emperors, citizenry and democratically elected officials, to today’s mixed model of central and local governance.

It is the dialogue between administrator and citizens that forges the city’s awareness of and empathy with its people. At the heart of that relationship lie the mechanisms of human decision-making, most importantly in vested interest, and empathy.

What is the motivation of a city to care for its citizens, and how does empathy inform an action of care?

This motivation stems partly from the confidence of citizens to demand more from civic space, emerging with the advent of nation-states and the weakening of feudal systems. The resulting political shift to republican, democratic, and socialist agendas empowered citizens to imagine what value public space could hold.

In 1968, Henri Lefebvre demanded ‘a transformed and renewed access to urban life’ (Lefebvre, 1996: 158). David Harvey endorses the concept: ‘The freedom to make and remake our cities and ourselves is, I want to argue, one of the most precious yet most neglected of our human rights.’ (Harvey, 2008).

The blurring of ownership boundaries poses a question about the balance between private and public spaces, questioning the threshold between individual and collective expectations and responsibilities. Civic engagement meant individual responsibility to the public sphere, and emerging universal rights meant the creeping of government responsibility into the home. This dichotomy of private/public and individual/collective forms the basis of the question of whom the city cares for, and how? How can a city care for the individual and the collective simultaneously without conflicting interests?
If the city is to understand our needs, we must first be able to define for ourselves the thresholds of individual and collective desires. Our need for novelty, for example, is different in the home than in the city. We might change our regular commute to raise the chances for serendipity in public space, but it would be strange to behave similarly in the home. Unexpected novelty in our private daily routine would be unsettling.

Our needs and trust, therefore, adjust as we cross these thresholds. The more people, and the greater diversity of people, we share a space with, the more difficult the question becomes of whom and how to prioritise. If a city cared equally for people who benefit from different experiences within a shared environment, how could it care democratically? Considering the finite nature of economic and material resources to be distributed, even a highly developed intelligence would run into ethical issues of this nature.

In the US, a 1981 Executive Order that institutionalised a cost–benefit analysis for Federal decision-making stated that ‘regulatory action shall not be undertaken unless the potential benefits to society from the regulation outweigh the potential costs to society’ (Shabecoff 1981). Supporters of the order explained that evaluating the outcomes of actions would lead to better decision-making. The flaw of such a system surfaces when benefits become difficult or even impossible to measure. Even the important cause of investing in health care becomes a quantitative obstacle, let alone issues that can evade measurement, such as aesthetic qualities of space.

The cost–benefit formula underpinning this policy depends on the measurability of each variable, even though those variables become almost impossible to assess democratically. To decide which neighbourhood a new air route should fly over, for example, one would need to calculate the total detriment of added noise to one local population compared with others. In reality, the perceived detriment to each population is extrapolated from the volume and intensity of objections raised, and their sources. Once again, the politics of ingroups and outgroups trump equality.

Even a well-intentioned formula for decision-making has the potential to discriminate. Accordingly, our confidence in political systems is based on a set of rights that guarantee to limit damage to individuals. Rights (and a robust acceptance of the notion of natural or inalienable rights) create thresholds between collective will and individual needs, limiting the ability of ingroups to ignore the needs they share with outgroups. It is perhaps those same mechanisms that would limit the possible harm AI could do, and allow humans to trust in the powers given to it.
Looking ahead

Seek, and Negroponte’s (1970) vision of an ‘architecture machine’ imagined the process of design as a ‘conversation’ between two intelligent entities, human and machine. The theory of conscious cities as spaces of care, has broadened the conversation. It calls not only for empathetic sensing, but intelligence in response, suggesting that one of our most urgent dialogues ought to centre on the relation between societal care, AI and governance. That we are faced with a lack of certainty about how to embrace this notional AI future, is perhaps not surprising given our inability to decipher what is best for ourselves.

Hello, Robot, an exhibition at the Vitra Design Museum in Germany in 2017, is one attempt to bring the public into this conversation. Exhibition curator Amelie Klein alludes to this tension when she notes, ‘The difference between well-meant care, patronising control and spying is very fluid’ (Aouf 2017). Exploring the relation between trust and care, the exhibition asks its spectators: ‘Do you want a robot to take care of you?’, and even, ‘How do you feel about objects having feelings?’

The dialogues raised in this chapter challenge society to define this relation more fundamentally—What is care for us? Dialogues 01–03, address the dimension of trust through this relationship. This is not just about providing belief in the voice of AI or faith in our own decision-making abilities; it runs deeper into our confidence in the aspiration of the city to empathise. It might seem like a contradiction to look to the present day, but dialogue needs trust, and for this to be enabled we need frameworks of governance and law. We might even consider a form of constitution.

This is not unusual thinking. In the science fiction story Runaround (1942), Isaac Asimov proposed three laws to govern the behaviour of robots:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings, expect where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Later in the story he introduced a ‘zeroth law’:

0. A robot may not harm humanity, or, by inaction, allow humanity to come to harm.
These laws might be seen to function as a constitution, one that ultimately limits the power of AI in the design of future cultures of care.

References


Possible Images
A Manifesto for Conscious Cities (2015)

Biographies

Claire McAndrew
Claire is Director of Research at the UCL Institute for Digital Innovation in the Faculty of the Built Environment. Combining social science insight and design-led thinking, Claire's research focuses on the possibilities of design and digital innovation within the built environment for transformative social and cultural effect. She has collaborated on national and international
digital innovation projects that traverse communication and interaction design, social science and the built environment. In 2014 she was awarded an AHRC UnBox LABS Fellowship on Future Cities. She joined The Bartlett in 2011, receiving her Chartered Psychologist status the same year.

**Itai Palti**

Itai is an architect and researcher focusing on designing with the human experience in mind. He is a Co-founder of the Centric Lab, the first cognitive neuroscience research lab for the built environment. In 2015, Itai founded the Conscious Cities movement, proposing a new field of research and industry for building people-centric environments that are aware and responsive. He is a fellow at the Centre for Urban Design and Mental Health, and a member of research and policy groups including Harvard’s Frontiers of Innovation and the Brookings Institution’s Learning Landscapes Committee. Itai’s writing on the intersection of design and science has been featured in leading publications such as *The Guardian* and *Quartz*, and he is a co-curator of the conscious cities publications and conference series.