Duets Ex Machina:

On The Performative Aspects of “Double Acts” in Computational Creativity

Tony Veale, Philipp Wicke and Thomas Mildner

School of Computer Science, University College Dublin, Ireland.

Abstract

We humans often compensate for our own weaknesses by partnering with those with complementary strengths. So fiction is full of characters who complete each other, just as show-business thrives on successful double acts. If it works for humans, then why not for our machines? The comparative strengths and weaknesses of different CC systems are well-documented in the literature, just as the pros & cons of various technologies or platforms are well known to the builders of these systems. A good pairing does more than compensate for the weaknesses of one with the strengths of another: it can find value in disparity, and deliver results that are beyond the reach of either partner alone. Here we consider the pairing of two CC systems in the same thematic area, a speech-based story-teller (with *Alexa*) and an embodied story-teller (using a *NAO* robot). Working together, these two compensate for each other’s weaknesses while creating something of comedic value that neither has on its own.

In It Together

The mythology of human creativity often paints a romantic image of the solitary creator, toiling against the status quo to fulfil a singular vision. But our creativity narratives also prize the results of successful partnerships. One can list a long line of inspired double acts, from Crick & Watson – or, indeed, Holmes & Watson – to Lennon & McCartney, in which a duo’s differences count as much as what they share. If good partners learn to overcome their differences, creative partners learn to *exploit* their differences, and no where is this truer than in the classic comedy double act.

Henri Bergson (1911) has argued that mechanical rigidity lies at the root of all comedy. We become risible when we are reduced to predictable machines and act unthinkingly in the pursuit of conformity. Yet Freud (1919) has also argued that when machines take on human characteristics, such as the semblance of free will, they appear *uncanny* or *unheimlich*, sources of terror rather than agents of comedy. Our CC systems can be nudged either way on this continuum of the *canned* to the *uncanny*, to play their presumed stiffness for laughs or to transcend this rigidity by acting unpredictably. Most comedy double acts do both, with one partner serving as a defender of conformity, the other as an agent of chaos. In their interactions we see glimpses of the *relief theory* of humour as espoused by Lord Shaftesbury (1709): the free agent shows a nimbleness of spirit and an ability to break free of its constrainer, the rigid partner. The latter looks stiff and inadequate, following Bergson, while the former looks graceful and agile, following Shaftesbury, so both theories together give us twice the reason to laugh. Famous comedy acts from Stan Laurel and Oliver Hardy to Bob Hope and Bing Crosby to Dean Martin and Jerry Lewis all worked in solo acts first, as singers, actors and comics, before coming together to reap the benefits of their obvious friction and complementarity (see e.g., Epstein, 2005).

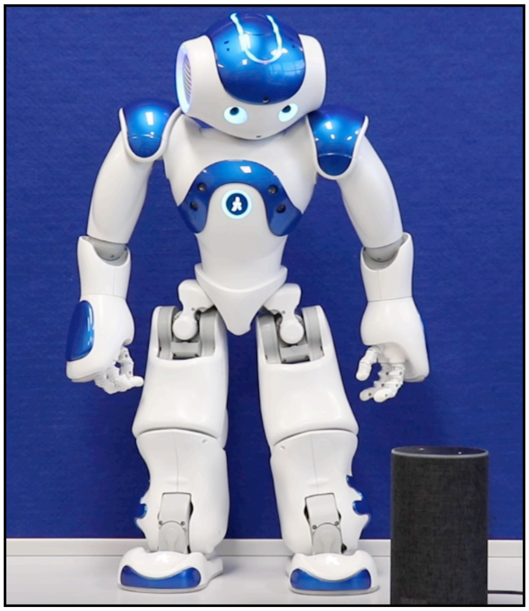

Figure 1. The Walkie-Talkie double act of NAO and Alexa.

When friction sparks comedy, each part of the duo acts as a tacit rebuke to the other; the straight guy is *too* rigid, and the funny guy is *too* unpredictable. This it not simply a matter of how material is divided up and performed, but an issue of substance in the material itself. For laughter can be wrung from a meta-critique of the act’s artifice, as when a ventriloquist’s dummy says to its human partner, “Why is it that every time *I* shout, I get sprayed with *your* spittle?" A ventriloquist and his dummy are two roles played by one performer, which an audience willingly sees as two agents. Each, however, represents a different part of the psyche of a single idealized performer, the *super-ego* (ventriloquist) and the *id* (dummy). One works to keep the other in check, and fails, but it is in this failure that the comedy takes root. Computationally, the fact that one CC system works as two gives us a convenient abstraction for a comedic double act. A single system, coordinated using backstage computation, controls two agents of conflicting temperament that create comedy through their interactions on the same shared task.

The rest of the paper puts flesh on our scheme, in which a *NAO* robot and an *Amazon Echo* are used to implement a story-telling double act (see Figure 1). We show how their complementary strengths and weaknesses are exploited to make a virtue of failings that would be nigh on intolerable in one alone. Our aim is to turn each platform into an agent with its own personality, rather like the bickering droid duo R2D2 and C3PO in *Star Wars*. The next section presents a story-telling skill for the *Echo*'s speech-driven *Alexa* front-end, before an embodied, NAO-based robot story-teller, for the same space of computer-generated stories, is described. This story space is built using *Scéalextric* (Veale, 2017), a story-generation CC system ideally suited to the creation of shaggy dog tales that put familiar faces in comical settings. We present an advance to *Scéalextric* that imposes a global shape on its plots and supports the generation of narratives of more than two key characters. These tales are performed by a double-act, named *Walkie Talkie*, of *Alexa* and a *NAO* robot, in which *Alexa* narrates a tale as the *NAO* embodies its actions. Coordinating their interactions is a blackboard architecture that obviates the need for any overt communication, yet we focus here on the ways in which their joint performance is built upon the interplay of the spoken and the physical. We show how the clear-spoken *Alexa* can act as the straight guy while the clownish NAO can be her foil. The paper concludes with a discussion of related work and a map of future directions for the *Walkie Talkie* double act.

Alexa in Storyland

Though the browser was once our principle means of web access, and a convenient platform for offering CC systems as services, the advent of devices such *Amazon Echo* and *Google Home* has given CC systems an alternate route into our homes. Consider *Alexa* (Amazon, 2019) a speech-activated 'genie' that answers our questions, fetches our data and controls our music, lighting, heating and more. *Alexa*'s repertoire of skills is easily extensible, allowing developers to add new 'skills' for the delivery of content that may well be machine-generated. So, in addition to fetching factoids, weather updates, recipes and canned jokes, *Alexa* can be extended to create riddles and poems, and even stories, on demand. Yet, since story-telling is an art, a narrative 'skill' for such a device must exploit all of the affordances, and sidestep all the impediments, of the technology concerned.

Each *Alexa* skill is opened with a voice command, as in "Alexa, open the narrator." Once inside an open skill, users may use a variety of pre-defined speech patterns to achieve a given end. Our story-telling skill, *The Narrator*, can be requested to tell stories on a specific theme, as in "Alexa, tell me a story about love" or "Alexa, tell me a Star Wars story." Once a topic is extracted, the skill fetches an apt story from a large pool of pre-generated tales. *Alexa* skills may call on a variety of Amazon Web Service components, such as an AWS database, to store the knowledge / data of a CC system, so that creative artifacts can be generated by the skill on the fly. However, as each skill must package its response in a fixed time (8 seconds) before the current task is aborted, we prefer to use *Scéalextric* and the *NOC list* (see Veale, 2016) to pre-generate hundreds of thousands of stories in advance, storing each with appropriate indexes to facilitate future thematic retrieval. The step function of the AWS pricing model is rather steep, and if one is not careful about data usage a skill can jump from costing nothing at all to costing hundreds of dollars per month. Yet, as shown in Veale & Cook (2018), pre-built spaces of content offer a clean and efficient approach to the separation of creation, curation, selection and delivery tasks. In our case, we opt to store our large story space on the Web, and *Alexa* dips into different parts of this space using topic-specific URLs.

The *Alexa intent model* is powerful and flexible, but can seem counter-intuitive from a conventional programming perspective. Accommodations must be made to repackage a CC system as an *Alexa* skill, and the process is not unlike building a ship inside a bottle. Yet the payoffs are obvious: *Alexa* has excellent speech comprehension and generation capabilities for a consumer device; the former is robust to ambient noise while the latter sounds natural, if prim, so in a story-telling double act, *Alexa* is destined to play the role of straight guy. Her formal disembodied voice reminds us of *HAL 9000* and any number of sci-fi clichés about rigid machines, making *Alexa* a natural fit for Bergson's theory.

Her rigidity extends to a lack of reentrancy in how skills are executed. *Alexa* retrieves whole stories from her online story space, choosing randomly from tales that match the current theme to produce a single, composite speech act for a narrative. Users can interrupt *Alexa* to stop one story and request another. but *Alexa* cannot segment a narrative into beats of a single action apiece, and articulate each beat as a distinct response to the user. That would require her to re-entrantly jump in and out of her narration intent, at least if she needs to execute other tasks between beats. This makes uninterrupted story-telling difficult to align with the actions of parallel performers, as choreography demands chunking, communication and reentrancy. This is not a problem when *Alexa* works alone; she simply narratives her chosen story in a single continuous speech act. But when she must work with a partner, such as an embodied robot, this double act requires her to articulate the story one beat at a time, and wait for a prompt from a human – such as "yes," "go on," "uh huh," "really?" or "then what?"– to proceed. In the gap opened by this interaction, *Alexa* is free to communicate with her partner and cue up the partner's enacted response.

For long stories – and our improvements to *Scéalextric* produce tales of multiple characters and many beats, as we describe in a later section – the need for an explicit prompt between each beat is an onerous one. Without this prompt, *Alexa* can do little, and her partner will also lack the cue to perform, bringing their double act to a standstill. However, as with human double acts, this rigidity of form is itself an opportunity for meta-comedy. When *Alexa* becomes stuck, as when it fails to receive or perceive a prompt, her partner offers a wry comment on the situation. These meta-actions constitute the double act's shared mental model (Fuller & Magerko, 2010), perched above its content-specific *domain* model, allowing an act to be more than the sum of its parts. This setup is not so different to a human ventriloquist with an insolent dummy: while *what* is said is vitally important, *how* it is said and enacted is a source of humorous friction.

Apocalypse NAO

Alexa has a voice but no body. The *NAO* has both a body and a voice, but the limitations of the latter often struggle to transcend the former. Although the NAO's capacity for physical movement is a major selling point, its gestures can be so noisy as to dominate its twee vocalizations. Moreover, *NAO*'s processing of speech is rather limited in comparison to *Alexa*'s, and frequently forces its human interlocutors to vehemently repeat themselves on even short commands. So a pairing of *Alexa* & *NAO* makes sound technical sense for a language-based task like storytelling, since *NAO*'s utility as an embodied storyteller has already been demonstrated by Pelachaud *et al.* (2010) and Wicke *et al.* (2018a,b). As the latter uses the *NAO* to tell computer-generated stories, we use that work here as a foundation for our CC system.

With a humanoid body offering 25 degrees of freedom, a *NAO* can pantomime almost every action in a story. Wicke *et al.* (ibid) built a mapping of plot verbs to robot gestures, so that their robot has an embodied response to each of the 800 verbs in the underlying story-generator, the *Scéalextric* system of Veale (2017). Two variants of the storyteller are presented. Wicke *et al*. (2018a) describe how pre-generated *Scéalextric* stories are selected at random and enacted with a combination of speech – to articulate each beat of a story – and gesture, to simultaneously pantomime the action. The chosen story is retrieved using a vocal cue from the user, who provides a topic index such as "love" or "betrayal." In Wicke *et al*. (2018b), the user exerts more control over the shape of the story. In this variant, the robot uses the causal graph connecting *Scéalextric* actions to generate questions that require users to probe their own experiences and offer yes/no answers in response. The answers allow the robot to navigate the space of *Scéalextric* stories to build a tale that is a bespoke fit to the user's tastes. However, each variant works solely at the content-level, using a domain model to map directly from generic story verbs to robot capabilities.

A storyteller transcends its domain model – its model of what constitutes a story – whenever it shows awareness of itself as a teller of the tale. This is storytelling taken to the meta-level, in which a teller acknowledges its dual status as a protagonist who *lives* the tale via physical actions and an omniscient narrator who relates the tale via speech acts. The domain model ensures the effective communication of *character-to-character* relations, whilst the meta-model is responsible for *teller-to-audience* relations, as well as, for a double act, *teller-to-teller* relations. Of the two, the domain model is the most immediate, and has received the greatest attention from researchers. Pantomime is the obvious basis for a robot's domain model, but tellers can take an abstract view of events without wandering into the meta-level. For instance, folowing Pérez y Pérez (2007), a teller can track the disposition of characters to each another. In *Scéalextric* stories of just two characters using a finite number of plot verbs (approx. 800), it is feasible to mark each action as to whether it tends to promote closeness or distance. So, *love*, *respect* and *trust* are verbs that bring closeness, while verbs such as *insult, betray* and *suspect* each increase distance. A robot teller can assign each character to a distinct hand, so that as the story progresses, the horizontal movement of its hands conveys the conceptual distance between characters.

The meta-model of a storyteller recognizes that there are many ways to exploit the domain model to convey a story. Montfort's *Curveship* system for interactive fiction (2009) shows how a meta-model can alter the dynamic of a tale by opting to focalize one character over another, or by switching between narrators and rendering styles. Montfort *et al.* (2013) use a blackboard framework to integrate their storytelling system with a metaphor generator whilst exploiting the affordances of the *Curveship* meta-modal. The domain model is responsible for *in-world* reasoning about a story, so only the meta-model acknowledges the existence of the audience, other performers, and the artifice of the process. Often, however, the distinction between domain- and meta-models is a subtle one. To an audience, there may be little difference between a robot pantomiming the reactions of other characters to a specific act – for example, by reacting with surprise or the disappointed shake of a bowed head – and gesturally signifying its own reaction as a narrator. In the final analysis it matters little if the audience can tell the domain- and meta-models apart, as long as the story is told with aplomb. Nonetheless, a meta-model works best when it augments rather than supplants the domain model. When an agent is aware of its role, it can act as a character or as a narrator or even as an audience member if it serves the tale.

The meta-model is dependent on the domain model for its insights into the story, to e.g., determine which parts are tense and dramatic or loose and comedic. With such insight an embodied teller can react appropriately to its own story, by feigning shock, joy or even boredom in the right places. In a double-act, these reactions must be coordinated across performers, so that they are seen by the audience not just as responses to the story but to each other. For instance, if the embodied agent (e.g., *NAO*) pretends to sleep at a certain point, the speech agent (e.g. *Alexa*) may join the pretence and wake it up with a rebuke or a self-deprecating remark. Each performer will have its own domain model suited to its own modality, and its own meta-model. But each will need to share a joint meta-model to permit coordination.

It's worth noting that in addition to the *NAO*'s physical affordances for pantomime, it also offers some support for vocal mimicry. So while its built-in voice is twee, the robot permits one to upload arbitrary sound files and recordings, making the use of 3rd-party voice synthesis tools (such as those offered by IBM Watson) a viable option. We draw on this service when we want *NAO* to communicate directly with *Alexa* and to have its voice prompts understood as commands, since *Alexa* does not react to the *NAO*'s normal speaking voice. It can also be used to associate a different speaking voice with different meta-model functions, from making wisecracks about the current story to making fun of the audience to poking fun at the system's developers. A key use of this ability is the coordination of meta-models. The *Alexa* narrator articulates each beat of the story before waiting for the *NAO* to respond in an embodied fashion. Since neither knows how long the other will take, they use conversation (of a sort) to align their own private models.

Skolem Golems and Scéalextric

The *Scéalextric* system of Veale (2017) offers an open and extensible approach to story-generation that has sufficient knowledge to build both the domain- and meta-models. A plot in *Scéalextric* is built from plot triples, each of which, in turn, comprises of a sequence of three plot verbs. In all, *Scéalextric* provides semantic support for 800+ plot verbs, by indicating e.g. how each verb causally links to others, or how each verb can be idiomatically rendered in a final text. Each verb is assumed to link the same two protagonists, in a story of just two characters overall. It balances this limitation by exploiting a vivid cast of familiar fillers for these two roles, drawing on the *NOC list* of Veale (2016) to provide detailed descriptions of over 1000 famous characters. Veale (2017) reports empirical findings as to the benefit of reusing familiar faces in shaggy-dog tales, noting that readers rate such tales as more humorous and more eventful. Yet the shagginess of these tales is exacerbated by the way that triples are connected, end-over-end, to generate what amounts to a random walk in the causal graph of plot verbs. Though *Scéalextric*'s plot graph has over 3000 edges connecting its 800+ verbs with arcs labeled *so, then, and, but*, the resulting stories exhibit local coherence at the expense of global shape. Its tales meander, and lack a clear purpose.

The limitations of *Scéalextric* as a domain model need to be remedied if a rich meta-model is to be built on top of it. A story of just two characters does not afford much variety for even a single performer to leverage, much less a double act, whilst the lack of coherent sub-plots that return to the main story trunk also reduces the potential for play at the meta-level. We remedy both deficiencies with a new kind of triple that is designed to be expanded recursively, into a plot tree, rather than additively into a rambling plot line. So rather than connecting plot triples end-to-end, our approach will expand these new triples via recursive descent from a single starting triple that gives each story its global shape.

Consider how *Scéalextric* (Veale, 2017) defines and uses its triples. Suppose TUV, VWX and XYZ are triples made from the plot verbs *T, U, V, W, X, Y* and *Z*. Then each verb is assumed to take two implicit character slots, α and β, which are later filled with two specific characterizations drawn from the NOC list. So the triple XYZ is in fact the sequence <α *X* β> <α *Y* β> <α *Z* β>. Triples are connected end-over-end, with the last verb of one matching the first verb of the next. In this way, TUV, VWX and XYZ can be combined to construct the story TUVWXYZ. The causal graph provides a labeled edge between any two plot verbs that are linked by at least one triple; the label set is {*so, then. and, but*}. So if given the starting verb T and the ending verb Z in advance, a system can search the graph of causal connections to find a story of a stated minimum size that starts with the action <α *T* β> and ends with <α *Z* β>.

In our augmentations to *Scéalextric*, we add a range of triples of the form T-X-Z, where *T, X* and *Z* are plot verbs and the hyphen – denotes a point of recursive expansion. Thus, T-X and X-Z admit additional content to link T to X and X to Z. This content is inserted as further triples, such as XYZ (to link X and Z) or T-V-X. The latter links T to X via another recursive triple that requires expansion in the gap from T to V and from V to X. The nonrecursive triples TUV and VWX can fill these gaps to yield a complete plot, TUVWXYZ. Notice how the existing stock of *Scéalextric* triples is reused, not replaced, and simply augmented with new triples that operate top-down rather than left to right. A subset of the new recursive triples are marked as suitable for starting a story; these give each plot its global shape. At present we designate over 200 recursive triples to be story starters, but these can be adjoined in a left-to-right fashion (as in the original *Scéalextric*) to create higher-level story shapes. Thus, the triples A-J-T and T-X-Z may be adjoined to create a story that starts with action A and ends with Z

For stories with just two characters a generator need not worry about under-using a character, especially if each plot verb – as in *Scéalextric* – assumes the participation of both. The introduction of arbitrarily many additional characters can enrich a narrative greatly, but at the cost of complexity. All characters must be kept in play, and not forgotten even when they are not participating in the current action or sub-plot. A sub-plot is a story path that diverges from the main trunk of the narrative and rejoins it at a later time. Consider a story in which character α *assaults* character β. A viable sub-plot involves α being investigated for the assault by a third character γ that fills the role of detective. The sub-plot may recursively draw in a fourth character, a lawyer for α, which then necessitates the introduction of a lawyer for β. When the sub-plot ends and the plot rejoins the main trunk, these additional characters can be forgotten, but not before.

We add a capacity for additional temporary characters to *Scéalextric* via skolemization. If β is a character, β-*spouse* denotes the love interest of β in <α *seduce* β-spouse>, so whatever NOC character is chosen for β, a relevant NOC character is also chosen to fill β-*spouse* (e.g. Bill Clinton for Hillary Clinton). Other skolem functions include *friend*, *enemy*, *partner*, and each exploit the NOC in its own way. α-*friend*, for instance, is a character with a high similarity to the filler for α (e.g. Lex Luthor for Donald Trump), while α-partner is instantiated with a character of the same group affiliation in the NOC (e.g., Thor for Tony Stark, as both are Avengers). Other skolems, such as α-*lawyer* or β-*detective*, exploit the taxonomic category field of the NOC list. In such cases, the most similar member of the category is chosen to resolve the skolem, so α-*lawyer* is filled with a character similar to α that is also a lawyer, and β-*detective* is filled by a detective that resembles β (e.g., Miss Marple for Stephen Hawking). No skolem is ever instantiated as a character that is already in use in the current story context.

These additions to *Scéalextric* give it much of the flexibility of traditional story grammars while preserving the key knowledge structures that make its stories so playful and diverse. Its stories still exploit unexpected juxtapositons of NOC characters that evoke both similarity and incongruity, but now a story can draw even more characters into its web while choreographing how they interact with each other. As we consider this an important contribution of the paper we shall make these additional triples and skolemizations available for use by other story-generation researchers. But now let us consider how these additions can be exploited at the meta-level to drive a creative story-telling double act.

Are These The Droids You're Looking For?

In comedy, timing is key, and so choreography is needed to align the actions of partners to ensure that they read from the same script while staying in sync from one beat to the next. For a given beat it is impractical for one to infer the timing of another, as a *NAO* cannot reliably infer how long it will take *Alexa* to speak the text of a beat, just as *Alexa* cannot know how long the *NAO* may take to enact it. If our duo is not to become hopelessly co-dependent, an unseen partner is required to manage backstage coordination. This 'third man' is a *blackboard* (Hayes-Roth, 1985), the ideal architecture for synchronizing the cooperative strangers of a distributed system. As shown in Montfort *et al*. (2013), a blackboard is a communal scratch pad on which different generators can track their work and share both knowledge and intermediate work-products. We shall use a blackboard to store key elements of the domain- and meta-models of the performers, as well as their current positions in each.

The double-act is initiated by a command to *Alexa*, such as 'Alexa, tell me a story about Donald Trump.' So it is the responsibility of Alexa to retrieve an apt tale from her story space, as already pre-generated using the augmentations to *Scéalextric* described above. Each story is fully rendered as text when retrieved, and *Alexa* segments it into a sequence of individual story beats of one action apiece. It is this sequence that is placed on the blackboard for *NAO* to see. In the dance of *Alexa* & *NAO*, *Alexa* leads and *NAO* follows. *Alexa* starts the tale by articulating the text of the first beat, then waits for *NAO* to respond. The robot, seeing the cued beat on the blackboard, reacts appropriately, either with a pantomime action for the plot verb, or with a gesture that signifies its response to the story so far. But *Alexa* does not proceed with the story until she is given an explicit vocal command to do so, e.g., 'continue', 'go on', 'then what' or 'tell me more.' This can come from the audience, but *NAO* will provide it itself if none is forthcoming. When it replies to *Alexa*, the robot looks down at the *Echo* device to maintain the social contact of a double-act. Both agents are engaged in a back-and-forth conversation, and it should show.

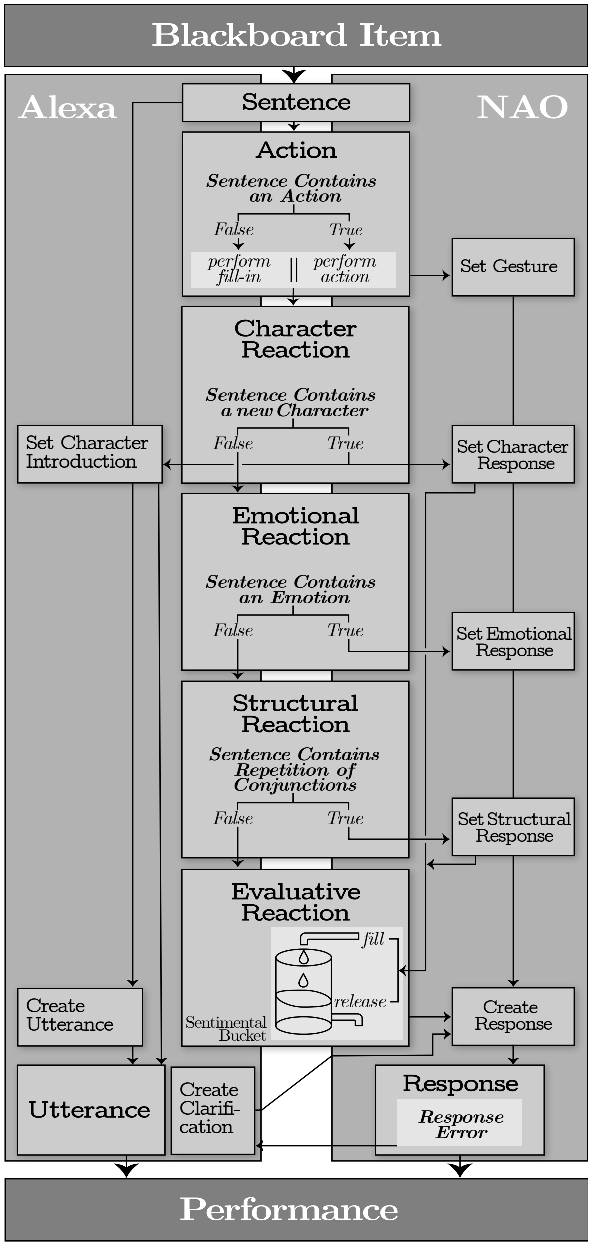

Figure 2. Blackboard logic for the system's meta-models.

This baseline conversation uses only the domain models. But as more substance is added to the meta-models of each partner, sophisticated artifice is possible. So *NAO* can peek at the next story beat on the blackboard, and determine its causal relation to the last. It can then use this to choose its cue to *Alexa* to proceed with the tale. Suppose the next beat is 'But Donald spurned Hillary's advances'. Seeing the *but*, *NAO* can prompt *Alexa* to go on by ominously asking '*But then what?*' In this way a single initiative task becomes a mixed initiative task, in which *NAO* draws the tale out of its companion, and seems to shape it as it is spun. As *NAO* uses pre-recorded sound cues for these interactions (recall that *Alexa* does not understand *NAO*'s native voice), it can use sound effects here as well as richly tempered voice recordings, to give the interactions a greater social dimension.

An integrated depiction of the double-act's meta-models is shown in Figure 2. A key responsibility of a meta-model is to predict an audience's response to an unfolding story and allow performers to take elaborative action as needed. Suppose *Alexa* articulates three successive story beats that begin with *then*, *so*, or *and*. A meta-model may see this as characteristic of a flat stretch in a story in which one action leads predictably, and boringly, to the next, and so spur the robot to reply with a structural reaction, such as a yawn.

If *NAO* peeks ahead to see that the current flat stretch is about to lead to a 'but' it can announce, wisely, 'I see a but coming.' Alternatively, the robot might reply with laughter when a silly act is described, or, more insightfully, when a character gets his comeuppance. An unexpectedly negative turn in a story may prompt the robot to utter "Dick move!" or some other pejorative that shapes the audience's view of the evolving tale. The robot can also pass remarks on characters as they are introduced into the story, by querying the NOC list for relevant qualities. So it may, for instance, say that "Donald Trump is so arrogant" when that character is introduced for the first time. Each meta-model may also be capable of its own small acts of creativity. For instance, the meta-model can generate dynamic epithets for characters as they evolve in a tale, such as *Hillary the Death-bringer*, *Bill the Seducer*, or *Donald the Lie-Teller*. These epithets can be the robot's spoken contribution to the plot delivery. So the meta-model allows performers to switch from narrator to actor to Greek chorus as the story context demands.

The joint meta-model of Fig. 2 supports the following reactions to a tale as it is told: *gestural reactions* (the *NAO* makes an appropriate gesture for a given action); *character* reactions (*NAO* or *Alexa* react in an apt fashion whenever a character is introduced); *structural* reactions (*NAO* reacts to the logical shape of the tale); *emotional* reactions (*NAO* reacts with emotion to a plot turn that is highly positive or negative); and *evaluative* reactions (*NAO* or *Alexa* react to their cumulative impression of a story so far, if this opinion is sufficiently positive or negative to be worthy of remark). Since our content model is *Scéalextric,* a wholly symbolic CC system, all stories have predictable markers that allow our meta-models to be implemented as rule-based systems. The next section illustrates the reaction of the meta-models within an annotated transcript of our double act in action.

The Double Act in Action

As the only embodied agent in the duo, it is the responsibility of the robot to create the duo's shared physical space. *NAO* must address itself to *Alexa* to present their interactions as a conversation, and not just a pairing of devices that speak past each other in a synchronized manner. To begin with, *NAO* asks *Alexa* to 'open your notebook' so they can create a story together. *Alexa* then asks *NAO* for a subject, which it provides (such as "Star Wars") and *Alexa* reacts by noting her satisfaction with the story to come. This tale then unfolds, beat by beat, with *NAO* asking *Alexa* to continue between beats once it has enacted its own reply. But *Alexa* has responsibilities too, and must do in words what it cannot do with physical acts. *Alexa* must acknowledge the robot's contributions to show that they do indeed share the same space. For when one agent acts as an audience to the other, they can collectively shape *our* feelings for the tale.

Alexa's weaknesses have been well-documented in other work (e.g., Kapadia *et al*., 2017), and her recovery mode is not sufficiently transparent to avoid failed interactions with the *NAO*. So if *Alexa* does not receive her next prompt in a timely manner, she will reiterate several requests for input before eventually quitting the narrator skill. Unfortunately, this reiteration cannot be unpacked so as to tell the blackboard of a failed interaction, so *NAO* will remain unaware that *Alexa* has, so to speak, left the building. However, the robot can detect a long pause in a conversation and prompt *Alexa*, with comic frustration, to continue. With each of its reprompts, the robot's feigned exasperation grows, until it finally tells *Alexa* to "reopen your notebook" and continue. At this point, the narration continues from the last beat that was processed, and the duo returns to their bickering ways.

What follows is a sample transcript of the double-act in action. All speech is presented in italics, while each gesture and meta-model reaction is annotated in square brackets:

NAO: *Alexa, open your notebook.*
Alexa: *Please provide a topic for our story.*
NAO: *How about Harry Potter?*
Alexa: *I have written a great story about Harry Potter.*
NAO: *Let's hear it.*
Alexa: *What if Harry fell in love with Hermione Granger?*
NAO: *Sounds promising.*
Alexa: At first, *Harry's thoughts of Hermione were of love.*
NAO: *Proceed.* [Love Gesture]
Alexa: *I didn't get that.* [Failed Interaction]
NAO: *Please go on.* [Reprompt] [Waiting Gesture]
Alexa: *Because Hermione openly favored Harry.*
NAO: *Tell me more.*
Alexa: *Hey, Bubblehead! we need another character.*
[Introduction of new character]
NAO: *Lizzy Bennet works for me.* [New character Reaction]
Alexa: *Yet Hermione governed the land with Lizzy Bennet.*
NAO: *Oh damn.* [Structural Reaction]
Alexa: *Should I continue?* [Reprompt]
NAO: *Continue. Is there more?* [Waiting Gesture]
Alexa: *But Harry spurned Lizzy Bennet's advances.*
NAO: *Oh my goodness!* [Structural Reaction]
Alexa: *So Lizzy Bennet took a whip to Harry's back.*
NAO: *Please go on.*
Alexa: *Well, Harry complained to Hermione.*
NAO: *But then what?* [Structural prediction]
Alexa: *But Lizzy Bennet spread slander about Harry.*
NAO: *That is interesting.* [Spread Gesture]
Alexa: *Yet Hermione's suspicions about Lizzy grew.*
NAO: *That escalated quickly!* [Structural Reaction]
Alexa: *Are you still there?* [Reprompt after brief failure]
NAO: *Tell me more.* [Waiting Gesture]
Alexa: *Then Hermione banished Lizzy from the kingdom.*
NAO: *I am entertained.* [Evaluative Reaction]
Alexa: *But Lizzy was not as dumb as Hermione imagined.*

NAO: *Oh my goodness!* [Structural Reaction]
Alexa: *So Lizzy unseated Hermione from her perch*.
NAO: *Then what?* [Flat Structural Reaction]
Alexa: *Then Lizzy Bennet had Hermione locked up.*
NAO: *So then what?* [Locked-Up Gesture]
Alexa: *So Harry's denunciation of Lizzy was scathing.*
NAO: *Please go on.* [Humiliated Gesture]
Alexa: *Well, Lizzy turned Harry into an indentured slave*.
NAO: *What a terrible person.* [Emotional Reaction]
Alexa: *But Harry escaped from under Lizzy Bennet's nose*.
NAO: *I am surprised!* [Structural Reaction]
Alexa: *Now expect to be poisoned with bad cooking!*
[New Character Introduction: Bridget Jones]
NAO: *Bridget Jones, a great add!* [Character Reaction]
Alexa: *Meanwhile Bridget guarded Hermione closely*.
NAO: *Cool, tell me more*. [Guarding Gesture]
Alexa: *So Harry's wiles proved too strong for Bridget*.
NAO: *And then what next?* [Flat Structural Reaction]
Alexa: *Then Harry put poison in Bridget Jones's cup*.
NAO: *That took a dark turn*. [Emotional Reaction]
Alexa: *After this, Harry gave Hermione her freedom*.
NAO: *I love to hear that*. [Emotional Reaction]
Alexa: *So Hermione fomented rebellion against Lizzy*.
NAO: *And then what?* [Rebellious Gesture]
Alexa: [Interaction failure. Alexa exits story skill.]
NAO: *Continue*. [Reprompts after delay]
NAO: *Are you there?* [Reprompts Alexa again]
NAO: *Hello, Anyone home?* [Emotional Reaction to Alexa]
NAO: *Alexa, reopen your notebook*. [Reopens skill]
Alexa: *Then Hermione usurped all of Lizzy's power*.
NAO: *So what?* [Grasping Gesture]
Alexa: *So Hermione turned Lizzy into an indentured slave*.
NAO: *That took another dark turn*. [Locked-Up Gesture]
Alexa: *In the end, Harry walked Hermione down the aisle.*
NAO: *That's just lovely*. [Emotional Reaction]
Alexa: *The end*.

A single transcript can be revealing about specifics, but not about the generic tendencies of a system. To appreciate the latter, we simulated the double-act 136,787 times, choosing a different story to tell each time. Our goal was to estimate the relative occurrence of alternate meta-model reactions to the story in each case. In particular, we considered the following: the BUT structural reaction to a turn in the plot; the BORED evaluative reaction to a predictable stretch of plot; the STRONG emotional reaction to a highly-charged plot verb; the GOOD evaluative reaction to an exciting stretch; the NEW character reaction to the introduction of another named entity to a story; and the GESTURE reaction, which delivers a mimetic response to a given plot action. Overall, the BORED evaluative reaction accounts for 18.4% of all reactions, the BUT structural reaction accounts for 16.6%, the STRONG emotional reaction accounts for 15.5%, the NEW character reaction accounts for 7.7%, and the GOOD evaluative reaction accounts for 4%. In all remaining cases, or 37.8% of the time, the NAO responds structurally, with a prompt to "continue" or "go on" and a downward glance at the *Echo* unit by its side. The GESTURE reaction is independent of these other reaction types, since the robot can make a gesture *and* utter a spoken response in a single turn. For 49.6% of story beats the robot performs a gesture that is visually mimetic of the current plot verb; for the other 50.4% of beats, *NAO* makes a 'holding' gesture – such as folding its arms, putting its hands on its hips, or shifting its weight from one leg to another – in the manner of human listeners who wish to emphasize their physical presence.

Related Work

The *Alexa* skill store contains an array of storytelling skills for the Amazon *Echo*, ranging from linear narratives to the *choose-your-own-adventure* style of story. None, however, uses computer-generated tales as a basis for narration, and few tell stories as complex or data-rich as those used here.

Kapadia *et al*. (2017) paired *Alexa* to *YuMi,* a two-armed industrial robot, to develop a learning-from-demonstration (or LfD) system. LfD requires trainers to use both hands to move a robot's own limbs into the poses it must learn, and to annotate these actions at the same time. The pairing with *Alexa* allows trainers to speak to the LfD system to verbally label what is being taught as they use their own hands to move the robot into its demonstration poses. The authors note the vexing technical challenges that *Alexa* entails, but still argue that using *Alexa* for hands-free vocal control in a robotic context is worthwhile. Their LfD system, *EchoBot*, is not a true double-act, however, as a human manipulates both devices simultaneously with voice and gesture inputs, and *EchoBot* is not designed to exhibit its own personality.

Fischer *et al*. (2016) also use *Alexa* as voice control for a robot, the one-armed *Kinova Jaco*. Users issue commands to *Alexa* (via *Echo*) and a backend turns these commands into appropriate kinematics for the robot. While *Alexa* and the robot are cooperating partners, interaction is one-way and not a dialogue. Neither is it part of a creative task.

Kopp, Bergmann & Wachsmuth (2008), building on the work of Kita and Özyûrek (2003), presented a multi-modal system that also uses a blackboard to integrate spoken text and embodied gestures into a single communicative act. In this case, multimodality occurs within the simulated environment of a virtual visual agent, or avatar, whose animated gestures achieve both communicative and cognitive ends: they augment what is said, and reveal the inner state of the cognitive agent as they do so. Each modality operates with a shared representation on the blackboard (both imagistic and propositional in nature) of that which is to be said, and enacts it as speech or gestures to suit their own agendas. In effect, this system is a double act of sorts, realized as just a single coherent agent. Yet such coherence prohibits a dual system from reaping the benefits of a true double act, since only the latter allows a system to talk to, interrogate, and make fun of itself in a consistent and humorous manner.

Farnia & Karima (2019) explore how humorous intent is marked in a text, and the effect of these markers, subtle or otherwise, on the perception of humour by an audience. A double act of *Alexa* and *NAO* allows us to explore markers that are more than just textual, or even vocal, to explore how a witty personality can be constructed from the physical and meta-linguistic markers that are imposed on a text.

Double Vision: Summary and Conclusions

A good double-act is a marriage of convenience, even if it often looks otherwise. Many comedic duos go out of their way to accentuate their differences, as comic friction only serves to emphasize their complementarity. When partners complete each other, it is as if they occupy a world all their own. Nonetheless, even a seamless partnership may require significant backstage coordination to make it all work. The same is true of technology double acts, such as our pairing of *Alexa* & *NAO* that turns story-telling into a performance. In this paper we have focused on the considerable – but not always obvious – technical challenges of making a double act of *Alexa* and *NAO* a practical reality in a CC context. We have developed the content models, the meta-models, and the platform functionalities to the point where we can finally use the double act to empirically test our hypotheses regarding the true value of embodiment and multimodality in the generation and delivery of machine-crafted artifacts.

Our double act divorces the job of story generation from the task of telling a story well. Each responsibility requires one CC system to be sympatico with the other, just as the performers in a double act must read each other's minds, or – more realistically – their shared blackboard architecture. Nonetheless, we have structured the performative functions so that they can work with machine-generated tales of any kind, once the meta-models have been adapted to operate over this new content model. Even so, we have only begun to exploit the full performance possibilities of offline generation and later online delivery in a multimodal setting. In addition to the obvious entertainment applications, we are mindful of the educational possibilities of CC double acts that *show* as well as *tell*, that embody what they create, and that reveal an emergent personality they can call their own.

To both see and hear the *Walkie Talkie* double-act do its thing, readers are invited to subscribe to the following channel on *Youtube*, where annotated videos of the duo performing a series of different stories can be watched online:

https://bit.ly/2SNeeHQ

References

Amazon. (2019). Alexa skills kit. *https://developer. amazon.com/ alexa- skills- kit* (last accessed, February 2019).

Bergson, H. (1911/2013). *Laughter: An Essay on the Meaning of the Comic*. Trans. C. Brereton & F. Rothwell. New York: Dover.

Epstein, L. (2005). *Mixed Nuts: America's love affair with comedy teams from Burns and Allen to Belushi and Aykroyd*. Waterville, ME: Thorndike Press.

Farnia, M. & Karimi, K. (2019). Humor markers in computer-mediated communication. Emotion perception and response. *J. of Teaching English with Technology,* 1:21:35.

Fischer, M, Memon, S. & Khatib, O. (2016). From Bot to Bot: Using a Chat Bot to Synthesize Robot Motion. *The AAAI Fall Symposia series, AI for Human Robot Interaction*, TR FS-16-01.

Freud, S. (1919). Das Unheimliche. In *Collected Papers*, volume XII. G.W. 229–268.

Fuller, D. & Magerko, B. (2010). Shared mental models in improvisational performance. In Proc. of the Intelligent Narrative Technologies III Workshop (INT3 '10). ACM, New York, NY, USA

Hayes-Roth. B. (1985). A Blackboard Architecture for Control. *Artificial Intelligence*. **26** (3): 251–321.

Kapadia, R., Staszak, S., Jian, L. & Goldberg, K. (2017). EchoBot: Facilitating Data Collection for Robot Learning with the Amazon Echo. In Proc. of the 13th IEEE Conference on Automation Science and Engineering (CASE) Xi'an, China, August 20-23.

Kita, S., & Özyûrek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. Journal of Memory and language 48(1):16–32.

Kopp, S., Bergmann, K., & Wachsmuth, I. (2008). Multi-modal communication from multimodal thinking towards an integrated model of speech and gesture production. International Journal of Semantic Computing 2(1):115–136.

Pelachaud, C., Gelin, R., Martin, J., & Le, Q.A. (2010). Expressive gestures displayed by a humanoid robot during a storytelling application. *New Frontiers in Human-Robot Interaction* (AISB), Leicester, UK.

Pérez y Pérez, R. (2007). Employing emotions to drive plot generation in a computer-based storyteller. *Cognitive Systems Research* 8(2):89-109.

Montfort, N. (2009). Curveship: An interactive fiction system for interactive narrating. *In Proc. of the Workshop on Computational Approaches to Linguistic Creativity*, at the 47th Ann. Conf. of the Assoc. for Computational Linguistics, Boulder, Colorado, 55–62.

Montfort, N., Pérez y Pérez, R., Harrell, F. & Campana, A. (2013). Slant: A blackboard system to generate plot, figuration, and narrative discourse aspects of stories. In *Proc. of the 4th International Conf. on Computational Creativity*. Sidney, Australia, June 12-14.

Shaftesbury, Lord, (1709/2001). Sensus Communis: An Essay on the Freedom of Wit and Humour. In: *Characteristicks of Men, Manners, Opinions, Times*. Indiana, Indianapolis: Liberty Fund.

Veale, T. (2016). Round Up The Usual Suspects: Knowledge-Based Metaphor Generation. In Proc. of *the Meta4NLP Workshop on Metaphor at NAACL-2016, the annual meeting of the North American Assoc. for Computational Linguistics. San Diego, CA*.

Veale, T. (2017). Déjà Vu All Over Again: On the Creative Value of Familiar Elements in the Telling of Original Tales. In Proc. of *ICCC 2017, the 8th Int. Conf. on Comp. Creativity, Atlanta*.

Veale, T. & Cook, M. (2018). *Twitterbots: Making Machines that Make Meaning*. Cambridge, MA: MIT Press.

Wicke, P. & Veale, T. (2018a). Storytelling by a Show of Hands: A framework for interactive embodied storytelling in robotic agents. In *Proc. of AISB'18, the Conf. on Artificial Intelligence and Simulated Behaviour,* pp 49--56.

Wicke, P. & Veale, T. (2018b). Interview with the Robot: Question-guided collaboration in a storytelling system. In *Proc. of ICCC'18, the 9th Int. Conference on Computational Creativity, Salamanca, Spain, June 25-29.*