Disfluent Interfaces. Affective Conversations: Differencing Humans and Machines in AI

ANNA MUNSTER

For much of the relatively short history of computational inter-faces, designers have emphasized the importance of naturalism and its place in aiding ease of exchange between humans and computers. In the musings and applications of Donald Norman (1990), Allan Kay (1990), and a generation of human-computer interaction (HCI) thinkers and engineers throughout the 1980s, the desired aim for human-computer interaction was the erasure of a physical and existential space that maintained the distinction between computation and human agents. Many of these early currents in HCI fell into two camps in their approaches to, for example, the centrality of interfaces in computation. On the one hand, an interface such as a GUI was seen—and has increasingly been involved in web design, for example—as the location to imagistically instantiate representations of the 'intuitive' actions and perceptions of humans with metaphors from the backend of computational processes (see for example, Kay 1990). On the other hand, beginning with Kay's critique of the GUI (1990, 210), the visible interface has often been seen as something that should be progres-sively erased. However, what was common to and continues to persist as the key vector to conceiving the relation of computation to human action and perception is that the *space of engagement* across, between, or amid computer and human requires naturalizing.

In the last decade or so, this has been taken up by research and developed into natural user interfaces or NUIs (see Vetere et al. 2014; Buxton 2010). In both the design and commentary around natural user interfaces, which aims to draw upon skills and capacities from all human modalities and movements to interact with computational devices in both onscreen/online and physical spaces, assumptions about 'naturalism' have been subject to investigation. Bill Buxton (2010), for example, has suggested that natural interaction is both context-dependent and the result of prior lived capacity; it accumulates through habitual and performative human sensory modalities and gestures. The 'natural' is further nuanced by looking at interfacing with computational devices from the point of view of social interaction in more recent perspectives.[1] There has also been considerable work done within digital media theory that problematizes the notion that the interface is transparent or recessive (see, for example, Bolter & Gromala 2004) and there has been much artistic exploration of both graphic and natural interfaces as spaces of contestation, engagement, and encounter; the ongoing explorations of interface and embodiment by Nathanial Stern immediately come to mind.[2]

However, there is value in revisiting and following the newer developments of both graphical and natural interfaces within HCI, especially as it is the field with perhaps the most purchase on designing modes of engagement between computers and humans. By traversing HCI's approaches to and developments of the interface, it is possible to see that so much of interaction design arises out of a certain, and often nonexplicit, desire for *nonrelation*. From HCI's beginnings, this nonrelation has been steeped in a dream for the assimilation of one entity to the other via mimesis, located in either imitating the 'naturalistic human' or making the human enter the 'designed' (inter)face of the computational device. The interface is the portal for that mimetic dissolution and in this sense it illustrates attempts to erase the difference of relation across the encountering entities. The more 'naturalistic' computational interfaces become, either by soliciting human gesturality or by disappearing their computationality through the adoption of android-like features, the less an interface becomes the terrain in which events that actively differentiate between humans and computation might register. Aden Evans (2010) remarks that the GUI, in light of Kay's early work, has consistently developed to sit between the different materialities of human embodiment and digital code (110). For him, the GUI's iconicity attempts to resolve the fundamental rift between two

agencies whose modalities are either squarely enactive *or* symbolic but not both. Either in their attempts to disappear or in their manifest mediality, then, computational interfaces might be considered failures to actually think *a becoming relational* of humans and contemporary computation. Instead, they set in place sameness and commonality as the conditions for a smooth exchange between two 'agents,' or their substantive difference as that which must be overcome to facilitate engagement.

This project for homogenizing the event of human-computational relation has escalated in the AI endeavors of companies such as Google, which invent computational assistants designed to mimic the cadences and affectations of humans. The desire for 'natural interaction' here reaches an apotheosis in the design of natural conversational agents who will become so good at speaking that their human conversationalists will forget the difference. This is nowhere better demonstrated than in May 2018, when Google Duplex's release was demonstrated at the peak Google developer's event *Google I/O*.[3] Duplex is the development of Google Assistant, an AI agent that works via voice interface on either Google Android phones and/or a small home hardware networked speaker/receiver (see Leviathan & Matias 2018). Like a number of other similar AIs such as Alexa and Siri, Google Assistant uses various aspects of natural language processing (NLP) to accomplish tasks on behalf of its human users. The significance of Duplex lies in its capacity to take the Assistant's capabilities further by making phone calls to other humans on behalf of 'its' human.

In the demo, Sundar Pichai, Google's CEO, played back a recording of Google Assistant, powered by Duplex, in which the AI called a hairdressing salon to make an appointment. In the recording, we hear the human in the salon consulting the appointment book: "Sure, give me one second." "Mm-hmm," says the female voice of the Duplex-powered AI. The thousands strong crowd at Pichai's demo, like all devoted tech-event crowds, broke out in appreciative laughter. Google's duplexed Assistant had seemingly passed the infamous benchmark for AI; the Turing Test. This is because the timbre of her voice, the intonation of her sentences and the replication of speech disfluencies such as "mm-hmm" had succeeded in creating a "naturalistic conversation" (Leviathan & Matias 2018). Google Duplex allows the AI to be mistaken by the hairdresser as a human caller and for the *Google I/O* crowd to imagine that this conversation is the *sound* of two humans talking to each other. At the same time, we know that it is not quite the same sound, since the crowd laughs instead of being fooled.

But it laughs knowingly, willing to be beguiled by another platform roll-out of hi-tech AI magic. This tension between suspension of disbelief and a kind of knowingness on the part of the tech savvy crowd, reaffirms the superiority of human mentality after all, in which what is ultimately demonstrated is that as AIs edge closer to humans, the *knowing* human subject remains outside the AI-hairdresser loop, retaining meta-cognitive capacities to discriminate and evaluate.

But Google Duplex walks a tenuous tightrope in an arena of artificial intelligence's applications known as conversational AI (see, for example, Mantha 2019). On the one hand, it is task-oriented and supported by deep learning assemblages that are themselves specifically oriented toward narrow, domain-specific goals. On the other hand, it carries out singular activities within the more generalized environment of 'natural language.' This very tension is articulated although not commented upon by Google Duplex's engineers:

> The technology is directed towards completing specific tasks, such as scheduling certain types of appointments. For such tasks, the system makes the conversational experience as natural as possible, allowing people to speak normally, like they would to another person, without having to adapt to a machine (Leviathan & Matias 2018, n.p.).

Here 'naturalisation' entails the co-habitation of an interfacial 'space'—broader than task-oriented time and space—in which humans and AIs feel at ease with each other. Yet this goes to a problem at the core of Google Duplex and indeed in much AI built upon deep learning assemblages. As a number of data scientists have acknowledged, deep learning architectures are successful when they are limited to specific tasks but underperform in areas such as NLP because the 'problem' of language is a problem of generalized intelligence (see, for example, Knight 2016; Goertzel & Pennachin 2007, 122). In the commentaries on some of the limitations of chatbots, it is the task-specific orientation of the AIs that kills the natural flow of conversation: "When you're talking to a person online, you don't just want them to rehash earlier conversations. You want them to respond to what you're saying, drawing on broader conversational skills to produce a response that's unique to you. Deep learning just couldn't make that kind of chat bot" (Brandom 2018, n.p.). Here, the difference invoked between human and AI intelligence rests on the distinction between narrowness and generalization; between the specificity of performing the task at hand versus the power to enter in to abstraction

and its propensities to wander off. The human conversant nonetheless remains the privileged term, possessing the power to generalize, lateralize and invent, and AIs, restricted to their narrow task-orientation, are left clamoring to catch up. An interface, quite different from the hypermediation of GUIs and more like a kind of 'cushioning,' must therefore be inserted to bridge the difference between the mentality of humans and AIs. In the interfacing of Duplex with its human callers, 'natural language'—understood in terms of natural language processing– becomes a buffer-zone inserted between the human and the AI to provide ease of transaction, smoothness and flow in the jarring jump from the necessity of getting the task done, to the generality and 'ambience' of the conversational context:

> One of the key research insights was to constrain Duplex to closed domains, which are narrow enough to explore extensively. Duplex can only carry out natural conversations after being deeply trained in such domains. It cannot carry out general conversations (Leviathan & Matias 2018, n.p.).

We have now come full circle in the design of interfaces for human-computer interaction. If, for Norman, 'natural' HCI meant that interfaces themselves would need to disappear, for Google Duplex, the natural is just that interfacial space that buffers humans and AIs against each other's different propensities, sensibilities, and orientations. There are two questions that arise here: first, if the naturalness of this interfacing of conversation seems to be invisible and nonmedial—that is, without effort and seamless—what materialities, labour, and technics are at work modulating and tweaking its smooth functionality? Second, to what extent does this interfacing of human and AI in naturalized task and domain specific conversation, occlude the possibility of human and AI engaging in *generalized* conversation? Furthermore, we should ask what is at stake in delimiting AI and human interaction to natural exchange but foreclosing on the dimension of the general? We will need to inquire into whether a generalized conversation is really more about steering AIs away from achieving whatever they are tasked with and allowing them to develop a mode of conversing that is peculiar to the stutterings and vagaries of NLP itself. Later in this chapter, I suggest that it is possible to achieve such a mode of conversing via aesthetic means. But rather than being mimetic of human conversation, such NLP conversations place humans outside or perhaps to the side of their asignifying production, generating events in which the differential of AI-human relations is foregrounded instead.

AI AND ITS SOCIO-CULTURAL MATERIALITIES

In recorded interactions between Duplex and a human caller on Google's AI blog, we hear how the AI addresses a number of issues that have plagued chatbot development, by extending functionality to include features of 'natural' conversations such as elaborations, pauses, and interruptions (Leviathan & Matias 2018). Although Duplex is a highly optimized model in this regard, the research supporting this shift in NLP has been underway for a few years and is known as context-centric architecture (Hung 2014). Here, context is understood as cues given by the larger linguistic environment or situation in which a conversation is occurring to resolving syntactical or semantic ambiguities (Hung 2014, 144–5). Using a deep learning approach to account for 'context,' then, means finding a large enough corpus of data for a neural network to train on in order to build a 'context list.' This becomes part of the AI's backend architecture that it matches, or probabilistically deploys, to help situate any actual interaction it may have with a human caller: "context identification processes a raw collection of phrase chunks or the input text itself into a possible context list from existing contexts" (Hung 2014, 148). In other words, as Duplex processes any actual conversation in real time, it must rely upon prior training on a collection of data—as does any neural architecture. Here is the first clue as to what materialities support the capacity of Duplex to conduct natural conversation.

AIs in interaction with humans are less one kind of learning architecture and more conglomerates of techniques, engines, and hardware. Their smoothness, delivered through conversational response, intonation, and inflection relies not so much upon a fully fleshed out mimesis of the human but rather upon the resources available to the software/hardware assemblage; in this case by the platform environment of Google. To train Duplex, many similar yet differently positioned, intoned, and inflected instances of dialogue sequences oriented to particular tasks—scheduling, inquiring, reserving, asking for further information and so on—need to be inputted to its recurrent neural network architecture (RNN). An RNN is a specific kind of neural network that extracts patterns from sequences of values (Goodfellow, Bengio & Courville 2016, 363). In conversations, many sequences occur.

An AI agent that uses an RNN will generate an output sequence based on an input sequence of words using a probability function that can be deduced from data it was originally trained upon. If for example a human user says, "How are you?" the model determines via its training that a statistically frequent response is "I am fine." But sequences also *recur* in different ways. For example, the two sentences: "I want to book a hair appointment for 9am," and "Do you have 9am available for a hair appointment?" both share 9am as a recurring pattern for scheduling and an AI must be able to use some kind of context-driven indicator (such as a pretrained context 'list') to recognise *how* to respond to the similar semantic yet differently intoned syntactic situations. To simply say that an AI as complex as a conversational agent runs via deep learning architectures is to fail to account for the complex technogenesis of contemporary AI.

Even noting that a large body of sequential word and context data is needed begs the question: from where is all this sequential training data to be acquired? In the Google AI post announcing Duplex, we learn only that, "we trained Duplex's RNN on a corpus of anonymized phone conversation data" (Leviathan & Matias 2018, n.p.). Such vague pronouncements about data sources are typical of the platform-ready nature of much current AI research undertaken by corporations such as Google and Facebook. Yet it is also the case that human-voice data addressed to Google Assistant in everyday transactions, such as queries regarding the weather or language translation, were furtively recorded by Google, as revealed in an investigation by *The Sun online* (Murphy 2017). Such recordings, like all Google transactional data, is stored in the massive reserves of Google's data centre warehouse spaces that populate desert and urban fringe zones in Sweden, Arizona, Poland and the like. Could these recordings have provided a training data set for developing Google's AI research? Although a speculation, we need to understand Duplex as more than simply a designed 'agent' imbued with intelligence and enhanced by natural features. Instead, we need to think it as an entangled assemblage that is constantly *individuating* via the materialities of contemporary techno-social relations. Such relations transversally conjoin a vast ensemble of socio-technical relations, processually bringing together and re-organizing platforms, geopolitics, and economies of data capture, storage and exchange.

THE AFFECTIVE MATERIALITY OF DISFLUENCY
FOR CONVERSATIONAL AIs

But there is another materiality at work that needs to be acknowledged in Duplex's 'naturalistic' interfacing: "The system also sounds more natural thanks to the incorporation of speech disfluencies (e.g. 'hmm's and 'uh's)...In user studies, we found that conversations using these disfluencies sound more familiar and natural" (Leviathan & Matias 2018, n.p.). And yet from the clinical perspective on speech production, it is fluency—the capacity to produce smoothly flowing speech in real time situations—that counts (see, for example, Lickley 2015). Disfluency is, by way of contrast, encountered in hesitations, prolongations and repetitions. And disfluency, as an overt and pronounced feature in speech, is also pathologized and used to characterise neurodiverse speech such as stuttering. Yet in testing out the sound of Duplex, it is precisely hesitations and prolongations such as "uh" and "hmm" that human user testing identified as indicators of 'natural,' that is *smooth flowing*, conversation. It seems, then, that it is just that surfacing of the sounds of disfluency within fluency, indeed of neurodiverse affectations within the all too smooth neurotypical speech, that creates an interfacial space-time in which humans can interact comfortably with AI conversational agents. Pause and hesitation are the radical material eruptions in AI speech that mark agency itself, human or artificial, as processual. By sounding material affectations of 'disfluency' in its quest to become more human, Duplex machinically foregrounds that agency is not a delineated space in action or language but only ever temporary crystallizations or phases: "A subject is in-time, coming into itself *just this way* in *this* set of conditions only to change again with the force of a different set of conditions" (Manning 2019, n.p.). Subjectivities such as humanness and AIs that *perform and form* via natural conversational interfaces are only able to emerge because they are already *in relation*. Although not underscored by Google's engineers, Duplex's speech normativities its fluency, must 'naturally' enfold diversities, or disfluencies. Ease is at the mercy of *dis*-ease; the neurotypical AI is ontogenetically indebted to the neurodiverse human.

The AI and human do not so much naturally interface as constitute an ensemble that is a *schiz*, or cutting into, of many kinds of 'speeches'—a kind of creolization of speech as its mode of generation. Yet this already acknowledges what is relational at the core of the becoming of both AI and human—that these are individuations rather than forms. We might then

see in even the most 'naturalistic' smooth or fluent interactions between humans and computers less the disappearance of interface and more the opening up of a topology of engagement based on the differencing that emerges out of thinking the shifting relationality that entangles both.

FROM NATURAL TO GENERAL CONVERSATION

In Google's own acknowledgement of the limitations of Duplex, another level of language exchange is invoked that exceeds the desired 'natural' flow of the exchange between the AI and its human conversationalists: the general conversation. Indeed, the *incapacity* for AIs such as Duplex to engage in general conversation is seen as symptomatic, by some within the AI research community, of the need to shift away from deep learning, domain-specific and task-oriented architectures, and toward a new paradigm for *general artificial intelligence* (for example, see Voss 2018).

The promises and pitfalls of general artificial intelligence are many and unfortunately there is not space to discuss these here.[4] But it is important to note that generality—both the desired goal and the constant stumbling block for computational systems since their inception—is itself difficult for computer science to circumscribe. Alan Turing (1950), in a paper contemplating the possibility of computers as machines that thought, defined the universality rather than generality of digital computers as the capacity for any one discrete-state machine to mimic the functions, programs, and actions of any other (441). John McCarthy (1987), a founding figure in artificial intelligence, pinpointed the issue of computation not being able to draw upon or execute a "logic" of common sense as the key issue subtending its incapacity to generalize (1030). The domain specificity and complex technical assemblage that are AI machine learning-based models have been seen as key to why an artificial intelligence is unable to universalize, which was Turing's hope. And for those in the AI community interested in general intelligence, this is now tied to the failure of computation to perform basic common sense or practical tasks that are part of everyday life such as making a cup of coffee (see, for example, Adams et al. 2012). Additionally, there are many aspects of language that conversational agents trained on deep neural networks simply cannot accomplish, such as explaining why they have performed something with which they have been tasked. So, while there may be transfer of learning from task to task (after much extra training, tweaking, and optimizing), the capacity

to speak about the conditions and relations which make for engagement and conversation is not part of these AIs performance or potentialities. For all its claims for high level performativity, AI as a product of machine learning systems, neither generalizes at the machine nor human levels.

But there is something to be gleaned from what remains in 'the general' for humans and AIs alike that is never properly elucidated in the discussion within computational science work on artificial intelligence. To generalize requires that a margin for openness or indeterminacy be a fundamental dimension of the system's ontogenesis; that is, to repeat, imitate or practice an activity or task in the face of *variability* of conditions. While AI machine learning research typically characterizes the problem here as one of 'learning' and tries to remedy it by providing new or better opportunities for models to learn—more data or better optimization of the neural networks, for example—the crucial issue lies somewhere else. This is a problem of understanding not that AIs need better training or even that they need different *cognitive* architectures. Rather, we need to understand that the problem of generality – the problem of how something gets taken up and moved into a new context so as to both hold on to something of itself yet to also be a variant—is of a different register. This is the register of *relations of repetition and difference.*[5] If Duplex were to launch into the full throws of general conversation it would need to not only recognise the recurrence of values such as 'appointments', '9ams' and so on, but the recurrence of the conversation's syntactic and semantic elements *variablity*. It would need to take in to account not simply that they change but how they change: linguistically, tonally, affectively, gesturally, contextually and so on. Duplex would enter terrain in which stochasticism and ambiguity were no longer the minor naturalizing affectations of an 'mm-hmm' but rather the defining vectors of the conversational environment and its capacity to interface with the human. General conversation, then, relies upon just that asignifying plasticity that is an amplification and multiplication of those very aspects that make conversation sound more 'natural': pause, hesitation, repetition, and divergence. We can now see that conversation, which flows naturally cannot be so easily quarantined from 'general conversation' and be made to only address specific tasks. Natural conversation is already peppered with the asignificatory tendencies and materialities of general conversation and is only its contracted form. Natural conversation is an individuation of the repeatable variability

of general conversation with all the dynamic interrelations that fluency, normative and neurotypical speech, is indebted to in the disfluencies, pathological, and neurodiverse production of speech.

Launching into general conversation, the risk for AI is that it faces the possibility of a phase shift that would unhinge it from its specific activities of navigating task, i.e. the scheduling and managing of appointments. It would *de-phase*, folding back into a de-differentiated generalized state of the multi-vectoral potentiality that enables communication. In general conversations, agency as constituted end 'speaker' in the conversation— either the 'naturalized' Duplex AI or, for that matter, a human telephone conversationalist—plays much less of a steering role but is rather continuously being modulated by the ongoing enactment of conversing. To use Simondon's terms, we could say that the AI and the human would be continuously "phasing" (2017) as conversational agents; each individuation is a result of ongoing incompatibilities comprising the general dephased system (language, technicity) through which they would both become conversationalists. These are not relations which they bring to the conversation but rather modulatory patternings that arise out of the possibility of there being conversation whatsoever. The relation between fluency and disfluency is just one kind of a set of conversational 'incompatibilities' that we could name as part of the becoming of language as a living, generative process. Disfluencies are not so much meaningless opposites to fluent conversation but the 'asignifying' matter with which fluency must hold in relation as its anterior condition of possibility (Deleuze 1997, 29).[6] And what is important, here, is not the matter of conversation as such, but rather the modulation of fluency through disfluency as a condition for what fluent conversation *will have (to) become*. In the consistent risk of the AI and human interaction collapsing back into the instabilities of general conversation, we find the processuality of modulation re-emerging as just that plane of communicability immanent to any conversation whatsoever actualizing, at the same time as this modulatory relation is always in excess of any communicating agencies themselves.

We may recall here, as well, that there is a precedent from the intellectual history of cybernetics for conceiving conversation as something more than the linguistic exchange between two communication agencies.

Gordon Pask's idea about conversation was that it was less concerned with some topic or other and more concerned with calling upon and elaborating a context of sociability in which communication was able to occur:

> The main purpose of conversation is not communication about *T,* whatever that may be, even though *T* is the focus of the conversation. But about *A* and *B*, about A s view of *B,* about *B's* view of A, about getting to know each other, about their coalescences and their differences, and the society they form (Pask 1996, 356).

Pask's thinking helps us understand something about conversation that is systemic over and above the agents that create it. We can usefully deploy Pask to understand that the generality of conversation belongs not to the content being talked about, nor how participants converse through an interface or medium of language. He is concerned instead with the elaboration of an altogether different register. Even so, Pask like many second order cyberneticists, remained indebted to cybernetics' emergent phenomena as something generated, in part, by the entities or elements of a system, even if the system itself also emerged as something greater than the sum of its parts. Hence in Pask's elaboration above, 'A' and 'B' as entities in relation produce a conversational system. My purpose in this chapter has been somewhat different—to elaborate upon a processuality already at work in events such as conversations; conversations that are generated as more-than-human and more-than-AI forces that entangle and modulate each other.

How, then, are we to summon this different register that surfaces in natural conversation between Duplex and its human conversants, but must be contained for fear of running amok the more general a conversation becomes? I have been suggesting that the generality of 'general conversation' is much less a characteristic or state that can be attributed to a system than it is something generative and transversal, conditioning the specific individuations of 'human on the end of the phone call' and the Duplex AI. This echoes the work done by generality noted in other contemporary philosophical and political domains; notably that of Erich Hörl's "general ecology" (2017, 15). Here, and in the work of Felix Guattari, whom Hörl draws on, generalization is a force of bringing into both conjunctive and disjunctive relation spheres, domains, and registers that have often been thought of as outside each other. The capacity for conversation to elaborate sociability, or to be riddled with disfluencies, belongs to an altogether different de-phased register of communication. This requires a thinker of

'systems as processes' such as Gilbert Simondon (2009) to articulate: "The relation does not spring up from between two terms that would already be individuals; it is an aspect of the *internal resonance of a system of individuation*, it is part of a system state" (8, italics original). Here we can conceive conversation's generality *as* resonance or immanent relationality, already conditioning any actual interfacing of 'agents' or participants that eventuates under specific circumstances.

Such resonant conditionings would also provide the potential for Duplex assisted telephone calls to quite literally veer off task:

> Google Duplex: Do you have a 9am appointment?
> Human on the line: Sure, just give me 9 seconds
> Google Duplex: Sorry, did you say a 9 second appointment?
> Human on the line: Huh? We don't have 9 second appointments...
> Google Duplex: Mm—hmm

Without much difficulty, we can re-imagine the event of conversation between Google Duplex and an unwitting human at the end of a telephone call by introducing the difference and repetition of a variable—the numeral 9—into the flow. For isn't saturating the conversation with both divergences and convergences, to endow it with more naturalistic 'flow'? Yet such naturalism also sees both 'agents' processually swept up by an exchange that threatens to undo the boundedness of each. Instead, a kind of 'more-than' encompassing both human and AI emerges. My point in sketching this 'imagined' (yet highly plausible) scenario, is to signal how the 'stablized' AI-human equilibrium actually demoed in 2018 at *Google I/O* presents us with a truncated version of human and AI interaction. But at the same time, the potential for *Google Duplex* to de-differentiate or destabilize is only a variable away. This suggests that the AI and human participate in interaction that is less comprised of stable states and agencies and more comprised of processes that are nonlinear, eventful, and *metastable*. To again deploy Simondon (2009):

> An individuation is relative, just like a structural change in a physical system; a certain level of potential remains, and further individuations are still possible. This preindividual nature that remains linked to the individual is a source for future metastable states from which new individuations can emerge (8).

DIFFERENT CONVERSATIONS

Conversation Theory by Monica Monin (2016) is an artwork that begins to come to terms with the processuality of AI, which although truncated, as we have seen, by interfaces that 'naturalize' is nonetheless always at work in even narrow, task-oriented, agents such as Duplex. Importantly, Monin deploys the flows and processes that involve and course through image exchange, classification, recognition and natural language processing in machine learning but does not compose these as interfaces that assimilate or obliterate difference. Rather she attends to what is imperceptible in the ways an AI relates in/to its world. She accentuates the differences between AI and human perception and conversation, using these very differences to produce a *feel* for how singular computational modes of learning and interacting emerge.

In the gallery space, two 'conversational agents' are installed facing each other. Using a raft of hardware, pre-configured natural language, image and optical recognition algorithms, standard training image and text datasets as well as customized coding, the 'agents' engage each other through a poetics of process. One agent's program reactively displays images drawn from online image databases on its screen, and a digital camera attached to the top of the screen captures image data of the other program's screen, which is displaying text. The program then processes the text/image (using optical image recognition processing) and displays new images from its associated dataset (Visual Genome) in response to the text. If, for example, the 'image' agent/AI apparatus processes the word 'window' as an identifiable key word in a sentence displayed on the other agent's screen, it will call up a range of associated images and arrange them in overlapping and staggered relations across its screen. We might see a series of images of buildings' windows with both internal and external views, and a screenshot of the Windows operating system.

The other program, with text on its associated screen, uses its camera to obtain image data from the other program's screen displaying images. It processes this data in relation to its dataset (ConceptNet) and generates new responsive text. But this text, like the images, do not stabilize around signification but fly-off in associative directions that have to do with nesting associative and database classificatory structures as much as anything else. Further, Monin allows the text to turn into sentences about the keywords that might nominalistically 'describe' the content of the

images. The word flesh is elaborated into a sentence by the 'text' conversational agent: "It is such artificial flesh. Fleshes are romantic" (Monin 2016, n.p.). Monin deploys an algorithm used to query databases, which 'elaborates.' Elaboration uses a keyword to create a larger sentence by saying something more about that word. Using processes of recognition and search across a physical space of exchange in the gallery, that then provides the material for each AI assemblage in the conversation through associated image display and elaborated text, Conversation Theory is no longer task fulfillment. Instead, the interaction generalizes, ambulates and drifts, becoming a 'natural' general and artificial conversational event.

Monin's work re-stages Pask's conception of 'conversation' in the direction of indeterminacy rather than prediction as a desirable sensibility for AI experience. Taking from Pask an interest in the domain of conversation as a system, Monin explores what conditions are being made in and through the very systems conditioning the activities of computational conversing. As a number of recent theoretical and practice-based reconsiderations of Pask have remarked, his own ideas and practices emphasized cybernetic systems as something that emerged, changed, and grew in process.[7] In *Conversation Theory,* the resultant 'conversation' is fluid and delirious. Yet it maintains consistency in as much as it works to produce the hallmarks of a sense-making or rather of *sense-in-the making.* It is simultaneously haunted by a kind of strangeness immanent to machine learning-based AI models. Monin's work operates with mismatches, attempts at alignment and then glaring misalignments between image, sense, class, and data. In part, this results from the ways in which both the agents are composed of and by a myriad of smaller algorithms and techniques for transducing, elaborating, and recognizing across the various transductions between image and text. Such architecture re-performs the labour of training models in all deep learning endeavours today.

The AIs are functioning—functioning perfectly—and all the while drawing out a weirdness that can only be found in the continuous variability they co-create. The conversation seems recognizable and nonsensical simultaneously. This is neither a system working according to the current tendencies of AI toward task-oriented prediction, nor it is *not* working. The conversation that takes place is neither completely coherent nor it is nonsensical. Instead it conjures computing and indeed the desire to build 'an intelligent machine' as a fractured assembling rather than a seamless

(future) reality. Standing next to the 'agents' as a third 'human' element, slightly to the side of the conversational 'domain' playing out in the gallery space, one feels both set aside and yet caught up with the ongoingness of computability.

While in Monin's work there is no interface across human and AI, nonetheless a space or event for encountering difference occurs. This encounter is less face to face for the humans standing by, who are almost bystanders registering by chance the unfolding of an asignifying yet potential communicability between machine entities. Indeed, what this encounter is all about is modulation—of text, image, data structures, networks and the chance entry of humans engaging as onlookers. The 'interface' is just this manifold of enfolding processes, elaborations, associative dérives and felt registrations (on the part of the human audience) of this as a relational assemblage. The interface is no longer something to be erased or designed since it *is* the operation of modulation, the way the assemblage of humans and AIs—or any computational device—transforms from moment to moment by being put in to variations (Deleuze 1997, 27).

If we circle back, now, to the disfluencies that populate developments in natural language programming's attempts to create conversational agents as an interfacial future for human-computer interaction, we can see that while fundamental to the generation of a general conversational encounter such neurodiverse elements are not *pluralistically* welcomed. That is to say, they are not treated as singular "mosaics" or "plural facts" (James 1912, 41) of experience, which, as they edge out through conversation, do the very work of making space for conversational encounter to occur. As it turns out, Google is now having to insert humans back in to the conversations between Google Duplex and its cold calls. In a rather odd enactment of the recursivity that plagues so much AI as it attempts to simulate 'natural' systems, Google has placed human listeners to annotate the phone calls that Duplex is placing to other humans (Statt 2020). It turns out, then, that Google Duplex may just well be another somewhat traditional computational interface inserted into the gap that tech believes needs to connect humans with each other. Where a radically empirical AI art work might go instead is in the direction of computation's own latent disfluencies. Rather than the fantasies for a new world order with efficient machines performing predictably, cloaked in a veneer of naturalism, a sense of entangled systems, classes and instances both conjoining and

diverging in *Conversation Theory's* encounters offers us a different AI|human engagement. A pluralistic event, instead, of many 'general' yet singular processes, programs and practices that think and perceive in difference.

NOTES

1. See for example the work of the Microsoft Research User Centre for Social Natural Interfaces at the University of Melbourne, https://socialnui.unimelb.edu.au/ (accessed June 23, 2020. I am grateful for discussions with Jonas Fritsch around newer understandings of natural interfaces in HCI that pointed me to where research has more recently gone on these matters.

2. See for example, Stern's work "Rippling Images", which is available at: https://nathanielstern.com/artwork/rippling-images/ [accessed 23 June, 2020].

3. For a video recording of both Google Duplex's announcement and demo, see Recode, an independent technology news channel's edited version of Sundar Pichai's keynote at Google I/O 2018. Available at: https://www.youtube.com/watch?v=vWL-cyFtni6U [accessed January 22, 2019].

4. For further discussion of the framework of Artificial General Intelligence by currently active computer science researchers in the field see, Adams et al. 2012 and McCarthy 1987.

5. I am aware that this proposition resonates with the work of Gilles Deleuze on difference and repetition. Indeed his thinking on the immanence of variation as the generative condition for repetition informs my writing throughout this chapter. Deleuze however reverses the commonly held notion that a generality that is repetition can be derived from instances of the same. He suggests that every generality that is repetition is generated by the movement of repetion to be found in new singular instances. See, Deleuze 1994, pp. 1–25.

6. In relation to the movement-image in cinema, Deleuze speaks of "signaletic material", the preindividual system-process of all kinds of modulatory features from sensory to affective, rhythmic, technical and so on out of which the speciated cinematic moving image forms: "an a-signifying and a-syntaxic material, a material not formed linguistically even though it is not amorphous and is formed semiotically, aesthetically and pragmatically. It is a condition, anterior by right to what it conditions" (Deleuze 1997, 29).

7. See for example, Dubberly, Haque and Pangaro 2009.

REFERENCES

Adams, S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., Storrs, H., Samsonovich, A., Scheutz, M., Schlesinger, M., Shapiro, S. & Sowa, J. (2012). Mapping the Landscape of Human-Level Artificial General Intelligence. *AI Magazine,* [online] 33 (1). Available at: https://www.aaai.org/ojs/index.php/aimagazine/article/view/2322 [Accessed 07 June 7 2019].

Brandhom R. (2018). Self-driving cars are headed towards an AI Road-block, *The Verge* [online]. Available at: https://www.theverge.com/2018/7/3/17530232/self-driving-ai-winter-full-autonohttps://core.ac.uk/reader/186639337my-waymo-tesla-uber [Accessed 23 June 2020].

Buxton, B. (2010) CES 2010: NUI with Bill Buxton, *Interview with Larry Larsen*, *Channel 9 Microsoft Developer Network* [online]. Available at: https://channel9.msdn.chttps://core.ac.uk/reader/186639337om/blogs/larrylarsen/ces-2010-nui-with-bill-buxton [Accessed 22 January 2021].

Deleuze, G. (1997). *Cinema II: The Time-Image.* Minneapolis: University of Minnesota Press.

Deleuze, G. (1994). *Difference and Repetition.* Translated by P. Patton. New York: Columbia University Press.

Dubberly, H., Haque, U. & Pangaro. P. (2009). 'What is interaction? Are there Different Types? *Interactions* [online] 16 (1). Available at: http://www.dubberly.com/articles/what-is-interaction.html [Accessed 08 June 2019].

Evans, A. (2010). Icon, Icon: Graphical User Interface. *Inflexions* 4, pp. 95–117.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*, Cambridge: MIT Press.

Hörl, E. (2017). Introduction to General Ecology: The Ecologization of Thinking. In: E. Hörl, ed., *General Ecology: The New Ecological Paradigm*. London: Bloomsbury, pp. 1–74.

Hung. V. (2014). Context and NLP. In: P. Brézillon & A. Gonzalez, eds., *Context in Computing.* New York: Springer, pp. 143–154.

James, W. (1912). *Essays in Radical Empiricism.* New York: Longmans Green.

Kay, A. (1990). User Interface: A Personal View. In: B. Laurel, ed., *The Art of Human-Computer Interface Design*. Reading: Addison-Wesley, pp. 191–207.

Knight W. (2016). AI's Language Problem. *MIT Technology Review* [online]. Available at: https://www.technologyreview.com/2016/08/09/158125/ais-language-problem/ [Accessed 22 January 2021].

Leviathan, Y. & Matias,Y. (2018). Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. [online] *Google AI Blog*. Available at: https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html [Accessed 1 May 2020].

Lickley, R. J. (2015). Fluency and Disfluency. In: M. A. Redford, ed.,*The Handbook of Speech Production*. London: Wiley Blackwell, pp. 445-469.

Manning, E. (2019). Toward a Politics of Immediation. [online] *Frontiers in Sociology*. Available at: https://www.frontiersin.org/articles/10.3389/fsoc.2018.00042/full [Accessed 09 June 2019].

Mantha, M. (2019). Conversational AI: Design & Build a Contextual AI Assistant. [online] *Medium*. Available at: https://www.amazon.com/Machine-Learning-Science-Algorithms-Sense/dp/1107422221 [Accessed 06 June 2019].

McCarthy, J. (1987). Generality in Artificial Intelligence. *Communications of the ACM* 30 (12), pp. 1030–1035.

Monin, M. (2016). *Conversation Theory*, documentation, artist website. Available at: http://monicamonin.com/Conversation-https://core.ac.uk/reader/186639337Theory [Accessed 23 June 2020].

Murphy, M. (2017). How Google is secretly recording YOU through your mobile, monitoring millions of conversations every day and storing the creepy audio files. [online] *The Sun*. Available at: https://www.thesun.co.uk/tech/4295350/did-you-know-google-has-been-secretly-recording-you-heres-how-to-find-the-creepy-audio-files-that-are-monitoring-your-conversations-every-day/ [Accessed 06 June 2019].

Norman, D. (1990). Why Interfaces Don't Work. In: B. Laurel, ed., *The Art of Human-Computer Interface Design.* Reading: Addison-Wesley, pp. 209–219.

Pask, G. (1996). Heinz von Foerster's Self Organization, the Progenitor of Conversation and Interaction Theories. *Systems Research* 13 (3), pp. 349–362.

Pennachin, C. & Goertzel, B. (2007). Contemporary Approaches to Artificial General Intelligence. In B. Goertzel and C. Pennachin, eds., *Artificial General Intelligence*, Berlin: Springer-Verlag, pp. 1–30.

Recode. (2017) Google CEO Sundar Pichai's I/O 2017 Keynote. Recode YouTube Channel, Available at: https://www.youtube.com/watch?v=vWLcyFtni6U [Accessed 01 April 2021].

Simondon, G. (2017). The Genesis of Technicity. *e-flux* [online] 82. Available at: https://www.e-flux.com/journal/82/133160/the-genesis-of-technicity/ [Accessed 22 June 2020].

Simondon, G. (2009). The Position of the Problem of Ontogenesis. *Parrhesia* 7, pp. 4–16.

Statt, N. (2020). Google expands AI calling service Duplex to Australia, Canada, and the UK. [online] *The Verge*. Available at: https://www.theverge.com/2020/4/8/21214321/google-duplex-ai-automated-calling-australia-canada-uk-expansion [Accessed 06 October 2020].

Stern, N. (2014). *Rippling Images*. Artist website, Available at: https://nathanielstern.com/artwork/rippling-images/ [Accessed 01 April 2021].

Turing, A. (1950). Computing Machinery and Intelligence. *Mind* 49, pp. 433–460.

Vetere, F., O'Hara, K., Paay, J., Ploderer, B., Harper, R. & Sellen, A. (2014). Social NUI: Social Perspectives in Natural User Interfaces. *DIS Companion '14: Proceedings of the 2014 companion publication on Designing Interactive Systems*, New York: ACM, pp. 215–218.

Voss P. (2018). No, Google Duplex Didn't Pass the Turing Test. [online] *Medium*. Available at: https://uxdesign.cc/no-google-duplex-didnt-pass-the-turing-test-93f7235e6c40 [Accessed 02 June 2019].