Personal details

Dr Stephen Rainey, PhD, Research Fellow, The Oxford Uehiro Centre for Practical Ethics, Faculty of Philosophy, University of Oxford

Delivery address, email address, and phone number

Oxford Uehiro Centre for Practical Ethics, Suite 8, Littlegate House, St Ebbes Street, Oxford OX1 1PT

Stephen.rainey@philosophy.ox.ac.uk

+44 (0) 7856736879

Acknowledgements

The author gratefully acknowledges funding from the BrainCom Project, Horizon 2020 Framework Programme (732032), in the production of this article.

Thanks to Dr Hannah Maslen and Dr Jan Christoph Bublitz whose extremely helpful comments served to help improve upon earlier drafts.

‘A Steadying Hand’: Ascribing Speech Acts to Users of Predictive Speech Assistive Technologies

Abstract

Neuroprosthetic speech technologies are in development for patients suffering profound paralysis, such as can result from amyotrophic lateral sclerosis (ALS). These patients would be unable to speak without intervention, but with neurotechnology can be offered the chance to communicate. The nature of the technology introduces a neuroprosthesis that mediates neural activity to generate synthesised speech. How word prediction coheres with speaker intentions requires scrutiny. Some future forms of prostheses, using statistical language models to predict word patterns, could be thought of as participating with communicative intent – not merely channelling it. Concepts relating to vicarious liability, may serve to clarify these issues. This paper is about showing how technology might interact with speaker intent in cases of delegated action, and how it should be seen as participating in the implementation of user ‘instructions’.

Keywords

Neurotechnology, language, synthesised speech, agency, vicarious liability, delegation

Introduction

This paper explores how speaker intent and assistive predictive speech technologies may combine in speech neuroprosthetic technology to generate puzzles about whether and how to ascribe acts to an agent who has delegated the carrying out of that act (at least in part) to another. In the specific case of a predictive speech prosthesis, the delegation is to hardware and software and the action they assist with is speech action.

Tamburrini,[1] discusses how existing legal concepts can be used in dealing with the apparently novel circumstances that may be thrown up by cases such as those in which neuroprosthetics are involved. His analysis suggests that cases of harms deriving from neuro-controlled devices can be dealt with by existing legal precepts such as employer responsibility for employee action; actions of well/badly trained pets as the owner's responsibility; Parent/child responsibility; producer responsibility in product harm, given their expectation of profit from sale. These responsibility and control issues are clearly vital, and illuminating. But, in order to identify the implications of ‘speech delegation’ in a wider range of contexts, we argue that conceiving of the utterance as delegated in accordance with the logic of vicarious liability is illuminating at the level of description of chains of events, but would not result in appropriate ascription of speech in contexts such as medical decision-making and will-making. Thus, contra Tamburrini, we ought to conclude that

[1] ‘Brain to Computer Communication: Ethical Perspectives on Interaction Models’ (2009) 2(3) *Neuroethics* 137, 7 ff.

existing approaches in the law may not straightforwardly resolve disagreements regarding attribution of speech acts in all cases involving predictive speech assistive technologies.

Correspondingly, the intent here is not to use these concepts in an analysis or application of liability law itself, but to use some general concepts from it in order to analyse the ascription of utterances to speech prosthesis users. This may seem somewhat incongruous, so some explanation will clarify things.

Vicarious liability holds employers accountable in various circumstances for actions of their employees that lead to harm. This is at least because:

- by hiring employees, the employer creates the risk of harm to third parties by its employees' negligence. Where it benefits from the use of employees, the employer should also accept all the risk that comes with those employees;
- it provides an incentive for employers to exercise care in the selection, training and supervision of all employees…
- most employers have "deeper pockets" than their employees, which means that if the employee does not have sufficient resources to pay for the injury, the employer's superior economic position will help ensure the injured party will be properly compensated...[2]

In context of employment strict (vicarious) liability applies because we care most about redress in cases of employee wrongdoing, or harmful outcomes. In such cases we are concerned with who is liable for the acts carried out that lead to the wrongs or harms. In other contexts most relevant to the present case, like drawing up wills, and medical decision-making, we care much more about whether the content of actions really can be attributed to the proper agent, e.g. the patient in medical decision-making. In cases like medical decision-making, we aren't primarily pursuing an analysis aimed at holding someone liable for outcomes. These contexts are not about redress, but rather respecting the wishes of an agent.

While this makes the assistive speech technology case seem pretty different, the parallel comes in the ways that goals can be achieved through delegation, that is, employing a 'helping hand'. Might there be cases where an utterance via speech technology should be taken as the agent's utterance, despite the helping hand making a substantial difference to the speaker's intent? What about the possibility of the technology overshadowing completely that intent? These questions turn upon how we are to understand the 'instructions' of the user having been processed by the machine.

More precisely, we need to know in what ways the helping hand provided by assistive technology can affect the connection between the agent's aiming for some end and the

[2] Adapted list from Bowal Peter and Bontorin Louis, *Vicarious Liability: The Legal Responsibility of Employers - LawNow Magazine* (9 August 2018) lawnow.org <https://www.lawnow.org/vicarious-liability-legal-responsibility-of-employers/>.

attaining of that end. A useful distinction here, one utilised elsewhere in discussing different types of neurocontrolled prostheses,[3] might be between 'goal-setting' and 'process control'.

Briefly, where the user of a prosthesis has goal-setting power, their prosthesis acts on their instructions somehow to bring about that goal. The 'somehow' means that the prosthesis uses some automation, or responsiveness to environment, of its own in order to realise the goal: the user thinks 'get a drink' and the prosthesis operates so as to get a glass of water. Where process control is centre stage, the prosthesis user operates their device in a fine-grained way to themselves realise their goal: thinking move left, move right, move up, in order to control a prosthetic arm's grasping a glass for instance.

In cases of goal-setting, process control can be seen as delegated to the prosthesis itself. But in being somewhat automated, we might wonder about the transmission of agency here. Could the prosthesis take *so much* process control that it effectively adopts some goal-setting power? Certainly this would be a worry if bad consequences were to come about via the action of the prosthetic. And in this kind of logic of delegation we can see the parallel with concepts from vicarious liability.

In being employed by a user, a prosthesis might overstep its bounds if it adopts goal-setting power. As long as automated process control realises the goals of the user somehow, it can be seen as functioning well. But there are limits here. If an employee were given a task to enact as a part of their employment, there are various ways in which they may complete that task. Along a spectrum lies *detour* and *frolic* – to be discussed a little more later – whereby the employee may be seen to complete the task by some odd means, or to deviate completely from acceptably realising the employer's goal.

It's not that vicarious liability applies such that we should see an assistive speech prosthesis as an employee, and the user as liable for whatever it says on their behalf. But perhaps some similar relationship exists in the use of a such technology, and so might generate an analogous problem to solve: how or whether to ascribe acts to an agent who has delegated the carrying out of that act (at least in part) to another. Does the action of the prosthesis realise the goal set by the user in a way acceptable to that user, or does its functioning impinge so much upon the user's intent that it effectively sets its own goals? We can look to some concrete cases, in more conventional terms, for examples of themes that further motivate our problem.

Cases

Barrett v Bem[4] discusses a literal helping hand. A testator made a will on the day of his death, his hand steadied from shaking by a relative party to that will. The case turned on whether the will was signed voluntarily if the signing hand was steadied by someone party to the inheritance, or if this amounted to someone else signing the testator's name. Given

[3] Steffen Steinert et al, 'Doing Things with Thoughts: Brain-Computer Interfaces and Disembodied Agency' [2018] *Philosophy & Technology* <http://link.springer.com/10.1007/s13347-018-0308-4>.

[4] *Barrett v Bem & Ors [2012] EWCA Civ 52 (31 January 2012)* <http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2012/52.html&query=(barrett+OR+v+OR+bem)>.

that in general, it is "undesirable that beneficiaries should be permitted to execute a will in their favour in any capacity" the question of 'helping' sign a will is a pertinent one. To what extent could a helping hand be seen as a determining factor in the production of a will, or any outcome?

In terms of the neuroprosthetic speech case, the technology could be seen as the steadying hand. Speaker intent is present, but the paralysed body is unable to utilise the conventional ways of realising it. The technology supplants this deficit, in bypassing the conventional articulatory mode, and creates a new means of realising speech intent. As in the signing of a will, the pertinent question is: to what extent does 'steadying' determine the content of the subsequent action?

Certainly, rephrasing some intended speech could be considered as potentially problematic. But not always. A user might like the 'autocorrect' sense in which a statistically-informed speech model could refine their language. They might enjoy an enhanced vocabulary, or power of expression. But, perhaps especially in contexts of advanced directives, the dividing of estates, and basic freedom of action, this could amount to a deviation from agency. How much autonomy could we, or should we, ascribe to a speaker if we suspect their speech prosthesis is acting as a quasi-agent between intent to speak and final speech production?

At an abstract level, we can see how this kind of intention transmission scenario could apply in general to cases of technology-mediated speech. A user of a speech device develops an intention to speak, triggers their device according to specific training, and verbal output results. This is in turn interpreted by an audience. What if a speech prosthesis says things the speaker didn't intend? Could we apply legal doctrines in order to contribute to answering this question? Certainly, we can see how more parallels may be drawn with a few other existing cases.

In X Health Trust v XB[5] the validity of an advanced directive to withdraw treatment from a man suffering with motor neuron disease was upheld, despite concerns being raised about an apparent 'valid until' clause. The advanced directive (AD) was compiled using a template from the internet that included a box labelled 'valid until' near which was placed a date. However, the AD was held to be a legitimate expression of the patient's perspective following examination of the way in which it was drawn up: in the presence of the patient's wife, his GP who was well-known to him, a mental capacity professional, and a carer. Given this, the apparent validity expiration was held to be not an indicator prevarication or uncertainty.

In the context of a speech neuroprosthesis, the content of this case is particularly relevant as it involves high-stakes expression of a perspective, without a natural means of communication. This patient's advanced directive had been produced not without difficulty. However, the judge was satisfied, following oral evidence from those that were present, that due care had been taken to represent faithfully the patient's views.

[5] *The X Primary Care Trust v XB & Anor [2012] EWHC 1390 (Fam)* (Unreported, EWHC (Fam), MRS. JUSTICE THEIS, 1 May 2012).

Where a neuroprosthesis might deploy statistical models to predict speech, the position of the speaker becomes extremely important as deviations from that perspective could be possible in the name of 'better' speech outputs. The genesis of the output, as the AD, might be problematic or contestable, but the output itself might very well be assessed as accurately portraying the person given further evidence of a proper sort.

In Sharp v Adam,[6] the case turned upon the rationality of a will alteration that saw an 'inexplicable' disinheritance of the testator's daughters. In being afflicted with a progressive disease, the willmaker's cognitive capacities were put into question. In general, they were held to have been adequate for the making of a will, yet to have been deficient in some specific respects owing to the progression of the illness. The final will was held to be irrational in part and an earlier version put forward as authoritative. Interestingly for the current context, the fact that the rationality of the decision was scrutinised allows us to draw a parallel with the pragmatic dimensions of speech.

Whereas a speech device restores a communicative practice to someone otherwise suffering without one, it is not outside the realm of possibilities that in some instances this capacity for communication may be ill performed. Slips of the tongue, in everyday speech for example, illustrate this kind of performance error. We don't doubt capacity in such cases, merely noting a mistake in speaker performance. Such errors are easily picked up and remedied. Mediated in technology, however, there is greater need perhaps for an audience to listen critically to produced speech. Where a speaker is known to their audience, and yet they appear to speak in an out of character way, there is likely good reason to scrutinise the speech in terms of its rationality, just as in Sharp v Adam.

Each of these cases represents an instance where an agent’s instructions are in some way questioned owing to their particular mediation. We will make the relation to speech prostheses more concrete next, in order to introduce a means of understanding what’s at stake. In order to consider ascribability for utterances in cases of prosthesis-mediated language, it will be useful to examine how this technology may or may not provide parallels with legal concepts and precedents. That being so, it is important to get a sense of the technology involved, and how it can be designed to produce speech for users.

The technology

Neuroprostheses for the production of speech in cases of profound paralysis, such as can result from amyotrophic lateral sclerosis (ALS), are increasingly being explored. These technologies record, process, and decode neural correlates of speech production in order to synthesise spoken language.[7] For example, they may involve the introduction of probes into

---

[6] *Sharp & Anor v Adam & Ors [2006] EWCA Civ 449 (28 April 2006)* <http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2006/449.html&query=(sharp+OR+adam)>.

[7] Florent Bocquelet et al, ‘Key Considerations in Designing a Speech Brain-Computer Interface’ (2016) 110(4) *Journal of Physiology-Paris* 392 <http://linkinghub.elsevier.com/retrieve/pii/S0928425717300426>; S Geva et al, ‘The Neural Correlates of Inner Speech Defined by Voxel-Based Lesion-Symptom Mapping’ (2011) 134(10) *Brain* 3071 <https://academic.oup.com/brain/article-

the brain itself, in locations known to be active during speech production [8]. The motor areas associated with speech articulators, the jaw, lips, tongue, and so on, are one set of such locations. In recording the electrical activity of these areas while a speaker concentrates on imagining speech, patterns of activity closely matching those produced when actually speaking can be captured.

This approach is effective even where the 'speaker' imagining speech is completely paralysed, unable to move their own articulators. Their motor neural activity can remain intact even without motor function at the articulators themselves. In externalising the neural activity by artificial means, the paralysed body is effectively bypassed and speech activity made possible.

Appropriately processed, and decoded, this information can be used to drive an artificial speech synthesiser. A synthesiser can then be used to produce the speech that the prosthesis-user would have, by outputting a waveform according to the pattern of activity indicated by the neural recording.[9] 'Appropriate processing and decoding' could involve the deployment of "…a statistical language model (giving the prior probability of observing a given sequence of words in a given language)."[10] This kind of model aids in word prediction and aims "To capture important syntactic and semantic information of language… by calculating probabilities of single words and probabilities for predicting words".[11] This gives the advantage of allowing a forward-predicting capability to the speech processing, which ought to make it faster and more reliable. The technical benefits offered by the inclusion of a language-recognising prediction element within the architecture of a 'brain-to-speech'

---

lookup/doi/10.1093/brain/awr232>; Emily M Mugler et al, 'Direct Classification of All American English Phonemes Using Signals from Functional Speech Motor Cortex' (2014) 11(3) *Journal of Neural Engineering* 035015 <http://stacks.iop.org/1741-2552/11/i=3/a=035015>.

[8] Shreya Chakrabarti et al, 'Progress in Speech Decoding from the Electrocorticogram' (2015) 5(1) *Biomedical Engineering Letters* 10 <http://link.springer.com/10.1007/s13534-015-0175-1>; Frank H Guenther et al, 'A Wireless Brain-Machine Interface for Real-Time Speech Synthesis' (2009) 4(12) *PLOS ONE* e8218 <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008218>; Beata Jarosiewicz et al, 'Virtual Typing by People with Tetraplegia Using a Self-Calibrating Intracortical Brain-Computer Interface' (2015) 7(313) *Science Translational Medicine* 313ra179.

[9] Florent Bocquelet et al, 'Real-Time Control of an Articulatory-Based Speech Synthesizer for Brain Computer Interfaces' (2016) 12(11) *PLOS Computational Biology* e1005119 <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005119>; Kristofer E Bouchard et al, 'Functional Organization of Human Sensorimotor Cortex for Speech Articulation' (2013) 495(7441) *Nature* 327 <http://www.nature.com/nature/journal/v495/n7441/full/nature11911.html?foxtrotcallback=true>.

[10] Bocquelet et al, above n 7, 396.

[11] Christian Herff et al, 'Brain-to-Text: Decoding Spoken Phrases from Phone Representations in the Brain' (2015) 9 *Frontiers in Neuroscience* 5 <https://www.frontiersin.org/articles/10.3389/fnins.2015.00217/full>.

processor could have an obscure but potentially very important side-effect associated with it.

Employer and Employee?

Recalling from above that Tamburrini suggests existing legal concepts can be used for BCIs, specifically those associated with employer/employee arrangements, it is worth concentrating on that here. The immediate sense in which a prosthesis user 'employs' their device, and the potential quasi-agency of the technology, give grounds to explore this. Specifically, we are interested in the kind of scenario where an element appears within an otherwise conventional causal chain that may cloud ascriptions of utterances to speakers and the extent to which 'employment' applies in an appropriately analogical sense.

In the advanced directive case above, the clarity of intention expressed is thrown into question through the particular makeup of a form. In the case of the will, the cognitive capacities of the testator are rendered doubtful by a progressive disease. In the case of the steadying hand and the signature, the legitimacy of the document is scrutinised. We suppose that a speech prosthesis could also be considered, in general, as a device that could appear as a point of contention within otherwise conventional causal chains. The nature of the contentiousness is like the cases mentioned as it is likely that the intentions of a prosthesis user, in being processed and expressed artificially, and in being inaccessible but for the device, could be thrown into doubt.

If we refer to very general precepts of vicarious liability law, the relations between a speech prosthesis and its user might be construed as parallel to a kind of employer/employee arrangement. The user *employs their device* in some sense in order to express themselves. This prompts our consideration of matters in terms of vicarious liability.

In producing speech the user can no longer themselves produce, the prosthesis may be taken as akin to an employee charged with a speaking function. All things being equal, where the device functions as planned, the user acts through the device as though acting themselves. Like with an employer/employee relationship, however, '*detour and frolic*' can become relevant. A *detour* represents the carrying out of instructions, but by somewhat unexpected means. Employers remain responsible for actions in such cases. *Frolic* occurs when the actions of the employee cannot be related meaningfully to their employment. The employer can be relieved of liability for acts that depart from employer-mandated action when an employee engages in a *frolic* of their own design.[12]

In terms of vicarious liability, *detour and frolic* are two ways in which deviations from the instructions of a superior, such as an employer, may be considered. As mentioned, '*detour*' is such that the instructions are carried out, though with some collateral action. The employer remains in such a case responsible for the actions performed in the course of events. '*Frolic*', on the other hand, suggests a departure from any reasonable conception of events that could be seen in the light of carrying out given instructions. In the case of frolic,

[12] Young B Smith, 'Frolic and Detour' (1923) 23(5) *Columbia Law Review* 444 <http://www.jstor.org/stable/1112332>.

the employee has 'gone rogue', and in doing their own thing relieves the employer of responsibility.

A *detour*, in the case of a speech prosthesis could be construed as the production of an utterance functionally equivalent to a speaker's intentions, though not necessarily exactly what they wanted to say. In such a case, the prosthesis user could still reasonably be ascribed the content of the utterance. A *frolic* in the case of a speech prosthesis could be construed as an utterance or set of utterances not serving the wishes of the user at all. This could be seen as quasi-agency of the device, suggesting that the learning algorithms of a neural net could happen to lead to speech outputs that are entirely coherent, understandable as speech acts, but divergent from the will of the user.

Could the use of a neuroprosthetic speech device introduce a gap between user intentions and utterances such that ambiguity could become a systematic problem? If so, this could lead to complex issues surrounding the ascription of utterances to the user. If a speech prosthesis user is like an employer (in some appropriate analogous respect) then we could use the analyses of delegated acts apparent in vicarious liability contexts to clarify issues here. This could help us come to a view of whether the ascription to the user of an utterance ought to be made, or whether to the activity of the speech device. In so doing, we would be asking: Could a prosthesis produce speech acts somehow akin to *detour and frolic*?

Respondeat superior?

Mediation of neural activity and the use of prostheses means that effort must be made to 'learn a new language'[13] and 'use the brain'[14] to trigger, then control the device. In this sense it appears that the device is an instrument put to use.[15] However, given this triggering and control is then processed, and an eventual output comes via language predicting software, it appears the device has an element of activity outside the scope of user control. A predictive element is arguably necessary for efficient functioning of neuro-controlled devices in general.[16]

Neural activity processing software may operate according to both generic rules, as discovered in a statistical model of a language, as well as specific rules developed from practice with the device. But this seems to raise the potential of a tension between the faithful reproduction of articulatory activity, and phrase prediction. Ought a prosthesis which uses a language prediction model to be considerable as a quasi-independent actor operating on behalf of its user?

---

[13] Nita A Farahany, 'A Neurological Foundation for Freedom' (2011) 2011 *Stan. Tech. L. Rev.* 11, 10.

[14] Bernard Baars, 'How Brain Reveals Mind Neural Studies Support the Fundamental Role of Conscious Experience' (2003) 10(9–10) *Journal of Consciousness Studies* 100, 9–10 <http://www.ingentaconnect.com/content/imp/jcs/2003/00000010/F0020009/art00008>.

[15] Jonathan R Wolpaw et al, 'Brain–Computer Interfaces for Communication and Control' (2002) 113(6) *Clinical Neurophysiology* 767, 769 <http://www.sciencedirect.com/science/article/pii/S1388245702000573>.

[16] Steinert et al, above n 3.

The answer to this question might depend upon the sophistication of the software used. For instance, that described in Herff et al.[17] appears to be based on the statistical likelihood of two words co-occurring based in specific texts. This is also complemented by a conditional probability of a word occurring given particular neural activity associated with phonetic features. As admitted by the authors, this produces 'very specialized models', but models that are verifiable in practice.[18] Given the use of artificial neural nets in language neuroprostheses in general as a control feature[19] a more sophisticated language model, based in an artificial neural net could be possible and desirable.[20] A suitably complex model and approach could offer the improvements in prediction given by the word co-occurrence method, but with more agility in terms of actual spoken language.[21]

All artificial neural nets can learn from example data and they can generalise from it, treating similar data similarly.[22] In so doing, however, they can become 'black boxes' in the sense that the specific rules they employ at a given time are opaque. Were a neuroprosthetic device to use a sophisticated model of language that could respond to novel data by generalising according to rules learned from training data, this could amount to the sort of quasi-agent problem indicated above. It would seem possible that a tension could emerge between prosthesis-users' speech intent, and the evolving rules of a sophisticated speech prediction model. The nub of the issue would come in how to discern when a model was 'smoothing' user speech that would otherwise be poorly expressed, or when it was departing from user intent to an unacceptable degree.

Where a sophisticated speech prediction model featured prominently, and very actively, in the overall makeup of a neuroprosthesis issues of utterance-ascription become salient as the balance between speaker intent and prediction software may be thrown into question. With an unbalanced relation between prosthetic device and user, not only speaker intent may be at stake, but agency more generally. The phenomenon of 'verbal overshadowing', wherein sincere reports made by an individual are subsumed by 'objective' descriptions given to them by the device, could come into play.[23] The objectivity of a synthesised speech output could overshadow user intent in this way. This perhaps mirrors concerns raised in

[17] above n 11.
[18] Ibid 6.
[19] Bocquelet et al, above n 9.
[20] Dzmitry Bahdanau et al, 'Learning to Compute Word Embeddings On the Fly' [2017] *arXiv:1706.00286 [cs]* <http://arxiv.org/abs/1706.00286>.
[21] Yoshua Bengio et al, 'A Neural Probabilistic Language Model' 1137; Bruno Golosio et al, 'A Cognitive Neural Architecture Able to Learn and Communicate through Natural Language' (2015) 10(11) *PLOS ONE* e0140866 <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0140866>.
[22] RJ Mitchell, John Bishop and W Low, *Using a Genetic Algorithm to Find the Rules of a Neural Network* (1993).
[23] JW Schooler and TY Engstler-Schooler, 'Verbal Overshadowing of Visual Memories: Some Things Are Better Left Unsaid' (1990) 22(1) *Cognitive Psychology* 36.

terms of deep brain stimulation about freedom and responsibility in action following neural interventions.[24]

Moreover, on a neurophysiological basis, the superior temporal gyrus is sensitive to the phonetic features of heard speech,[25] which could be problematic were the effects of the prosthetic's playback to interfere with its recording activity through stimulating neural activity. The speech system, or the users themselves, could become confused over what as being intended as speech, and heard as playback,[26] especially in cases where medication or cognitive issues may play a role. Not least, this would seem highly salient where complicated issues such as informed consent are at stake.[27]

The point here is that a tension could emerge between the faithful recording of neural activity, and the effective decoding and processing of that activity as speech. Given the technical challenge of producing real time speech based on this neural activity, this is not a trivial concern. The predictions of a neural net process geared toward communicative speech might well in principle diverge from intended speech where a set of descriptive rules derived from prior language use, and constituting a model, are used. Ought we to consider the speech as having come from the agent, their device, or some combination thereof?

The difference between something as simple as 'Hello', and 'Hi' could have consequences in terms of formality, depending upon speaker intentions. More complicatedly, the difference between simple and complex forms of assent could be of vital importance. Were an algorithm, in the balance of overall prosthesis function from neural triggering to speech production, to prioritise elegance of speech output based in predicted sentence meanings it could amount to a frolic of its own.

The causal chain from developing an intent and a propositional content issuing from the user is elongated in neuroprosthetically-mediated speech, compared to the more familiar organic state. The device hardware and software can change *in media res* – as it is used, via the proper functioning of learning algorithms for instance – such that the user themselves may not know or notice that things are not the same.

Speech outputs could relate to the user's intent as:

---

[24] Laura Klaming and Pim Haselager, 'Did My Brain Implant Make Me Do It? Questions Raised by DBS Regarding Psychological Continuity, Responsibility for Action and Mental Competence' (2013) 6(3) *Neuroethics* 527 <https://link.springer.com/article/10.1007/s12152-010-9093-1>.

[25] Nima Mesgarani et al, 'Phonetic Feature Encoding in Human Superior Temporal Gyrus' (2014) 343(6174) *Science* 1006 <http://science.sciencemag.org/content/343/6174/1006>.

[26] Timothy J Perfect, Laura J Hunt and Christopher M Harris, 'Verbal Overshadowing in Voice Recognition' (2002) 16(8) *Applied Cognitive Psychology* 973 <https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.920>.

[27] Walter Glannon, 'Ethical Issues with Brain-Computer Interfaces' (2014) 8 *Frontiers in Systems Neuroscience* 2 <http://journal.frontiersin.org/article/10.3389/fnsys.2014.00136/full>.

1. Isomorphic, or identical
    - The speech output by the device is straightforwardly that of the user
2. Functionally or instrumentally equivalent
    - There is a chance of *detour*
3. Deficient in some respect
    - There is a chance of *frolic*

Especially in cases of legally important scenarios, such as end of life decision-making, or the provision of informed consent, these possibilities are in need of expansion, before some general discussion and concluding remarks.

In the case of 1, there appears to be no problem at all. Where the synthesised speech output matched speaker intent perfectly, the job of the system is well done. There remains a question as to the conceptual adequacy of a picture such as this wherein speakers' intent is transparent to them and verifiable in speech behaviour.[28] That issue is beyond exploration here, but it might suffice to say for now that there is an indication here of how tacit presuppositions behind this kind of neuroprosthetic technology can feature.

In the case of 2, a functionally equivalent output could amount to the difference between two utterances that lead to the same outcomes. Saying 'yes' to some proposition has the same effect as saying 'of course!', in many circumstances. In others, it could be more like, 'ok, then.' This would be a problematic area for speech prosthetics given the worries being evolved here on ambiguity and liability. A proposition amounting to a very hedged 'yes' in response to an end of life decision is very significantly different to an unequivocal 'yes'. Binary sorts of communication, as is likely clear anyway, are insufficient for end of life decision-making. Nuanced discussion is required where the legitimacy of the decision situation is to be established.[29] Phraseology here could be 'functionally equivalent' in being affirmative, but significantly different in terms of force or tone. This kind of possibility of detour could be very serious.

In the case of 3, there is a serious problem in that the process control undertaken by the prosthesis has overshadowed the intent of the speaker in some respects and could amount to goal-setting. This kind of concern ought not to be simply disregarded with the thought that one would correct oneself in the event of systematic misrepresentation in speech. The points raised about the role of verbal overshadowing and the role of the superior temporal gyrus in both auditory processing and verbal production serve to underwrite the risks to agency posed here.

---

[28] Peter Carruthers, 'On Knowing Your Own Beliefs: A Representationalist Account' in *New Essays on Belief* (Springer, 2013) 145.

[29] James L Bernat, 'Medical Decision Making by Patients in the Locked-in Syndrome' [2018] *Neuroethics* <http://link.springer.com/10.1007/s12152-018-9358-7>; EJO Kompanje, 'Ethical Decision-Making in Two Patients with Locked-in Syndrome on the Intensive Care Unit' (2009) 4(2) *Clinical Ethics* 98 <https://doi.org/10.1258/ce.2009.009005>; Samuel Maiser et al, 'Locked-In Syndrome: Case Report and Discussion of Decisional Capacity' (2016) 51(4) *Journal of Pain and Symptom Management* 789 <https://www.jpsmjournal.com/article/S0885-3924(15)00637-5/fulltext>.

Interestingly, there is a parallel between 2 and 3 in that each may become problematic in light of speaker intent and responsibility for speech output. That they are distinguishable as detour and frolic respectively, yet each may cause similar problems seems to reflect jurisprudential debate in defining just what constitutes these concepts in vicarious liability.[30]

Consequences for real life scenarios including end of life decision-making are potentially great. By referring to some elements of vicarious liability we may be able to provide a neat analytic method to approach these scenarios. The logic of delegating acts apparent in the prosthesis case, with the potential for quasi-agential prostheses, is already well developed within vicarious liability law. It's not that we ought to apply vicarious liability to prosthesis cases in order to say of users that they are liable for the output of their technologically mediated speech acts as an employer would be for harms wrought by their employees. This would certainly not be the case in medical decision-making contexts, where the patient's perspective ought to be considered very carefully. We ought to ask, when considering these sorts of systems, how much room is there for the retraction of apparent wishes?

The ascription of utterances to prosthesis users may well be complicated or elucidated by family intervention or that of medical professionals. While family may know the patient well, perhaps their own judgement may exhibit the 'disability paradox' whereby quality of life is underestimated for the patient.[31] This underestimation possibility also arises among medical professionals. The cognitive capacity of patients communicating through artificial means may also be downplayed too much, even when intact.[32]

Discussion

Are we communicating with a user of a speech prosthesis, via an informational representation of that user's intent, or is it possible that a neuroprosthesis might systematically depart from the user's intent to various degrees such that it could amount to *detour* or *frolic*?

Where a speech prosthesis uses a sophisticated statistical language model in its architecture this would provide benefits in terms of processing speed and accuracy. Alongside this, however, it could diverge to various degrees from the communicative intent of the user. Where such deviation amounted to *detour* this could stand as an improvement upon the user's intent. Drawing upon a statistical language model could serve as an enhancement to a user's vocabulary, and phrasing. This is an interesting dimension that perhaps requires more scrutiny in neuroprosthesis-mediated language literature.

Beyond *detour*, however, a prosthesis with sophisticated prediction capabilities could in principle take so much process control that it diverts into *frolic* of some kind. Whether a speech prosthesis acts as an instructed employee, or by *detour*, or *frolic*, affects how it relates user instruction to speech outputs. This would be especially so when ascriptions of

[30] William O Douglas, 'Vicarious Liability and Administration of Risk I' (1929) 38(5) *The Yale Law Journal* 584 <http://www.jstor.org/stable/789704>; Smith, above n 12.
[31] Bernat, above n 29; Maiser et al, above n 29.
[32] Kompanje, above n 29.

utterances to agents were acutely important, like informed consent, end of life decision-making. We ought therefore to grant special latitude to users of speech prostheses for determining *post hoc* what they did or did not mean, despite apparent utterances. This ought to come in parallel with careful conduct of discourse with such users.

In terms of medical decision-making, if we consider an agent who wants to say A, but their speech machine says B, we don't as a matter of fact have consent for some procedure. If a physician in good faith believes consent is present, they may be exculpated for acting in the face of such an error. In a conventional case of unclarity or misunderstanding, it perhaps isn't really clear how vicarious liability in general could add something here since the doctor wouldn't be liable for her actions anyway. There is no third party we need to consider. But given the possibility of speech prostheses with sophisticated predictive functions, and with them *detour* and *frolic*, we ought to consider the prosthesis as a potential quasi-agent in the otherwise conventional causal chain. At an analytical level at least, the contribution of this potential quasi-agent can be considered in terms of *detour* and *frolic* in order to clarify ascriptions of utterances to the device user. This same sort of analysis ought to affect the ways in which technologically-mediated ADs are considered.

While it is true that people form contracts all the time, including verbal contracts, ADs may be complicated by various factors such as questions of capacity, procedure, and validity.[33] So there are existing analyses of how we ought to ascribe responsibilities in cases of dictated ADs with unclear utterances. How ought assistive devices to fit in these kinds of scenarios?

The cases above show that precedents exist, and these give us a basis for potentially novel circumstances. The basis in precedent can be bolstered by thinking of an appropriately sophisticated, predictive speech prosthesis in terms of concepts from vicarious liability. We ought not to prescribe too much latitude for questioning the authenticity of contracts assented to via technological means. But given the logic of delegating acts as captured neatly in vicarious liability law, and the theoretical potential for prostheses to overshoot process control and enter a goal-setting mode, we ought to approach these scenarios with eyes open to the specifics of the situation.

It seems that we have to do some projection forward from the state of the art in speech prosthesis technology in order to develop the problem looked at here. Specifically, the inclusion of a complicated speech prediction model forms a central part of what could cause problems. As it has been shown, however, despite this departure into mild science fiction, there is enough to support the projection as very possible. Neural nets are an established technology, even within speech prosthetics, and language prediction models are used. The convergence of these technologies within a future prosthesis appears likely.

It might also seem simply implausible that one could systematically misunderstand someone in a prolonged conversation. The general idea of ambiguity as a feature of language, and the

[33] Guy Widdershoven and Ron Berghmans, 'Advance Directives in Psychiatric Care: A Narrative Approach' (2001) 27(2) *Journal of Medical Ethics* 92 <https://jme.bmj.com/content/27/2/92>.

potential opacity of the rules of neural nets, ought to provide some basis on which this implausibility yields. Moreover, the psychological phenomenon of verbal overshadowing, as well as the neurophysiological role of the superior temporal gyrus, ought to lend credence to the possibility of very sustained misunderstanding between speaker and audience. Taken together, there appears to be good enough grounds to warrant thinking now about how near-future technologies in speech prosthetics ought to be prepared for, especially in terms of important decision-making.

Various convergences of technologies could very well prompt novel types of liability questions, most acutely perhaps when there is an element of quasi-autonomy in predictive language devices. The possibility of *frolic* may represent a limit case, but between that and unproblematic technology-mediated communication lies a spectrum of possible cases. The use of these general principles of liability law may appear somewhat metaphorical, it might be objected. But this is a strength of the analysis at hand. Given the beyond the horizon nature of the likely technological advances under discussion, it seems that broad appreciation of emerging themes is an appropriate way in which to pre-empt potential issues. In this general approach, metaphor allows us to project into a possible scenario without yet being bound to specific details. This in turn permits a flexible approach that can be agile enough to adapt to actual technological changes when they do happen.

Baars, Bernard, 'How Brain Reveals Mind Neural Studies Support the Fundamental Role of Conscious Experience' (2003) 10(9–10) *Journal of Consciousness Studies* 100 <http://www.ingentaconnect.com/content/imp/jcs/2003/00000010/F0020009/art00008>

Bahdanau, Dzmitry et al, 'Learning to Compute Word Embeddings On the Fly' [2017] *arXiv:1706.00286 [cs]* <http://arxiv.org/abs/1706.00286>

Bengio, Yoshua et al, 'A Neural Probabilistic Language Model' 1137

Bernat, James L, 'Medical Decision Making by Patients in the Locked-in Syndrome' [2018] *Neuroethics* <http://link.springer.com/10.1007/s12152-018-9358-7>

Bocquelet, Florent et al, 'Key Considerations in Designing a Speech Brain-Computer Interface' (2016) 110(4) *Journal of Physiology-Paris* 392 <http://linkinghub.elsevier.com/retrieve/pii/S0928425717300426>

Bocquelet, Florent et al, 'Real-Time Control of an Articulatory-Based Speech Synthesizer for Brain Computer Interfaces' (2016) 12(11) *PLOS Computational Biology* e1005119 <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005119>

Bouchard, Kristofer E et al, 'Functional Organization of Human Sensorimotor Cortex for Speech Articulation' (2013) 495(7441) *Nature* 327 <http://www.nature.com/nature/journal/v495/n7441/full/nature11911.html?foxtrotcallback=true>

Carruthers, Peter, 'On Knowing Your Own Beliefs: A Representationalist Account' in *New Essays on Belief* (Springer, 2013) 145

Chakrabarti, Shreya et al, 'Progress in Speech Decoding from the Electrocorticogram' (2015) 5(1) *Biomedical Engineering Letters* 10 <http://link.springer.com/10.1007/s13534-015-0175-1>

Douglas, William O, 'Vicarious Liability and Administration of Risk I' (1929) 38(5) *The Yale Law Journal* 584 <http://www.jstor.org/stable/789704>

Farahany, Nita A, 'A Neurological Foundation for Freedom' (2011) 2011 *Stan. Tech. L. Rev.* 11

Geva, S et al, 'The Neural Correlates of Inner Speech Defined by Voxel-Based Lesion-Symptom Mapping' (2011) 134(10) *Brain* 3071 <https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awr232>

Glannon, Walter, 'Ethical Issues with Brain-Computer Interfaces' (2014) 8 *Frontiers in Systems Neuroscience* <http://journal.frontiersin.org/article/10.3389/fnsys.2014.00136/full>

Golosio, Bruno et al, 'A Cognitive Neural Architecture Able to Learn and Communicate through Natural Language' (2015) 10(11) *PLOS ONE* e0140866 <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0140866>

Guenther, Frank H et al, 'A Wireless Brain-Machine Interface for Real-Time Speech Synthesis' (2009) 4(12) *PLOS ONE* e8218 <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008218>

Herff, Christian et al, 'Brain-to-Text: Decoding Spoken Phrases from Phone Representations in the Brain' (2015) 9 *Frontiers in Neuroscience* <https://www.frontiersin.org/articles/10.3389/fnins.2015.00217/full>

Jarosiewicz, Beata et al, 'Virtual Typing by People with Tetraplegia Using a Self-Calibrating Intracortical Brain-Computer Interface' (2015) 7(313) *Science Translational Medicine* 313ra179

Klaming, Laura and Pim Haselager, 'Did My Brain Implant Make Me Do It? Questions Raised by DBS Regarding Psychological Continuity, Responsibility for Action and Mental Competence' (2013) 6(3) *Neuroethics* 527 <https://link.springer.com/article/10.1007/s12152-010-9093-1>

Kompanje, EJO, 'Ethical Decision-Making in Two Patients with Locked-in Syndrome on the Intensive Care Unit' (2009) 4(2) *Clinical Ethics* 98 <https://doi.org/10.1258/ce.2009.009005>

Maiser, Samuel et al, 'Locked-In Syndrome: Case Report and Discussion of Decisional Capacity' (2016) 51(4) *Journal of Pain and Symptom Management* 789 <https://www.jpsmjournal.com/article/S0885-3924(15)00637-5/fulltext>

Mesgarani, Nima et al, 'Phonetic Feature Encoding in Human Superior Temporal Gyrus' (2014) 343(6174) *Science* 1006 <http://science.sciencemag.org/content/343/6174/1006>

Mitchell, RJ, John Bishop and W Low, *Using a Genetic Algorithm to Find the Rules of a Neural Network* (1993)

Mugler, Emily M et al, 'Direct Classification of All American English Phonemes Using Signals from Functional Speech Motor Cortex' (2014) 11(3) *Journal of Neural Engineering* 035015 <http://stacks.iop.org/1741-2552/11/i=3/a=035015>

Perfect, Timothy J, Laura J Hunt and Christopher M Harris, 'Verbal Overshadowing in Voice Recognition' (2002) 16(8) *Applied Cognitive Psychology* 973 <https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.920>

Schooler, JW and TY Engstler-Schooler, 'Verbal Overshadowing of Visual Memories: Some Things Are Better Left Unsaid' (1990) 22(1) *Cognitive Psychology* 36

Smith, Young B, 'Frolic and Detour' (1923) 23(5) *Columbia Law Review* 444 <http://www.jstor.org/stable/1112332>

Steinert, Steffen et al, 'Doing Things with Thoughts: Brain-Computer Interfaces and Disembodied Agency' [2018] *Philosophy & Technology* <http://link.springer.com/10.1007/s13347-018-0308-4>

Tamburrini, Guglielmo, 'Brain to Computer Communication: Ethical Perspectives on Interaction Models' (2009) 2(3) *Neuroethics* 137

Widdershoven, Guy and Ron Berghmans, 'Advance Directives in Psychiatric Care: A Narrative Approach' (2001) 27(2) *Journal of Medical Ethics* 92 <https://jme.bmj.com/content/27/2/92>

Wolpaw, Jonathan R et al, 'Brain–Computer Interfaces for Communication and Control' (2002) 113(6) *Clinical Neurophysiology* 767 <http://www.sciencedirect.com/science/article/pii/S1388245702000573>

The X Primary Care Trust v XB & Anor [2012] EWHC 1390 (Fam) (Unreported, EWHC (Fam), MRS. JUSTICE THEIS, 1 May 2012)

Barrett v Bem & Ors [2012] EWCA Civ 52 (31 January 2012) <http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2012/52.html&query=(barrett+OR+v+OR+bem)>

Peter, Bowal and Bontorin Louis, *Vicarious Liability: The Legal Responsibility of Employers - LawNow Magazine* (9 August 2018) lawnow.org <https://www.lawnow.org/vicarious-liability-legal-responsibility-of-employers/>

Sharp & Anor v Adam & Ors [2006] EWCA Civ 449 (28 April 2006) <http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2006/449.html&query=(sharp+OR+adam)>