**The Digital Subject: People as Data as Persons**

**Olga Goriunova**

**This is a pre-print of an article (author’s original draft) accepted for publication in *Theory, Culture and Society*, special issue ‘Transversal Posthumanities’, forthcoming 2018.**

A few years ago, Chiara Bernardi, at the time a doctoral student of mine, was working as a consultant in digital marketing. She attended a lecture delivered by a representative of MaxPoint, an American advertising company entering British and European markets. MaxPoint combines census data, data from Telmar (a large database of media and consumer information), bank transaction data, private browsing data, cookies and other more advanced online activity data, and data from a mapping algorithm that tracks behaviour based on location, to name just a few sources. Based on this information, MaxPoint offers a service called “hyperlocal advertising.” For as little as £5000 the service can group people with similar online behaviours living in the same neighbourhood into “Digital Zips” and deliver personalised ads to them. Though hyperlocal advertising information is scaled to neighbourhoods exhibiting commonalities at the household level, the MaxPoint lecturer Bernardi saw proceeded to describe an individual located on a map in terms of available data. That individual was most likely a woman who lived on a particular South London street, ordered organic vegetables and many books, used London Overground, and was not on Twitter. As more details followed, Bernardi realised the person was me.

What exactly is this digital entity that she identified as me? What relation does it have to me? How do I relate to it? How is it able to stand in for me and construct a me that attracts advertisements and thus alters me, while still being reliant on my activity? How is it produced outside of my awareness, mobilised, and recruited? It is clearly not I, and yet it is no one other than I. What other Is are out there, labouring in the legal, medical, industrial, and aesthetic spheres?

To engage with this encounter, I propose the term *digital subject*. This concept includes a subject of a data profile or of a Facebook stream, a history of browsing or search engine queries, mobile phone positioning records, bank transactions, sensor data, facial recognition data, biometric movement recognition data, or email inboxes, among other things. The digital subject thus moves between captured, unique, and persistent biological characteristics and premeditated forms of symbolic expression, judicially inferred subjects of actions, and performed identities. It is this very entanglement of physical, legal, sensual, and cultural elements that warrants the use of the term.

Various digital subjects connect and feed off the experiences of living people. The digital subject is therefore distinct from the living self, addressed in the continuity of its experience in critical psychology and psychoanalysis, though one does not necessarily need psychological tools to talk about it. The digital subject is an abstracted position, a performance, constructed persona from data, profiles, and other records and aggregates. While it would be a fruitless task to disentangle the living self from the digital persona, in defining the digital subject I follow a tradition of looking for forms of connection between subject (subject position) and subjectivity (seen as “more than the sum total of discourses from birth,” a complexity of experiences possessing a non-derivative ontology) (Henriques et al., 1984; Blackman et al., 2008). A digital subject comes after the subject, requiring new ways to understand how it connects to the subjectivities of living persons.

It could be argued that subjectivity itself, partly an effect of cultural and disciplinary individuating techniques (Foucault, 2005) and partly a product of work on the self through technology, such as the diary (Kittler, 1990), is formed in structural coupling with its computational environment, making all subjects digital today. Digital hybridity is the de facto mode of contemporary existence. On one hand, the omnipresence of hybrid technical forms might seem to render the modifier “digital” redundant. On the other hand, the modern “subject” as a site of thought, moral judgment, and action in its very conceptualisation has contributed to patriarchy, colonialism, and many other practices of subjugation, and therefore its return is not unproblematic. However, digital subjects are new forms of subject construction that arise out of computational procedures and are employed by various forms of power to distinguish, map, and capture not only subjectivities, but also non-humans and physical things that inhabit the world. The arena of digital subjects has become a new site for struggle. In this paper I aim to demonstrate the distinctness and usefulness of the term by discussing digital subjects in terms of some of their modes of production and critical perception. I will look at how they perform by focusing on the subjects of myself that are constructed digitally, and ask how they connect to the living I that is always a multiplicity (Braidotti, 2008). The concept of the digital subject therefore builds on the established critique of the subject to engage with the processes of its construction, recruitment, and enactment that occur in the computational regime. Even if digital subjects are solely a result of the neocapitalist desire for construction of fixed personhood, we first need to study them in detail to be able to imagine the alternatives.

Scientific notions of identity (such as the unique chemical composition of a molecule or the inner structure of atoms) are employed in the construction of digital subjects in order to capture people or things. For example, the face becomes the new fingerprint, and movement recognition technology claims it is impossible to move in ways nonidentical to one’s unique self. Here, digital subjects are mapped onto living persons in an attempt to eliminate excess and non-coincidence. Yet digital subjects also offer sites of reinvention, liberation, and play. Fake accounts and performed identities testify to that. It is this combined promise of identitarian uniqueness, intimate access, and legal accountability, as well as the creative potential for misidentification and theatrical performance of an aspect of one’s subjectivity, that characterises digital subjects.

In recent years, notions related to the subject and its digital counterpart have resurfaced with great force and proliferated. In the exact sciences, the digital subject often appears as an unproblematized extension of self or person: the digital self and digital personhood. In social sciences and cultural studies it is discussed as identity (constructed in social media) and as the data citizen (Isin and Ruppert, 2015). In critical theory, it is often a “data double” or a “data shadow” (Raley, 2013). Drastically different in their implications, these varied terms indicate that new computational subjectification and subject-construction processes trigger ontological and epistemological anxiety, while new modes of measurement, evaluation, and prediction offer significant monetary incentives and promise new kinds of politics. As a term, the digital subject is intended to bridge critical theory-based considerations of selfies, Facebook streams, and search queries in terms of “mediated digital selves,” and big data analytics-based debates on the construction of data profiles in terms of probability, prediction, and control. If the former is often narrated as “prob[ing] the gap between the perception of our own identity and its distributed representation via social media platforms” (V&A, 2015), the latter is typically framed as a critique of data analytics as operationalizing and preempting the world largely outside of existing systems of accountability, whether scientific or conceptual (Rouvroy, 2013).

The digital subject thus poses questions that cannot be answered within one discipline. While the humanities have a long tradition of thinking subject and subjectivity, for computer science the human is a mathematical array. When critical theory engages with data production, it often veers all to soon into discussions of surveillance. This topic is massively important, and yet such a well-trodden rut effectively prevents close, conceptually rich examination of the processes comprising data analytics. It is a political choice to not dismiss certain forms of rationality or formalism too easily or too early. As many paradigms and disciplines currently collide, calling for a radically interdisciplinary posthuman enquiry, the digital subject asks for new ways to understand causality and determination, unity and disunity, interiority and exteriority, as well as reason, expression, and sense. In this paper, I focus more specifically on the connection between the living person and its data—its digital subject—which I explore by addressing the promise of indexicality when it comes to this connection. Instead of indexicality, I argue, there is distance. I ask again and again: what is the digital entity that Bernardi identified as me? What relation does it have to me? How do I relate to it? How is this relation constructed?

**Distance**

The notion of distance is key to conceptualizing the digital subject. In Clement Valla and A.E. Benenson’sartwork *Some Sites and Their Artefacts:* *123D Catch* (2014), an automated camera glides through 3D rubble, presenting decaying digital objects, such as a lion’s head and antique busts, created using Autodesk 123D Catch software. An automated voice narrates the decay of 3D objects in the virtual space, starting with the following statement: “A digital artefact is neither an object nor its representation but a distance between the two.” For my purposes this could be adapted as: “A digital subject is neither a human being nor its representation but a distance between the two.” This phrase is key to capturing the anxieties around humanness, digitality, and representation.

I propose the notion of *distance* to discuss the relation between digital subjects and the humans, entities, and processes they are connected to. I choose distance as opposed to relation for an important reason. Distance is not representational; it induces change. In this context, distance is Gottfried Wilhelm Leibniz’s distance: luminous and thick. As non-empty, qualitative distance, it can expand and contract, stretch and collapse; it is full of interference. In his writing on Leibniz, Deleuze specifies that the fold comprises many planes, as it is always a fold of folds. The fold does not consist of points or parts, but is elastic and unites different forms of matter (Deleuze, 2006: 4-6). The distance of the fold, for Deleuze, must contain a metamorphosis: there is always a transformation occurring within the distance. The distance is also temporal; its limit is not quantity, but space-time. Such a distance is not an arrow from 0 to X: it is vast and twisted, found between two constantly changing entities. The “person,” the starting point of such a distance, is not necessarily a human being. It could be a human body moving through transport networks, but also a smart home, a sensor on a pigeon’s foot, car license plates, a fictional story or figure, a collective of anarchists sharing one mobile phone, or hundreds of other alliances, long and short-term. This “person” can be composed of things that form a relation without necessarily possessing subjectivity. At times a “person” might be a compulsively socialising teenager and at other times a SIM card shared among a village.

The person’s “representation” is not a representation. Data are not traces (Rouvroy 2013, Gitelman 2013). The causality operative in the distance is computational, at least partially. The legacy of the debate about whether data constitutes representation shares concerns that surrounded the medium of photography at its inception. As Honoré de Balzac wrote in *Cousin Pons* (1847): “If any man had come to Napoleon to tell him that a building or a figure is at all times and in all places represented by an image in the atmosphere, that every existing object has a *spectral intangible double* [italics mine] which may become visible, the Emperor would have sent his informant to Charenton for a lunatic […] Yet Daguerre’s discovery amounts to nothing more nor less than this” (Balzac, 1847: 584-5). In the current situation, the new medium is data, and its genealogy is to be found in the science labs, scientific experiments and statistics, rather than the legacy of art, in which landscape and aura are both forms of distance (Benjamin, 1936). After decades of discussion about the status of indexicality in photography, there are new ontological concerns about the image following a change in the material production of a digital photograph (Osborne, 2013: 127). Peter Osborne’s argument, for instance, is that technological specificity is always “fantasmatic” and its social uses of indexical signification are never secure (2013: 124).

While abstraction in data analytics is not linearly causal, it operates as distance, obtaining models, deducing statements, making inferences. Apps offer to cross-reference Instagram feeds and IP addresses, and then IP addresses with street addresses, orchestrating multiple alignments that are beyond representation. The ways they are aligned create digital subjects. Still, in posing the question in the manner of Benenson and Valla, there is a danger of constructing the (human) subjectivity and the digital subject categorically and in the manner of a one-to-one correspondence. To counter this tendency, distance must be understood as non-dualistic: its forms of contraction and twisting are grounded in the specificity and variability of embodied techno-cultural and socio-political practices rather than any idea of essential qualities. Enquiry about distance must also be based on the wonder of encounters with digital subjects similar to the one I had (Stengers, 2010). In such an event, I am confronted with what someone or something recognises or uses as me—a record and a model of whatever I could be. As digital subjects are constructed not only to sell products but also to imprison, medically treat, or discriminate against individuals, the non-coincidence and spatiality of the distance become urgent political matters.

Therefore, the choice to claim representational, indexical, correlationist, or reinventive distance is not only a matter of funding or disciplinary norms, but of politics. A humanist would claim the distance as the site of rivalry, while a media scholar might focus on the distance as the producer of authenticity. Hackers hide in the distance, while artists make work from it. Some sciences would accept the distance, while others claim there is no distance, but rather equivalence (a human equals her Tweets; we can understand society by studying Twitter). Struggle over the distance is political through and through: to claim and maintain distance is a matter of keeping and developing alternative options of becoming and living with our digital subjects.

In this paper, the question of how my digital subject is produced in relation to me and what it produces in return is ultimately a question of how to establish and understand the operation of distance. The distance can be interrupted, recruited, intersliced. When one is abused online, the distance collapses; when a security-mad enthusiast employs an echelon of tools to hide electronic footprints, it may become tense and dense. Distance can be maintained and manipulated: a technical disturbance or other processes can be disguised in the distance; hence it can serve as an artificial enhancement (for instance, an iris scan performed by border control can obscure multiple other checks and cross-references occurring at the same time). Playing with a camera’s image filters, probing the limit at which the face would become non-recognisable, is maintaining the distance, pulling on it and relaxing it. These are important acts when it comes to the technical-subjective maintenance of distance.

To sum up, the forms of production of distance, whether by neural networks, platform infrastructures, projects, or other engagements are key to the generation of digital subjects. The fluctuating distance is manufactured, created, and played with. It is through the actualisation of the distance that digital subjects acquire value, become of service, or gain aesthetic qualities. It is also through the distance that digital subjects become more or less personalised or multiple—put together and disaggregated. This distance is a concept that specifies a new relationality, and that allows one to address both ontological and epistemological questions brought about by data modalities. These questions can be approached through various genealogies, allowing for a meeting point between disciplines,.

The distance must be established; it can then be claimed. Less “valuable” people, such as women, people of colour, or poor people can be assigned digital subjects that continue discriminating against them. Some people in the Global South do not get to have digital subjects at all. The computational production of digital subjects is not “naturally flowing,” “objective,” or transparent. Claiming the distance brings attention to the processes of production of digital subjects and the ways in which their relation to people, their indexicality, is established. It can then be interfered with, redirected, played with, and reinvented.

In what follows below, I will discuss this distance in relation to existing notions of indexicality. I will then examine the use of neural networks as a specific case of the production of distance. I will conclude by drawing upon art practices as a distinct form of maintaining distance and argue for hybrid forms of actualizing distance that I believe the posthumanities are uniquely positioned to develop.

**Distant indexicality of data**

Constant data generation is the condition of today. While data generation is not direct measurement, it is unclear whether and what it represents; and although data generation is not exactly an ongoing occurrence such as photosynthesis, its “naturalness” and scale often invite biological metaphors. As things and humans go on about their business, leaking data, this informational condition can be considered a new ecology of computation. The key to understanding ecology is that there is no simple causality (Stengers, 2010: 32-33). Claiming, establishing, overcoming, and doing without linear forms of causality, and questions surrounding correlation, remain some of the core themes in the critique of big data.

Discussions around data generation are often framed in terms of demography and the vocabulary of statistics. Rita Raley, for instance, writes about the “digital demographic self”: a person with data produced by her that includes approximate age and gender, changes in bodily states (getting pregnant, getting ill), location, consumption behaviour, interests, social behaviour, and other aspects (Raley, 2013). Raley includes demographics and psychographics (interests and behaviours, often called “proprietary interest data” in advertising) into the demographic category. Raley’s digital demographic self is a digital census citizen, a consumer, possibly a patient or a suspect.

There is a lineage in documentary modernity, as Ronald Day argues, from documentation that works as an indexical sign to prove a phenomenon, that is “any type of ontological substance that acts as evidence” (Day, 2014: 5) to information: a transfer of knowledge from the document, a process of informing, a response to the need (for knowledge) (Day, 2014: 37, 42). Descriptions of the generation of digital subjects, which are themselves responses to requests, are well rehearsed. Only the recent, hot “data footprint” matters. If one has browsed an online store’s sofa department, one will be served advertisements of exactly those or similar sofas on websites they visit next. As it is highly unlikely that the interest in sofas will persist indefinitely, the value of this interest data will decline with time. With only recent activity valuable in this context, a holistic picture is not strived for: this digital subject does not constitute a full psychological or economic portrait.

What kinds of digital subjects are there? One visits a website to look for a sofa and is assigned a unique number that is matched with a specific product and/or interest category of sofas and living room furniture. This match is already a digital subject. An identifier could be an IP address, a cookie, or a coordinate derived from Wi-Fi triangulation. The matching expands from deductive to inductive. As more data is processed, correlations are found within it. These correlations are also digital subjects: interest in sofas, together with purchasing power, geographical position, age, preference for particular products and styles. This data is turned into profiles: models applied to correlations in the individual data, rules and data on other people. Here, induction and abduction become possible: judging by similar behaviours, it is possible to infer that one might be inclined to wish for a designer coffee table from a specific store or compute a prediction linking a computational model and an individual data. Profiles are also digital subjects. Duration and consistency are gone here. The fact that one may need a reading light by the sofa will not necessarily be part of such a profile, as it is not a spy’s report on one’s personality as a compulsive reader.

Raley points out that data and subject are indexically and repeatedly linked: the digital subject is constantly produced, re-instantiated, re-engendered, refreshed (2013: 123). The digitally, demographically, and psychographically centred digital subject is in fact a set of dynamic processes that have the structures of computational actions, models, and socio-political cultures. It is a process in which no exact or stable state is significant or valuable: what matters is the algorithmic interpretation at the moments data can be used, sold, or otherwise acted upon.

According to Day, documentary signs function within a broad socio-technical regime of information organisation that gives order, value and meaning to “texts and person as documents” (2014: 20). For Day, the index, as a mode of modern documentary tradition similar to Osborne’s distributed photographic complex as described above, changed from “explicit structures” to “implicit devices,” where both documents and users merged into data (2014: 2) at the same time as the socio-technical regime or the infrastructure itself disappeared from view (2014: 38, 47). To some extent for Day, as for Raley, social media act as “evidentiary fragments” in modern information infrastructures (Day, 2014: 29). Assigning an identity through a documentary process, he argues, is core to modernity, the process that today extends to asserting individual existence through attribution, where individuals are reduced to “conjoined data points” and indexed in the current computational information infrastructure (Day, 2014: 59-61). However, in his wonderfully argued book, Day also writes: “The problem […] is that it may appear […] that what is given by the information is a ‘fact:’ that the constellation of references is closed, that the text […] is a document” (2014: 66). Here, Day’s “problem” is distance.

While individual data points can be tendentially evidential, they are not documents or evidence, and profiles are even less so. Digital subjects are values, dynamically re-instantiated correlations, rules, and models, shreds of actions, identities, interests, and engagements, which are put into relation with each other, disaggregated, categorised, classified, clustered, modelled, projected onto, speculated upon, and made predictions about. Digital subjects are unconnected and entangled, distributed and distributive. As models can be relaxed or tightened, profiles are evaluated in terms of probability. Digital subjects are future oriented. Computationally, they span different spatio-temporal scales: they can differ in length of alphanumerical strings, in complexity, in forms of composition and proposition, in proximity of evaluated parameters, in number of units, in frequency of occurrence, and in the types of future they propose.

When it comes to assessing risk in insurance, for instance, with its specific periodicity and temporality, an individual does not live long enough to correspond to enough points at which data is acquired to offer accuracy. While with other kinds of data (such as Facebook data or supermarket buying patterns) the degree of correlation can be more accurately decomposed and thus more meaningful for specific products, the individual scale does not necessarily or universally give better indications. The digital subject here is not personalized; it corresponds to an aspect of a population. A digital subject, in fact, rarely corresponds to a classically constituted individual: *it is always more and less than a human*. I would argue that digital subjects are not ever computationally aggregated into an order and composition that corresponds to the classical modern subject. Such aggregation is possible, though laborious, and relies on turning data into evidence and establishing indexical links (more on this below).

While promising extreme individualization, and indeed capable of granularity, digital subjects are constituted by processes that scale up and down in their distributed-ness. As one example, the momentary requests of web-based advertisement services and the continuous data aggregation and permanent storage of the Five Eyes surveillance programmes pose radically different temporalities and kinds of threat, which nonetheless rely on the some of the same sets of computational procedures. Ranging from a single value to lots of data, from rules to models proposing meaning about a profile or a population, digital subjects change form or identity: they become different things. Yes or no in a row; the table itself; the model—these are not spatial and temporal distributions of digital subjects, but forms of recording, aggregating, establishing relationality, and prediction, where each subject has its own spatio-temporal framework. These are the processes and frameworks that comprise and characterise the distance under discussion, though they are not able, on their own, to exhaust or explain it away.

Digital subjects offer new forms of singularity and multiplicity. The basic dialectic of today is the promise of individualisation, of standing out via the creation of a singular occurrence; yet such a singularity can only be produced through constant aggregation, comparison, sorting and re-arrangement of other singularities in the operation of multiplicity (Goriunova, 2015). Singularities can only exist through their relations within the multiplicity: they are determined from the interactions of multiplicities. The oscillation between singularity and multiplicity is not circular. In the case of profiles, correlations in the data linked to an individual are used in a model that refers to a group of people, from which something can be inferred about the individuals whose data points might be missing. Such individuals are neither the people whose data is crunched, nor exactly specific people at all (van Otterlo, 2013: 44). They might not exist at all, yet might still be *likely* candidates to fit a profile. The determination of singular event (such as targeting a likely candidate) is enacted from the operation of multiplicities, in which singularities pulsate in and out of existence. The likelihood becomes the condition under which a singularity might be inferred from multiplicities, so it does not even need to exist or be proven right. Yet a multiplicity is aggregated from singular points, rules, and models.

On request, digital subjects are pulled together into plastic aggregates of subjects, which are sets of associations, propositions, and probabilities. While digital subjects are produced and distributed by software, they are also pulled together within computational socio-political infrastructures. They are “flecks of identity” (Fuller, 2005), distributed computational processes, with the capacity to become concentrated, singular, and linked to individuals, whether precisely or likely. The subject here has been lost in abstraction, reduction, scales, and directionality, and yet is constantly arranged to present forms of correspondence, correlation, and association. Performative and processual action produces, compares, updates, and manages digital subjects dynamically and constantly: they are co-constructed by computational infrastructure, by the way in which software records, marks, and structures activity.

How do digital subjects map onto data-generating individuals? Digital subjects enact realities and forms of action. What do they enact? Subjects born every time anew, which, at a distance, somehow manage to maintain continuous relations with the individual or her multiple non-unitary subjectivity? If the beginning point of this data generation is a body, a demographic self, a citizen, a psychic, cognitive, or communicative process, or an aggregate thereof, and the end point is a constantly changing collection of identifiers, values, patterns, models, and profiles, how are both parts continuously affected by each other and reinstantiated through the distance that links them? Digital subjects are something arising out of data generated about something, and become active in the computational infrastructure that enacts something else in turn. Do these three somethings manage to coincide through a distance? Are they connected tightly or loosely, and through which mechanisms?

In response to these issues, Raley suggests the notion of the “data double.” For Raley, data doubles are discrete and constantly aggregated data bodies, Deleuzian dividuals, rather than the “figures and products of modernity” (Raley, 2013: 127). According to this line of thinking, we become our data, granulating our subjectivities in the data worlds we inhabit.

Yet the data double is like a shadow. This shadow arguably belongs to the representational paradigm in science. Andrew Pickering writes: “The representational idiom casts science as, above all, an activity that seeks to represent nature, to produce knowledge that maps, mirrors and corresponds to how the world really is. […] Within the representational idiom, people and things tend to appear as *shadows* of themselves” (Pickering, 1995: 5-6) [emphasis mine]. One could argue that a radical critique in the tradition of the humanities corresponds conceptually to the well-critiqued representational paradigm in the philosophy of science. And worse, such a paradigm is put into service by big business and politics, assigning the “usefulness” of science and data regimes on the grounds of the modern quest to conquer through abstraction, reduction, and representation. At risk of confirming what they argue against, the critical humanities face an acute interdisciplinary challenge: the need to develop a notion of distance relying on the performative idiom of science.

Digital subjects do not have absolute representational correspondence to humans and their collectives. They are not doubles. *Indexicality comes from elsewhere*. Certainly, Facebook’s advertising service insists that it gives access to “real people,” maintaining that there is a relation back from the double, an indexing relation. It is, however, clearly a speculative relation, one that is beyond indexing. Because the return and blossoming of behaviourism across fields is undeniable, questions of privacy and human rights are urgent in this context. With an equivalence habitually claimed between “digital demographic selves” and data-generating individuals, there is definitely a danger of coordinated control. But when indexicality is appointed from outside data relations, it is destabilised, pushing distance to the fore.

Distance, here, is topologically twisted time-space. A human, or a “nondecomposable solidarity of occurent existence” (Massumi, 2009: 45), and digital subjects, or particular computational formalisations, form a Moebius strip: one follows the path but always ends up on the other side. One might make an analogy with writing: we are writing a story about ourselves and in the story are the subjects we become, the subjects of the story as we write it. We are the subjects of a book that becomes concrete and real through its writing. This is the subject as a position, abstraction, or construction—it subordinates, controls, and acts. However, there is no collective author doing the writing; the process is computational. For this reason the analogy with writing may be misleading. The story is made of patterns, similarities, models, and clusters, which are sorted, re-arranged, stored, and sold. Therefore we write ourselves by generating data that is worked upon and then produced as digital subjects, which are inconsistent and not very coherent, and serve different purposes: advertisement, secret services, or consumption. These digital subjects do not coincide with any originating “we.” They are rather at a distance. Yet, as I return to this point again and again, there continues to be a legal, industrial, and techno-scientific pull to map computed digital subjects onto human beings.

It is easy to discern why the techno-scientific complex or jurisprudence would have an interest in attaching information to a specific identity. After all, an identifiable person can be assigned debt or a prison sentence. But is there something more in the desire to do so? Data values can have some evidentiary relationship to reality but it is not absolute. Moreover, as I’ve shown, models abstracted from data or, in some cases, coinciding with data (such as the k-nearest neighbour algorithm) are not factual at all. Data can be actualised as pointers or witnesses only through meticulous work—as in the projects of the Forensic Architecture group. In their work, it requires bespoke methodologies, which Eyal Weizman and his team call techniques of the interrogation of images, along with metadata and social data. The techniques they have developed allow them to interpret data (still and aerial images, video footage, maps, reports, social media, etc.) as evidence that they supply to the UN, the International Criminal Court, and other high courts and commissions. Here, distance can be actualised into an evidentiary relation, producing, as Day argues, “the documentary … representation that … transforms the possibility of identity … into … fact” (2014: 3). However, data is not a priori testimonial.

Indexicality arises elsewhere. Forms of indexicality may include the collapsing of differences, assigning sites of authenticity, and stitching together data worlds, via tropes of unity and persistence. Key to the digital subject is that it is formed in the distance that, while distributing the individual, can and sometimes does simultaneously establish it. Indexicality looms upon us, creating individuals, because modernity calls upon individuals. Using Louis Althusser’s interpellation (in which subjects are constructed when hailed by authority as in: “Hey, you there!”) and algorithmic interpolation, Day suggests that data singularities are pulled together into an individual precisely because modernity has trained us to become subjects via a response to that call (2014: 78). To function within orders of modernity, we must respond to the call upon us. “It is a sociocultural, technological, modern documentary system that is calling one, as one […] First of all, it demands that one has formed a notion of the documented subject within one’s self […] One must have internalised the law of the subject before the call of the law or any other moral order” (Day, 2014: 80). It is upon this training from birth that one responds to the hail of algorithms.

It is then with the proposition of the posthuman and posthumanities, which dispose of the subject as the metaphysical centre of the modern order, where ascertaining the distance can begin. Citizen, body, demographic self, identity, and individual are usually terms of unity: subjectivities as assemblages of complex and intensive forces are a form of resistance to the logic of modernity. The generation of the digital subject in the multimodal distance that is itself a site of contention incorporates resistance to The Subject, continuing the project of Deleuze and Félix Guattari, Rosi Braidotti, and other scholars. Digital subjects can be established as evidence, or they can be used to reconstruct the modern subject, collapsing the distance and making the claim of digital surrogacy. But the options are multiple. Digital subjects are illusory subjects, imagined by networked machines as profiles, likelihoods, and probabilities. The non-coincidence of the digital subject, either with its data sources or with its actions, establishes the distance. Sometimes the distance is vast and at other times it collapses. Sometimes it is evidentiary and at other times it is purely speculative. It is misleading, and yet in this very fog lie the possible routes to escape.

**What do digital subjects look like? On neural networks**

The specificity and range of operations generating digital subjects require a close reading of modelling, data structures, services, and infrastructures. To understand digital subjects is to develop posthumanist descriptions and cross-disciplinary vocabularies for the computational socio-political operations that bring them about. In the short section below, I focus in particular on Lookalike Audiences, a Facebook advertising service, and on some of the neural networks that, together with other machine learning engines, run this kind of product.

Facebook offers numerous advertising services. The data it holds about its users include mail order and online purchase frequency, credit card balance and “tendency to make indulgent purchases,” level of spending, types of technology used at home, interest in fashion, life stage, and affluence, among others. The Lookalike Audiences service offers a potential expansion of the customer base for each advertiser. For example, a company sells products online. Facebook places a “pixel” on the purchase webpage and tracks customers’ behaviour elsewhere on the web, building a database. It then runs the collected data against its own databases, looking for profiles with behaviours that might resemble the company’s customers, subsequently targeting them. Here, a group of attributes (Facebook claims that it has over 100,000 attributes) is compared and matched according to different degrees of proximity, using different algorithms. More matching attributes will offer a good fit (best for small-scale campaigns) and less will increase the lookalike yield while decreasing the “lookalikeness.”

How does distance operate here in order to produce the digital subject of the lookalike? The “lookalike” is most likely produced by a range of neural networks. Neural networks are part of machine learning, a branch of computer science that studies algorithms and systems that “improve their knowledge or performance with experience” (Flach, 2012: 3). Models that see what entities look alike are results produced by machine learning algorithms as they are trained on and work through data. Neural network models are indeed called models, and yet they differ from the mathematical statistical modelling intended to produce accounts that provide causal explanations that could be mapped onto “real systems.” David Byrne, for instance, argues that modelling “is formalising by abstracting,” where abstracting is the ability “to manipulate mathematically” (2002: 113). Neural networks are computing systems that, like simulations, can deal with complex non-linear cases. Byrne claims: “when we deal with non-linearity and emergence we can’t turn to any formal modelling because we won’t be able to establish an analytical solution to equations. Differential calculus will fail us in a non-linear case. Neural nets can handle non-linearities” (2002: 131, 136). Whereas the variables that were used in statistical models were causes, in neural networks they are cases. Neural networks are case-centred, and “what emerges from the procedures are sets of cases rather than models” (Byrne, 2002: 95). Neural networks’ models can be seen as an aggregation of individual cases; they act as propositions and probabilities. In one lookalike construct filed for patenting by Facebook in 2015, such propositions and probabilities are described in terms of “likelihood,” “expectation,” “possibility,” and “possibility of preference” (Cheng at al., 2017).

Continuously optimised neural networks are abductive, prospective, and predictive (Mackenzie, 2017), prescribing a new form of action. As neural networks specify a mode of learning, rather than a mode of action, the action is understated. There is an analogy here to be drawn with the new military techniques focused on the withdrawal of categorically understood military action. Just as stated in the introduction to this special issue, there is no specific difference between the state of war and the state of peace. With a focus on patterns and correlations instead of linear causality, action is driven by information sourced opportunistically and dynamically. When various tactics and strategies are merged together, a desired effect can be reached through a combination of influences, none of which is decisive or openly decided upon. Result-driven data science corresponds to result-driven interventions, which lack strong causality but allow rapid, anticipatory self-optimisation.

As causality in models gives way to analogies, continuous measurement, and the production of sets of cases, models’ validity is established by their usefulness. Jimmy Lin in particular argues that when big data substitutes causation for correlation, it acts within a paradigm of engineering (what works best) rather than science (with its regime of objectivity) (Lin, 2015). Aiming for better engineering rather than better (scientific) understanding also means that what works does not necessarily correspond precisely to anything too specific: this is certainly the case with the Lookalike service. Byrne summarises similar findings: “while neural nets are complex systems, we really have no way of establishing if they correspond in any meaningful way to any real complex system […] no correspondence can be proven between outputs of neural nets and the real world” (2002: 139 -140). The distance built by neural networks is particularly stretched; operations in the world of their own, they have a “working” relationship with something they model and simulate.

Byrne chooses to describe neural networks as “idiot savants,” tools, and “icons” (2002: 141), which share in what they stand for. Digital subjects produced by such operations—sets of cases, matches between attributes—seem to have various degrees (from some to very little) of relationship to humans: representative correspondence is not on the menu. They are more separate and further away than demographic selves, doubles, and shadows. Yet, these almost separate entities partake in our lives. There is a kind of iconicity in these “sealed angels” (Leskov, 1873), which work as engineering tools. The digital subject, as neither an entity in itself nor a scientific fact, cannot be understood on the basis of identity, causally determined from the living being. If there are indeed elements of an iconicity produced through correlation and other procedures in the stretching and twisting distance, it is only glimpsed through moments of exclamations: “It works!” Such iconicity has no relation to the angels, but to the process of its own manufacturing or even fabrication.

**Performing a digital subject**
In the *Politics of Aesthetics*, Jacque Rancière briefly writes about the logic of fact as opposed to the logic of fiction (2006: 45). If the logic of fact is established via the apparatus of objectivity and is operative in science-related work, the logic of fiction guides the creation of literary or artistic work and is no less rigorous. Literary scholarship, for instance, offers analytical tools of high value that can be employed to understand the construction and power of a literary oeuvre. The logic of fiction guides writing that becomes operative and established in a field of text beyond authorial power, as much as the production of scientific fact depends on the infrastructure of science to enliven it. Both the logic of fiction and the logic of fact are constructions of forms of engagement with reality and of its production, while reality itself still somewhat manages to escape analysis (Stengers, 2010). The logic of fact and the logic of fiction operate in domains that have previously been distinctly different. But today, science and art, model and novel, data and narrative enter into new kinds of partnerships. The logic of fact arranged by data science carries some forms of fiction, and the logic of fiction within which contemporary digital artists operate calls upon the logics of fact.

In art, it is in digital media and feminist work that some aspects of the digital subject resurface quite spectacularly today as digital identity or persona. Characteristic of work by feminist artists working with digital media is the notion of a self to be over-performed, and a distance to be filled in with such baroque eagerness that it becomes a place to hide. In what follows, I will discuss one particular project, an Instagram performance by Amalia Ulman, *Excellences & Perfections* (2014), which propelled her to stardom while triggering significant debate.

The project consists of photographs of Ulman posted on Instagram over five months that tell a story of a 25-year-old provincial girl. She moves to Los Angeles, breaks up with her high school boyfriend, and wants to be a model, dying her hair blond. She then runs out of money, “gets a sugar daddy,” attains a lavish lifestyle, gets depressed, and starts doing drugs. Then Ulman’s alter ego “gets a boob job” at the request of her lover and has a nervous breakdown. After that, she apologizes to her followers, dies her hair dark, and goes back home, posting photographs of fancy cooking and hinting that she now has a caring boyfriend. The project uses the real name and feed of the artist and consists of her photographs of herself in carefully chosen environments, yet is totally fictional. Both the narrative and the visuals are beyond stereotypical: they correspond to the story of a female positioning herself within a patriarchal social regime. Ulman is presented as very young, vulnerable, and powerless in her glamour. Yet documentation shows that many young women see Ulman’s project as a reflection on how they are compelled to manufacture their digital persona through digital portraiture.

The project has caused significant debate precisely because it re-enacts the dominant repressive visual regime, and its satiric nature rests solely in the author’s claim that it is a critical art project. Ulman claims to have taken time and effort to master the craft of the construction of feminine beauty online; the critical message lies in this exposure through re-enactment. Mixing registers of art discourse and popular social and visual cliché, Ulman’s project creates a certain digital persona as a figure, a sensibility, that transposes and takes part in creating the field of vision from which it emerged. Ulman’s fake identity functions within a logic of fiction that employs the visual terror of the stereotypical.

What digital subjects corresponding to Ulman would data analytics produce according to her data, within the logic of fact? Ulman used hospital photographs from a real, previous hospital stay, and wore medical tape in other staged photos to pretend to be recovering from breast-enlargement surgery. She was repeatedly at many of the different locations to where she claimed to be, hunting for spots with good lighting, luxurious hotels and shops, and employing a battery of tools to manicure her fictional photographic persona. According to the logic of fact, her digital subjects would be partly inaccurate (for instance, the artist would not actually desire the products advertised to her on the basis of her performance), though the data informing the subjects would not be limited to her Instagram feed. The answer, however, lies elsewhere. For data analytics, the real Ulman has no more importance than the figure with the sensibility that she produced. Ulman constructed the “girl next door,” “sugar baby ghetto girl,” and “life goddess” as three dominant mainstream trends in how women apparently present themselves online (Kinsey, 2016; Corbett, 2014). The girl, the baby, and the goddess function well enough in the symbolic regime of American and other western societies to be successfully employed in the logic of fact and become digital subjects, i.e. streams of data to be trained on, to be fed into models, to be produced as profiles. When it comes to probability scenarios, the real and fictional, the authentic and imagined, are all equally useful.

The logic of fact and the logic of fiction are arranged very differently in Ulman’s project than in my encounter with my digital subject. And yet, the logic of fiction is fed into the logic of fact without any restrictions in both examples. While the logic of fact promises factuality, in many registers it is filled with and thrives on fiction, even if this fiction is mostly culturally boring and aesthetically reductive. One might interpret this conundrum as a call for the production of exciting fiction. Also, radical changes in mutually determining arrangements of fact and fiction bring back discussions of the real.

Discussed in terms of “authenticity” in relation to the questions of digital identity or stereotypical scripts in pattern recognition-based culture, the question of the real is answered differently depending on discipline. On one hand, the scientific real of the human as physical reality that can be known is itself a metaphysical claim. On the other hand, the real that operates through culture and its narratives, figures, and metaphors is today merged with the “real” of other domains. Isabelle Stengers writes: “For each practice, it is on the basis of the definition of what is designated as ‘reality’ and what will be asserted as ‘value’ that the scope, implications, and problems of requirements and obligations can be specified” (2010: 53). Data regimes do not distinguish between bodies and novels, nature and culture; they rely on a process of recording and manufacturing data about everything. Symbols or representations of reality have always been contested constructions. The peculiar character of data creates a bizarre situation of having to defend established designations of reality in science, law, linguistics, or psychology, or to offer a blanket critique of data management of the real. As data modalities threaten traditions of defining reality, it is now data’s turn to invent its own modes of knowing the real. The question of how the real will be constructed must also be seen as an invitation to take part.

**Conclusion**

The digital subject is neither necessarily an extension of the human into digital networks (it is not a self), nor a representation of the I, but comes into being at a distance between the living being and the data pointers, profiles, models, and active propositions that it may prompt. Such distance does not have the power to create indexical digital subjects, as our data shadows, or map onto humans precisely, and may not even be concerned exclusively with human selves—either in particular or in general. There is no given or natural factuality or evidentiality in digital subjects. The first part of this paper explored such distance against representational claims, while paying close attention to the shape of some digital subjects, dispelling any idea of their unity, persistence, or human scale. The second part looked more specifically into some products that operate using neural networks, among other machine learning approaches. Constructing digital subjects in terms of possibility is also an operation of distance that is not linearly causal.

Another aspect discussed in this paper is an arts and humanities encounter with data sciences, and a contribution to digital media theory’s forms of enquiry. As such, it makes a methodological proposition, one concerned with forms of humanity while working with the changing notion of the human. Both the subject and technology are the focus of this paper as a contribution to this posthumanities special issue.

If modernity calls upon us to claim indexical relations between data, models, profiles, and living beings, it is not the only force influencing us. There are other calls, passions, knowledges, organisational forms, and layers that operate or can still emerge in the distance, produce digital subjects, and enact realities. Different purposes and uses will construct different digital subjects with the possibility of reformulation. We are in the process of calibrating ourselves in relation to new digital data forms now being introduced, and the posthumanities can become a form of alliance that claims the distance and calls upon digital subjects to manifest beyond the human, and to invent new practices of reality formation.

**Olga Goriunova** is Reader and Director of Research at the Department of Media Arts, Royal Holloway University of London. She is the author of *Art Platforms and Cultural Production on the Internet* (Routledge, 2012), editor of *Fun and Software. Exploring pleasure, paradox and pain in computing* (Bloomsbury, 2014) and co-editor of *Readme. Software Art and Cultures* (University of Aarhus Press, 2004). She is a co-founder and co-editor of *Computational Culture, A Journal of Software Studies* ([computationalculture.net](http://computationalculture.net/%22%20%5Ct%20%22_blank)).

**Bibliography**

Balzac, Honoré de (1847) *Cousin Pons.* Available in French at: https://fr.wikisource.org/wiki/Le\_Cousin\_Pons, accessed 7.09.2016.

Benjamin, Walter (1936) *The Work of Art in the Age of Mechanical Reproduction.* Available at: https://www.marxists.org/reference/subject/philosophy/works/ge/benjamin.htm, accessed 7.09.2017.

Blackman, L., Cromby, J., Hook, D., Papadopoulos, D., Walkerdine, V. (2008) “Creating Subjectivities,” *Subjectivity* 22: 1. doi:10.1057/sub.2008.8.

Braidotti, Rosi (2008) *Nomadic Subjects. Embodiment and Sexual Difference in Contemporary Feminist Theory*. Columbia University Press.

Byrne, David (2002) *Interpreting Quantitative Data.* London: Sage.

Cheng, Haibin, Xu, Xian and Yang Pei (Inventors) (2017) *Loolalike Evaluation. United States Patent Application 20170140283.* Assignee: Facebook, Inc. Filed 13.11.2015. Publication date: 18.05.2017. http://www.freepatentsonline.com/20170140283.pdf, accessed 7.09.2017.

Corbett, Rachel (2014) How Amalia Ulman Became an Instagram Celebrity. In: *Vulture*, December 14. Available at: http://www.vulture.com/2014/12/how-amalia-ulman-became-an-instagram-celebrity.html, accessed 7.09.2017.

Day, Ronald E. (2014) *Indexing It All: The Subject in the Age of Documentation, Information and Data.* Cambridge, MA: MIT Press.

Deleuze, Gilles (2006) *The Fold. Leibniz and the Baroque*. London; New York: Continuum.

Flach, Peter (2012) *Machine Learning. The Art and Science of Algorithms That Make Sense of Data.* Cambridge: Cambridge University Press.

Foucault, Michel (2005) *The Hermeneutics of the Subject. Lectures at the College de France, 1981-1983.* New York: Palgrave Macmillan.

Fuller, Matthew (2005) *Media Ecologies. Materialist Tendencies in Art and Technoculture.* Cambridge, MA: MIT Press.

Gitelman, Lisa (ed.) (2013) *“Raw Data” Is an Oxymoron*. Cambridge, MA: MIT Press.

Goriunova, Olga (2015) “The Ragged Manifold of the Subject: Databaseness and the Generic in Curating YouTube” in Ikoniadou, Eleni and Wilson, Scott (eds.) (2015) *Media After Kittler.* London: Rowman and Littlefield.

Henriques, J., Hollway, W., Urwin, C., Venn, C. and Walkerdine, V. (1984) *Changing the Subject: Psychology, Social Regulation and Subjectivity*. London: Methuen.

Kinsey, Cadence (2016) The Instagram Artist Who Fooled Thousands. In: *BBC Culture*, 7 March. Available at: http://www.bbc.co.uk/culture/story/20160307-the-instagram-artist-who-fooled-thousands, accessed 7.09.2017.

Kittler, Friedrich (1990) *Discourse Networks 1800/1900.* Stanford: Stanford University Press.

Leskov, Nikolai (1873) “The Sealed Angel.” *Russian Messenger* No.1, January.

Lin, Jimmy (2015) “On Building Better Mousetraps and Understanding Human Condition: Reflections on Big Data in the Social Sciences” in *The Annals of the American Academy,* 659, May.

Mackenzie, Adrian (2017) *Machine Learners. Archaeology of a Data Practice.* Cambridge, MA: MIT Press.

Massumi, Brian (2009) “’Technical Mentality’ revisited: Brian Massumi on Gilbert Simondon.” Interview with Arne De Boever, Alex Murray and Jon Roffe, *Parrhesia* 7: 36-45.

Osborne, Peter (2013) *Anywhere or Not At All. Philosophy of Contemporary Art.* London; New York: Verso.

Pickering, Andrew (1995) *The Mangle of Practice. Time, Agency and Science*. Chicago; London: The University of Chicago Press.

Raley, Rita (2013) “Dataveillance and Countervailance.” In: Gitelman, Lisa (2013)

Rancière, Jacques (2006) *The Politics of Aesthetics*. London; New York: Continuum.

Rouvroy, Antoinette (2013) “The End(s) of Critique: Data Behaviourism Versus Due Process” in Hildebrandt, Mireille and de Vries, Katja (eds.) *Privacy, Due Process and the Computational Turn*. London; New; York: Routledge.

Stengers, Isabelle (2010) *Cosmopolitics I.* Minneapolis: The University of Minnesota Press.

Ulman, Amalia (2014) *Excellences & Perfections*. Available at: https://instagram.com/amaliaulman, accessed 7.09.2017.

Valla, Clement and Benenson, A.E. (2014) “*Some Sites and Their Artifacts: 123D Catch*” in *Rhizome,* 8.05.2014. Available at: http://rhizome.org/editorial/2014/may/8/some-sites-and-their-artifacts-123d-catch, accessed 7.09.2017.

van Otterlo, Martijn (2013) “A Machine Learning View on Profiling” in: Hildebrandt, Mireille and de Vries, Katja. *Privacy, Due Process and the Computational Turn*. London; New York: Routledge.

“V&A Digital Futures—Self Versioning: Exploring the Potentials of the Digital Self” (2015) Announcement. Available at: http://www.spacestudios.org.uk/art-technology/va-digital-futures-self-versioning-exploring-the-potentials-of-the-digital-self, accessed 7.09.2017.