Article | Received 24 October 2025; Revised 14 February 2026; Accepted 24 February 2026; Published 11 May 2026
https://doi.org/10.55092/let20260005

Unspoken testimony: why neural speech decoding does not produce testimonial evidence

Tristan Baujault Borresen[1,*] and David Levy[2]

[1] Faculty of Law, University of Ottawa, Ottawa, Canada
[2] School of Philosophy, Psychology and Language Sciences (PPLS), University of Edinburgh, Edinburgh, UK

* Correspondence author; E-mail: tristan.b.borresen@uottawa.ca.

Highlights:

- Reconstructing someone's inner speech via neural speech decoding is not testimony in law.
- There is an intrinsic epistemic distinction between decoding thoughts and articulating thoughts.
- Ignoring the distinction leads to the fallacy of mistaking someone's mere thoughts as beliefs.
- Falling into this fallacy leads to the injustice of disregarding an individual's epistemic authority.
- The authors suggest NSD resembles hearsay in law, rather than testimony.

Abstract: This article warns against the idea that information from neural speech decoding (NSD) can be treated as testimonial evidence. We contend that NSD and testimony differ in their epistemic status. While the evidential validity of testimony depends essentially on someone's sincerity, by contrast NSD depends upon the accuracy of decoding the person's inner speech. To show this, we first explore how NSD poses a problem for rights against self-incrimination by testimony in, for instance, the Fifth Amendment to the US Constitution. Building on Pardo's epistemic account of the distinction between physical and testimonial evidence, we argue that since NSD produces ambiguous information, it does not qualify as testimony. Conversely, denying our claim makes one liable to what we call the informational symmetry fallacy (ISF), viz. the conflation of decoded output from someone's brain with that person's speech act. We then set out the consequence of falling into the ISF by criticizing Farahany's proprietary conception of thought. Her conception is shaped by goal of providing a new taxonomy of evidence better suited for relating constitutional rights to brain data and privacy. Here, we argue that Farahany must presume that we *already* own and are thereby accountable for all of our intercepted thoughts. Hence, she precludes the possibility of our *disclaiming* some thoughts and thereby justifies others in treating those thoughts as our beliefs, without considering the dependence of belief our epistemic authority. Lastly, we discuss a practical problem entailed by our account, namely that users of neural speech prostheses will not qualify as witnesses in a legal setting. In response, we argue that although NSD prosthetics enable users to communicate, there are insuperable limits in assessing someone's testimonial sincerity solely on the basis of an NSD readout. For this reason, we conclude that NSD is better analogized to hearsay evidence.

Keywords: the epistemology of mindreading; neural speech decoding; testimonial evidence; law of evidence and technology; hearsay; the right against self-incrimination; epistemic authority; epistemic injustice; neuroethics in the age of AI

1. Introduction

Ludwig Wittgenstein stated, "If I listened to the words issuing from my mouth, then I could say that someone else was speaking out of it" [1].

The advancements permitting the inferring of inner speech from brain signals, neural speech decoding (NSD), challenges the notion of testimony in law and epistemology. To testify is typically to put into words with one's voice one's convictions, memories, desires, *etc.* But what if by the scanning of one's brain someone else could do this independently of one's vocalizations, not to mention one's will?

In this article, we warn against the idea that information from NSD could produce testimonial evidence. Our contention is that NSD and testimony differ in their epistemic status. While the evidential validity of testimony depends essentially on a person's sincerity in reporting their beliefs, NSD is evaluated on the accuracy of decoding that person's inner speech.

To show this, we first explore how NSD poses a problem for rights against self-incrimination in, for instance, the Fifth Amendment to the US Constitution which protects individuals from testifying against themselves. Building on Pardo's epistemic account of the distinction between physical evidence and testimonial evidence, we argue that since NSD produces ambiguous information, it does not qualify as testimony. Conversely, in rejecting our claim, one becomes liable to commit what we call the informational symmetry fallacy (ISF): the conflation of the decoded output from someone's brain with that person's speech act.

We then set out the consequence of falling into the ISF by criticizing Farahany's proprietary conception of thought. Her conception is shaped by goal of providing a new taxonomy of evidence better suited for relating constitutional rights to brain data and privacy. Here, we argue that Farahany must presume that we *already* own and are thereby accountable for all of our intercepted thoughts. Hence, she precludes the possibility of our *disclaiming* some thoughts and thereby justifies others in treating those thoughts as our beliefs, without considering the dependence of belief our epistemic authority.

Then we discuss a practical problem entailed by our account, namely that users of neural speech prostheses will not qualify as witnesses in a legal setting. In response, we argue that although NSD prosthetics enable users to communicate, there are insuperable limits in assessing someone's testimonial sincerity solely on the basis of an NSD readout. For this reason, we conclude that NSD is better analogized to hearsay evidence.

2. Why NSD does not yield testimonial evidence.

It is through speech that we make ourselves known. It is in formulating words that I confess my sins to my confessor; I engage in small talk with my neighbor about what I did last week; I text my girlfriend to say that I love her; and others ascertain my beliefs and intentions.

This ability for self-expression is especially necessary when testifying. Judges expect that witnesses can formulate their own account of events; that they do not speak on somebody else's behalf; and that they are, thus, able to speak for themselves.

Self-expression is naturally impracticable for individuals with complete speech paralysis, such as locked-in syndrome patients—people who are lucidly conscious but lack all or nearly all motor abilities to communicate. Trapped inside their body, alone with their thoughts, they face the total impossibility of expressing themselves to that external world which they are also unable to act upon. Jean-Dominique Bauby used the sole mobility in his body, his eyelid, to spell his novel *The Diving Bell and the Butterfly*, noting tragic-comically the alienation of his peers when they tried to decipher the blinking of his eye:

I understood the poetry of such mind games one day when, attempting to ask for my glasses [in French, *lunettes*], I was asked what I wanted to do with the moon [*lune*] [2].

New neurotechnology may, however, bridge the gap between people with speech paralysis and the outer world. NSD promises to reconstruct one's inner speech by establishing a correlation between brain activity and speech processing via neuroimaging and artificial intelligence. More precisely, NSD involves a reverse inference model that encodes the correlations between speech perception and the activation of areas in an individual's brain to predict that individual's inner speech based on their recorded brain activity.

In 2023, through fMRI recordings of subjects' brains while listening to 16 hours of podcasts, Tang *et al.*, succeeded in non-invasively encoding the neural activation of speech perception [3]. Then, using a large language model, well-formed English sentences were extracted from the encoded language stimuli to decode the subjects' brain signals into continuous word sequences. Another recent study from 2025 succeeded in decoding speech from the subject's brain almost instantly, thereby coming closer to realizing the goal of returning a voice to the speechless by using a neural speech prosthesis [4].

2.1. The testimonial problem

Whatever the medical application of these technical advances, the possibility of decoding somebody's inner speech suggests that these studies in NSD are developments towards a form of mindreading[1]. Legal scholars have raised concerns therefore about whether today's legislation can prevent the state from using NSD to prosecute citizens by retrieving information about their mental life without their consent or even awareness through brain decoding. The most ink has been spilt arguing for extending or revising current civil rights regarding privacy and freedom of thought to protect against illicit uses of mindreading technology for forensic evidence [5–7]. However, surprisingly few have discussed what mindreading *means*. While some have noted that 'thought' is difficult to define in a legal framework [8], they have ignored whether the interception of thought is also the interception of speech, and therefore of testimony. For instance, in a scenario where law enforcement uses NSD to retrieve information about your thoughts in the interrogation room, are there good grounds for accepting that the police officer has coerced an attestation from your allegedly criminal mind?

Our answer in this article is no, we do not. Mindreading does not produce testimonial evidence. Indeed, mindreading may well undermine the witnesses' own authority in determining their beliefs. Our focus here is on the conceptual possibility of using NSD as a discursive form of mindreading that

[1] "Mindreading" here refers to inferences from brain activity, not from outward behavior as with psychology. We elaborate on why NSD can be characterized as mindreading in section 1.4.

produces compelled evidence rather than technical questions regarding the putative limitations in current neurotechnology to realize mindreading[2].

To make our case we will first explain how NSD threatens rights against self-incrimination. Our argument will proceed using Pardo's account of testimony as evidence[3]. More specifically, we use Pardo's distinction between physical and testimonial evidence to argue that NSD output cannot meet the defined criteria for testimony [11]. Pardo's account of testimony is especially apt for our discussion because of his focus on the distinctive epistemic character of testimonial evidence. In particular, Pardo insists that testimonial evidence is constrained and shaped by its connections with confrontation requirements and self-incrimination protections [12]. The discussion of putative testimony derived from NSD mindreading directly raises concerns about self-incrimination and whether testimony unspoken by a witness in a court can meet confrontation requirements. Indeed, NSD implies the possibility of testimonial evidence that has *never* been uttered or otherwise expressed by the witness, which evidence can also be sincerely disclaimed by the witness at the moment of its extractive presentation. So, the distinctive foci motivating Pardo's account are to the fore in our context.

Pardo's account of testimonial evidence is of course one of many. Lackey, for instance, offers a *philosophical* account of testimony that, unlike ours or Pardo's, downplays the importance of the witness' belief in her statements. She holds that a hearer can acquire knowledge of a fact through a speaker's testimony despite the speaker failing to know or believe that very fact [13,14].

Similarly, our treatment of testimony may conflict with some *legal* definitions of testimony. As Veas reports, in Germany 'oral statements' have been extended to other acts that require the defendant to perform an activity such as producing incriminating documents, actively participating in a psychological or psychiatric examination, providing a voice or handwriting sample, or providing a urine sample. Even exhaling has been considered an "activity," meaning that no one can be compelled to provide a breath sample for analysis [15].

However, in the second part, we present problems for both this philosophical and legal understanding of testimony. Unlike Pardo's account, they are problematic right where one might expect, namely in *assessing* violations against rights such as that against self-incrimination. For if speakers are said to transmit testimony to hearers without regard for the speaker's knowledge or intent, or that *any* activity might be taken as speaker's testimony, then there is no constraint on what might be attributed to speakers as relevant, though legally protected, information. This also makes obscure why testimony is treated differently compared to, e.g., a blood test. Moreover, treated in this way, why accord any authority to a speaker in clarifying the meaning of their testimony? Indeed, there would be no basis for thinking of any information thus derived as being outside the ken of others.

For these reasons, we base our arguments about NSD on Pardo's account of testimonial evidence. We extend and elaborate that account to demonstrate and analyze the nature of the problem for testimonial evidence posed by NSD, laying special emphasis on the role of epistemic authority. After a fashion, our article is a proof by elaboration of the challenge posed by NSD. While our elaboration is fine-grained in developing the consequences of using NSD mindreading for evidence, testimonial or

[2] Scepticism towards forced mindreading arises since, as Haynes [9] notices, subjects must during brain scanning be completely cooperative. Moreover, the term 'mindreading' has also evoked a fear of hyping neurotechnology. Gilbert and Russo [10] note that the use of this term has increased in recent years. In this paper, we do not address mindreading as such but rather the implication of *presupposing* that one has got 'access' to another's mind through sophisticated neurotechnology.
[3] See [11].

otherwise, we do not address the other harms or wrongs that might attend the use of NSD or similar neuro-technologies.

2.2. Mindreading and Self-incrimination

There is a long-standing discussion in jurisprudence around the definition of testimony as technology increasingly pervades our lives. From India, Germany and the EU to the United States, the standard concept of testimonial evidence as an oral statement is currently contested by cases of compelled decryption of smartphones via fingerprints and other biometrics[4], and search and seizure through brain imaging to test an individual's innocence[5].

In reality then, the use of NSD as coerced mindreading in courtrooms is a controversial, albeit hypothetical, continuation of the debate about what kinds of evidence are admissible for establishing the guilt or innocence of the accused. Especially in India, this debate has been heated because of the increasing employment of neurologically-based guilty knowledge tests in legal cases. Here, as Lokaneeta reports, since the early 2000s, there has been a proliferating use of brain electrical oscillation signature (BEOS) and brain fingerprinting (BF) to test whether a person has participated in, or has experiential knowledge of, a crime (see [19], pp. 1–2)[6]. Such testing is sometimes done voluntarily. Chandler describes a case where an individual who maintained he was falsely accused insisted on undergoing a polygraph test to demonstrate his innocence [21]. Provided with the possibility of coercively decoding someone's thoughts with NSD, the state might not only obtain information about *whether* the accused has knowledge of the crime but also *what* they know about it.

This is why NSD provides a deeper jurisprudential challenge than prior neurological methods already used in law, most notably to freedom of expression. This right not only grants citizens the ability to freely express themselves in public and private but also *not* to speak. In the United States, the detained can refer to the Fifth Amendment preventing individuals from testifying against themselves to avoid self-incrimination. According to Fifth Amendment jurisprudence, this prevents a defendant from being compelled to testify "out of his own mouth" [22], though they may have to provide physical or real evidence, like X-rays, DNA, or saliva[7].

But what about information derived from brain scans? Adding scans to the latter category of physical evidence entails, as Murphy and Greely comment, that "A court might plausibly conclude that the results of a mindreading device would not be 'testimonial' evidence and so would not be covered by the Fifth Amendment" [24].

Remarkably, Murphy and Greely reject this conclusion, finding it more likely that mindreading will be classified as testimonial evidence and thus protected by the Fifth Amendment [24][8]. While their claim is *predictive* in terms of whether juries will employ mindreading to criminalize individuals, there seems

[4] For the discussion on self-incrimination and forced decryption in the US, see [16]; in the German perspective, see [17].
[5] See Ligthart [18] for the discussion on the European perspective on expanding the right against self-incrimination to coerced brain imaging.
[6] After the landmark decision in 2010 from Supreme Court of India of the case *Selvi & Ors v. Karnataka & Anr*, brain mapping was ruled out as unconstitutional. However, Nair, Weisse and Chandler [20] found that the incriminatory use of neurological evidence between 2002 and 2023 has only been increasing.
[7] See the Supreme Court commentary in *Schmerber v. California* [23].
[8] Murphy and Brown's [25] claim is more affirmative. They write, "If the methods for producing the [brain] images involve the subject declaring or testifying to something as part of the experimental stimuli, the resulting [brain] image may very well be categorized as a declaration or statement."

to be no prescriptive argument for why they *should not*. For can the decoding of imagined speech from a person's *brain* ever be identical to words produced from the person's own *mouth*? Or does the decoded outcome rather reflect the *neurophysiological* imprint of speech? In other words, does the policeman's intrusion into the individual's mind conflict with the individual's right to remain silent?

To settle these questions, we need to clarify why and how we establish this dichotomy between physical and testimonial evidence.

2.3. The physical-testimonial distinction

At first, the reason for this distinction can be found in the presumption of innocence by which "everyone charged with a criminal offence shall be presumed innocent until proved guilty according to the law" [26]. In that regard, the burden of proof must be calibrated so that it does not contravene this presumption while not making the proof of the defendant's guilt impossible. Hence, the burden of proof must be *independent* of the defendant in the prosecution. As Laudan explains, the presumption of innocence can be interpreted using a material or probatory understanding [27]. In the former understanding, the defendant just *is* innocent of the crime unless proven guilty; whereas the latter requires that fact-finders make the prosecution establish guilt beyond a reasonable doubt, free from any prior suppositions about culpability.

On the probatory understanding, Pardo argues that the physical-testimonial distinction should depend on what or who holds the epistemic authority to determine the evidence [11]. For instance, if the nurse tells Isabel:

(1) "The thermometer shows that your body temperature is 39 °C," we rely on, or, as Pardo words it, "pass the epistemic buck" to the accuracy of the device in measuring thermodynamical changes. But if Isabel then tells the nurse:

(2) "How strange! I don't feel feverish."

We would grant that Isabel has the last word in determining the truth about whether she has the sensation of feeling feverish.

Since in (1) the epistemic authority derives from physical changes *independent* of Isabel, neither she nor anyone else can determine the truth of (1). Drawing from Grice's pragmatism, Pardo argues that an example like (2) is testimonial evidence as the ultimate epistemic authority derives from Isabel's *communicative intention*. According to Grice, a testimonial utterance depends on intending to mean something to an audience who form a belief based on the testifier's intent [28]. For example, by uttering (2), Isabel most probably would have occasioned the nurse to believe that Isabel is surprised about having a 39 °C body temperature as that would mean having a fever contrary to her own sense of being in good health. And in that case, the nurse would not only have recognized the meaning in uttering (2) but also relied on Isabel in establishing the *truth* about her sensed well-being.

We argue that Pardo's epistemological account for the physical-testimonial distinction can be clearer if we also observe that not only do (1) and (2) differ over the authority of the evidence, but also in *meaning*. Compare, for instance, the two different uses of 'means' in (3) and (4):

(3) 'That you have a 39 °C temperature' *means* that you have fever.

(4) 'How strange! I don't feel feverish' *means* that Isabel is surprised about having fever.

The difference between (3) and (4) is to be found in what Grice calls *natural* and *non-natural* meaning [26]. Where 'means' in (3) indicates a natural correlation between body temperature and inflammation, 'means' in (4) covers the correlation between the speaker's utterance and the speaker's

intended meaning. If the meaning of (3) is natural, it is because denying that "'39 °C temperature' means 'having a fever'" would in all circumstances be inconsistent with its meaning. However, denial is still consistent with the non-natural meaning of (4) since in denying that Isabel is surprised about having a fever; she might have been sarcastic, or that, on another occasion, it would have been a comment about watching the film *Saturday Night Fever*. Where the falsehood of (3) would make our current understanding of fever unintelligible, denying (4) only entails a reconsideration of the speaker's intention, not the meaning of the utterance by which the speaker manifests her intention.

This difference between natural and non-natural meaning aligns with what we saw with (1) and (2). As Pardo observes, we would grant the assertion in (2) as testimonial evidence given only Isabel's *sincerity*. The question of sincerity arises because Isabel might lie about her health. However, if it were false, it would still not change the meaning of (2). Rather, the possibility of being false is a condition on its meaningfulness. On the other hand, the condition for relying on (1) is *accuracy* in measuring the subject's body temperature. Since error is intolerable here, no question about sincerity arises at all.

It is important to notice that for an audience, sincerity is not about trusting that a person is speaking the truth but rather that she says what she truly *believes*. We grant that Isabel is sincere by trusting that she is truthfully conveying what she believes to be the case about her own health. Granting epistemic authority to a witness and trusting the witness' sincerity are therefore two sides of the same coin: it as, as Moran writes, the process in which one is "making oneself *accountable* for the truth of what is claimed, so that the speaker shall be held to account both for the meaning of the claim and for its truth should it turn out to be false" (see [29], p. 138).

Thus, together with Pardo's distinction, we propose therefore two necessary criteria for characterizing the difference between physical and testimonial evidence. Let *o* be a piece of evidence and *S* be a subject, then:

(i) *o* is evidence of *S's testimony* if and only if the authority to determine the truth of *o* is *S.* If otherwise, then *o* is *physical* evidence.

(ii) *o* is evidence of *S*'s *testimony* if and only if *o* has non-natural meaning, *i.e.*, the possibility of denying *o* is consistent with the meaning of *o*. If otherwise, then *o* is *physical* evidence.

These two clauses highlight the principal difference between physical and testimonial evidence, namely that they are evaluated using two different modalities of reliability: *accuracy* and *sincerity*. This is the difference between presuming that the source of evidence is right in stating a fact and asserting a belief about a fact. Crucially, 'asserting a belief' encompasses the idea of *agency*—viz. that testimony must be an intended *speech act* necessarily directed to an *audience*—in contrast to facts that obtain whether acknowledged by others or not.

To illustrate this distinction, Chandler and colleagues report a case where the defendant was asked to present information stored in his pacemaker about his heartrate at the time of a fire in his house [30]. The defendant refused, referring to his testimonial privilege, but his appeal to privilege was rejected. Given Pardo's account, we see here that rejecting his appeal to privilege was correct because (a) the defendant had no personal authority in determining the heartrate; and (b) the authority of the evidence rested upon the accuracy of the pacemaker, not the defendant's sincere assertion of his belief about his heartrate.

With a clearer understanding of the physical-testimonial dichotomy in hand, the question is how well NSD fits into one of the categories. To that end, we shall make a further distinction between 'brainreading' and 'mindreading'.

2.4. Mindreading is not testimonial

Following Roskies, reading minimally requires "a mapping of a physical pattern to meaning" [31]. She argues that the spectrum between brainreading and mindreading concerns the granularity of the content discerned by measuring brain activity. Brainreading comprises "rough and brute-force empirical correlations between measurements of the physical state and mental functions." This could be detecting the correlation between a hyperactive hemodynamical flow in the amygdala and stress responses such as anxiety [32]. Here, given a correlation between stimuli and the activation of particular brain areas, a brainreading would indicate *that* the subject is anxious, but not *what* the subject is anxious *about*.

Learning the object of one's mental state requires the reading of one's mind. Mindreading, Roskies argues, would include a reliable correlation between propositional content and brain activity patterns. An example of mindreading would be identifying people's memories, such as in the 'fMRI eyewitness test' conducted by Uncapher *et al.* [33]. In one of their trials, participants saw a series of faces and their memory-based recall of these faces was tested in two distinct ways while their brain activity was being scanned with fMRI. Such techniques might be used to retrieve information about a crime scene or to determine the reliability of an eyewitness.

However, as Murphy and Greely notice, the problem with mindreading is the specificity of the propositional content [24]. If Isabel had been thinking of 'getting an ice-cream sundae,' would we accept as accurate a device's retrieval of 'getting food' from Isabel's mind? Indeed, in the famous NSD study by Tang *et al.*, the model was most successful when reconstructing the *gist* of phrases from the participants' inner speech [3]. Yet such approximations could have cruel consequences. About an article in *Nature*, Roskies remarks about the findings of the study that "'I just jumped out [of the car]' was decoded as 'I had to push her out of the car'", adding that "The differences are stark enough they could make an enormous difference in a legal case […] I'm afraid they will have the ability to use this stuff when they shouldn't" [34].

If we sideline these scruples about the practical limitations of mindreading, we see that NSD qualifies as a discursive form of mindreading in that its decodes at least the gist of someone's imagined speech and, accordingly, the propositional content of that speech formulated in one's head. Stipulating the possibility of coercively intercepting imagined speech with NSD, would the decoded readout qualify as testimonial evidence given the definition above?

At first blush, compelled testimonies have just the same epistemic authority as voluntary ones. Consider coercion with threats. The interrogator may tell Isabel that in another detention room sits her whole family. If she does not tell him the names of those in her anti-government cell, he will kill her family[9]. Foreseeing the unbearable consequences of remaining silent or telling a lie, Isabel decides to confess the truth:

(5) "Pablo was the head of the bomb attack planning!"

In the interrogator's eyes, (5) is a sincere statement because it relies on Isabel's *choice* to be truthful and her perfectly human desire to keep her family alive. And these presumptions reinforce the interrogator's belief that he can take Isabel's word as good. Thus, the presumption of sincerity in compelled testimonies stands.

[9] In the US legal system, it is permissible for law enforcement agents to lie to the detained in order to pressure or deceive the detained into telling the truth (see [35]).

But once a bit of language is detached from the person's speech act, its meaning becomes ambiguous. This is salient for evidence taken from NSD. In the parallel situation with NSD, the police officer takes the decoded speech as evidence of the accused's testimony. Here, the officer will not rely on the accused's sincerity, depending instead on the *accuracy* of the (hypothetical) NSD device to reconstruct her thoughts, without regard for the individual's unwillingness to speak and consequent silence. However, this is to mistake the evidential modality of physical evidence as applicable to testimony. That is, the police officer, by using the wrong evidential criteria, treats as testimonial evidence what is physical evidence.

This conflation also extends to the semantics of testimony. If the NSD device were apt for intercepting Isabel's thought as formulated in (5), then (5) has natural meaning because it is inferred from physiological processes in the brain. Thus, it would be inconsistent with the (natural) meaning of (5) that it be false, if we further assume that Isabel knows who the head bomb planner is and is incorrigible about this fact. If as a matter of face (5) as given by the NSD were false, this would demonstrate that the NSD was no longer accurate in decoding people's inner speech and in consequence would lose its standing as a source of evidence.

Conversely, if (5) had non-natural meaning, then by definition it will have been conveyed with a communicative intention. But if the evidence is assessed by the exactitude of decoding the person's thoughts using NSD, then it does not matter what the person's communicative intention was because none of those will bear on the meaning. Indeed, the rise of lie detectors was precisely to spare the trouble of having to trust unreliable witnesses [36].

Considering this ambiguity in both the modality of evaluating the evidence and the semantics of the evidential information, we argue, contrary to Murphy and Greely, that the NSD readout cannot be considered forced testimonial evidence. Crucially, because the epistemic status is the same for compelled as well as voluntary testimony, we maintain that even if Isabel were to consent to having her mind read by an NSD device as a part of the interrogation procedure, the outcome would still not count as testimony. When trusting the NSD the question of its decoding accuracy is the sole question, not the witness's sincerity. Conversely, if we take Isabel at her word, the NSD cannot make a further contribution to uncovering the content of her mind. *A fortiori,* prosecutors would not conflict with the individual's testimonial privilege by using coerced NSD, because what will have been coerced is not testimony.

2.5. The informational symmetry fallacy

Now, someone who insists against us that an NSD readout may have the character of testimony must presume that sincerity is not a necessary condition for testimonial evidence. But this is to presume that intercepting the propositional content of a *thought* is sufficient to characterize what a person intends to *assert* in the performance of a speech act.

Imagine then I am having the thought in (6) that was intercepted by the police:

(6) The Holocaust didn't happen.

Granting the mindreading-as-testimony view, this suffices to attest that I am a Holocaust denier—albeit I might never agree to the statement, nor would I ever *intend* to say that. Indeed, as in (6), there are many thoughts that I consciously and meticulously formulate "in my head" that I would never express, nor attribute to myself as among my beliefs. But mistaken attributions of this kind would readily occur if the interception of thought is mistaken as the interception of an avowal, an assertion or of testimony.

Hence, by assuming the testimonial character of mindreading, one commits what we labelled above the ISF. It is to equate the information that there is such-and-such mental content with the information that there was a mental act of commitment to that selfsame mental content; to confuse *entertaining* a thought with *endorsing* it; to confuse my *thinking* of *p* with my *judgment* that *p*. In short, ISF is the conflation of the decoded output from someone's brain with that person's speech act.

Defining the ISF reinforces the essential difference between bare thoughts and considered beliefs. As Hampshire observes, thoughts to which we are committed are *beliefs* insofar as "A belief is a thought from which a man [sic] cannot dissociate himself" [37]. While we can dissociate ourselves from silently thinking (6), once we are willing to utter it as testimony, we are committed by others and ourselves to take it as expressing *our* belief. Obviously, just as one could entertain (6) when considering someone else's view, one could also covertly commit oneself silently to (6) in one's head, so to speak, thus making it part of one's own view. However, *evidence* for one such commitment must be overt in acts of speech or deed. It is by this overt commitment that I make myself accountable for the meaning and truth of that statement.

To summarize: we have argued that given the physical-testimonial distinction, NSD as a discursive form of mindreading does not produce testimony for two reasons: (a) the witness' epistemic authority is absent; and (b) the content of the NSD readout has natural meaning in contrast with the non-natural meaning required in testimony. Arguing that NSD *is* testimonial leads to the ISF, viz. conflation of the decoded output from someone's brain with that person's speech act, which shows itself by conflating the presence of a thought with the commitment to that thought. In the next part, we will demonstrate what follows from ignoring this fallacy.

3. The unacceptable consequence of treating NSD as testimony.

We noted above that in taking NSD as testimonial evidence, the error of law is not that arguing for a person's guilt on compelled, mind-decoded evidence moves the burden of proof onto the defendant. Rather, it lies in *bypassing* the defendant by making a claim about her guilt that should depend on her epistemic authority alone. Thus, the legal problem is not that the police force her to endure a mindreading, but that they allege knowledge of what she believes with no dependence on her intention to form, or to have formed, such a belief. In consequence, what is at stake is not the threat of involuntarily testifying against oneself, but the loss of freely determining what one believes.

We will show below the intolerable consequences of falling into the ISF by discussing Farahany's views, taking her as a proponent of the mindreading-as-testimony view. She proposes replacing the physical-testimonial distinction with a new classification on the basis of the *origination* of the evidence. Here, we emphasize that in order to discard the distinction, Farahany rejects our contentions that:

(iii) Speech is non-physical (non-natural) evidence, and

(iv) testimony is necessarily a communicative act.

In considering Farahany, we show that her account not only leads to the ISF but also a much greater injustice: precluding the disclaiming of our own thoughts, thereby vitiating people's epistemic authority. To this end, we will first rehearse the motivations for Farahany's account and then its logical inconsistencies.

3.1. A proprietary taxonomy of evidence

Farahany observes that neurological evidence such as NSD cannot ground the physical-testimonial distinction. Nevertheless, in contrast to our approach of focusing on whether and how NSD might be legitimate evidence, Farahany takes for granted the legitimacy of NSD as a form of evidence by introducing a new taxonomy that encompasses this type of information in law.

The main criterion for this classification is the *origination* of evidence. Here, Farahany associates 'origination' with one's *control* in bringing about the evidence. This is schematized as a hierarchy with four categories of information, in which the individual's control over the production of evidence increases progressively from the first to the last category. Her taxonomy begins with 'identification', comprising samples of physical matter that identify a person, like fingerprints or, in the neuroscientific field, brainprints; then 'automatic' which Farahany associates with automatic responses, like the case of detecting sexual preferences in the brain, or uncontrolled reflexes; followed by 'memorialization' comprising the extraction of memories from an individual's brain; and finally, 'utterance'. Utterances include "thoughts, visual images, words, or statements that are verbalized or recalled to the conscious mind, whether spoken aloud or ruminated on silently in the brain" (see [38], p. 1298).

Farahany argues that authorities only commit a legal offence in compelling one to deliver information in cases where one is the originator of the evidence. This licenses delivering identifying and automatic evidence as this includes facts in the world that can be uncovered, whose production occurs without any controlled intervention by the subject. Conversely, since subjects have the most control in producing utterances and memorizations, any involuntary retrieval of these kinds of information would be unlawful.

What ensures such protections, Farahany argues, is not only the right against self-incrimination covered in the Fifth Amendment to the US Constitution [39] but also the Fourth Amendment's provision against unreasonable search and seizure of private property [38]. She finds that the Fifth Amendment and Fourth Amendment are intrinsically linked since compelling one to confess against oneself seems not only an act of self-incrimination but also an unwarranted interference with one's privacy [38].

Therefore, in formulating this taxonomy, Farahany gives special concern to how such constitutional rights apply to individuals' *mental privacy*. Farahany argues that mental privacy should be protected under intellectual property laws, such as copyright. Just as a writer owns the information contained in her novel, what originates in the brain and mind is one's *informational property*: […] even if thoughts are kept sacrosanct in the brain, they are intangible effects secured to individuals by the Fourth Amendments (see [38], p. 1261).

Crucially, this claim to own one's thoughts starts through an act of "fixation," defined as "preserving something, even if only temporarily, in a tangible medium of expression" (see [38], p. 1296). Consequently, Farahany rejects our contention (iii) above by claiming instead that *all* informational evidence, including speech, is physical. If all verbal communication is fixed in a physical form, maintains Farahany, so too are thoughts fixed through being correlated with neurological changes in the brain.

What matters for informational property, thus, is not only the protection against *seclusion* but also one's *secrecy* interest—one's control over disclosing the *content* of one's property. The difference between seclusion and secrecy here is between preventing the state from unwarranted, physical invasions upon one's possessions such as searching one's cerebral cortex, and the *information* that is retrieved

from a seized brain scan. For instance, for an unpublished novel copyright follows the author's seclusion interest in excluding alien seizure of the physical manuscript, but also secrecy by securing against unconsented publication or acknowledgement.

As a result, Farahany's account includes two senses of 'authorship' in which one becomes the author of one's expression. The first covers *authoring* or *originating* an expression, starting by producing and *fixing* that expression in a physical medium. She observes, however, that this fixation need not be *deliberate*. It suffices that it is created by the author by mere "fortuity". The second sense is *authorizing* the disclosure of one's expression; the right to, in Moran's formulation, "delegate or transfer something of one's own outside of oneself, to another person or agency" (see [29], p. 10).

Conceiving this in line with copyright, Farahany must assume that the first sense always involves the second; that *origination* implies *authorization* and that both together constitute *ownership* of one's expressions or personal information. Thus, in her account of informational ownership, for a person *S* exerting the control over creating some piece of information *p*:

A. *S* originates *p*, and

B. *S* owns *p*, and therefore

C. *S* has the exclusive control over disclosing *p*[10].

In short, Farahany's stance makes focal the idea that originating evidence entails its ownership. This, nevertheless, does not require an *intentional* act of creation, but only a certain level of control in the production of the information in a physical medium, for example remembering, imagining, or uttering something either in one's mind or in a public manifestation.

We now expose three shortcomings of Farahany's approach. First, if intention is not in fact necessary in controlling the creation of evidence; then, second, mere thoughts can be taken as speech acts; and this, third, excludes one from disclaiming one's thoughts.

3.2. Control without Intention

The first shortcoming of Farahany's account is her silence about why controlling the production of evidence does not require the intention to do so. Simple logic would suggest that a criterion of whether someone is attempting to control something is not success in controlling it, but the intention to control it. That said, there are cases where one utters something without an intention or any control, like a slip of the tongue. If so, what then is the difference between the information conveyed by one's autonomic blinking and Isabel inadvertently saying (7) out of the blue?

(7) I'm not well!

Farahany must presume that while one has no control over identifying information, such as a blood test, one does, *in principle*, control what arises in one's mind—whether in words or images, contrary to the example mooted in (7). However, this raises the question: if 'control' does not require the 'intention

[10] Notice that in the copyright framework, the derivation from A to C is not unconditional. Under normal circumstances, ownership is established at B, and authorization seems to follow from that: if you are to own something as you do when you have the copyright on something, then you must not only have created it, but you must also have the authority to disclose it. Evidently, these can come apart. First, I can sell the copyright, in which case while I created it, someone else has authority to disclose it. Second, likewise I can inherit authority to disclose it (e.g. my father's novel) without having created it. Here it is the distinctive mental content context that holds the two senses (ownership and authorization) of authorship together as 'inheriting' content decoded from someone else's brain seems inconceivable.

to create information,' then what meanings can 'intention' have in this context? Perhaps, one can intend to create it *as* evidence, to bestow upon (7) an intended *meaning*.

We take Farahany therefore to accept that one owns an utterance if one has, in principle, control over the production of the utterance in its physical medium, even if the production is 'stored' in the brain, irrespective of any attendant intention to communicate anything by it. This reading makes sense since Farahany rejects (iii), but it also entails the rejection of (iv). That is, in Farahany's sense an utterance need have no attendant intention to convey any meaning for its author to claim ownership of it as potential evidence, with all the protections that might entail.

This rejection of (iv) also aligns with Farahany's claim that one potential intrusion on one's mental privacy is via *evoked utterances.* These arise, for instance, silently "when no audible communication is made […] causing the suspect to consciously but silently respond" (see [39], p. 390). Evoked utterances include distinct physiological changes in the brain that, like those detected with NSD, "can be decoded into contemplated words, thoughts, recalled memories, and visual imagery" (see [39], p. 390).

As we read her, Farahany's explanation for an intrusion here is this: If Isabel is the author of an evoked utterance in her brain (*i.e.* she has the fullest control over the production of the mental information that is inferable from its neurophysiological form) then she has the right to claim ownership of that evoked utterance, *regardless* of an intended meaning she may have had for it or its lack of one. That being so, the police would violate the Fourth Amendment when they evoke an utterance from Isabel and in seizing it would thereby intrude on her mental privacy; while they also contravene her Fifth Amendment rights by circumventing any intention she may have *not* to speak.

3.3. Unuttered speech acts

Farahany's mention of 'evoked utterances' exhibits a second shortcoming. The mere presupposition that an utterance, so called, need never be communicated leads to the Fourth and Fifth Amendments protecting an absurdity, namely 'unuttered utterances'.

This presupposition conflates unintentional behavior with intentional action. That is, it fails to distinguish the thoughts that stream through Isabel's mind with those of Isabel's thoughts that she intends to utter. Farahany's notion of 'utterance' commits her to supposing that sentences we formulate in our heads *are already instances of a speech act* and thus to affirming that they have the same status and protection as voluntary expressions. So, on Farahany's view, even where there is neither a planned nor an actual uttering of the sentence in question, that sentence once originated is *already* a speech act. This leads to the absurd conclusion that one could be taken to have avowed an intention to murder simply by having "ruminated" upon the thought of that murder.

What is entirely overlooked in Farahany's account, evidently, is that the same speech *content* (*i.e.* the sentences) can feature in various speech *acts*, each with different meanings and force, in different contexts, even when uttered by the same subject. Thus, the mental content of a thought is of itself neither an utterance, nor an act, nor a speech act.

3.4. Precluding the disclaiming of one's thought

The final shortcoming stems from the observation in the first part above that people do not necessarily endorse a thought simply by entertaining it. If people can, therefore, dissociate themselves from what

they think, they can disclaim a thought. But if Farahany presupposes that Isabel owns the intercepted thought by originating it, how can she ever *repudiate* her ownership? Surely, if a person can claim their ownership of an evoked utterance that has been neither intentionally created nor communicated, then *others* on becoming aware of that evoked utterance can also attribute ownership of it to her. of one of your thoughts since, after all, you *are* its originator. Rather, you can disclaim a *propositional attitude* (e.g. believing or disbelieving) toward that content of your thought when the attitude is attributed to you. Consequently, not only does Farahany's account presuppose that people's intercepted thought qualifies as a speech act, it also gives a free license to attribute propositional attitudes such as *belief* independently of someone's actual propositional attitudes.

This problem with Farahany's account stems from assuming that *origination of evidence* implies *ownership of evidence*. But this implication does not always hold. While I might be the originator of the evoked utterance,

(8) "I killed Patrick at Waverley Station last night."

I am by no means *obliged* to endorse this evoked utterance as my own in the sense of holding an affirmative propositional attitude toward it, though I could. Just as I can prescind from endorsing an evoked utterance, I can entertain the thought in (8) before rejecting it not only as untrue but repellent—as described above when I rejected the thought that the Holocaust did not happen. Conversely, to endorse or associate myself with (8) and thereby 'own' it, I must establish my accountability for (8). To do so I must have an intention to *manifest* the thought, to act so as to do so, and thereby to *make of* it an utterance, properly speaking. In other words, originating a thought does not imply that I have produced it *for* expression, nor therefore would I be justified in claiming it as mine just because I produced it. To assume accountability for it, then, I must imbue (8) with an *intended meaning*. Moreover, others must understand what I mean by (8) in virtue of understanding my communicative intention. Consequently, by rejecting the idea that testimonial utterances are necessarily intentional speech acts, Farahany undermines the very possibility of becoming intelligible to each other through meaning, *i.e.* by saying who we are through the things we express and our willingness to account for having said what we have.

Farahany's view is based on the premise that an intruder can hold me responsible for evidence merely because I originated the evidence (without intending to endorse it). Following this logic of origination, Farahany extends Fourth and Fifth Amendment protections for my mental privacy and testimonial privilege, respectively. This premise can and should be rejected. If it were, it would be for the intruder to vindicate whether an intrusion has occurred or not and, therefore, whether the information can reasonably be used against me[11].

The difficulties in Farahany's proprietary account derive, we suggest, from the ISF. The fallacy arises here from assuming a symmetry between what the intruder can claim to know from the victim's evoked utterance and the victim's attitude toward what the intruder claims to know. This symmetry would presuppose a shared, *a priori* understanding of how intrusively obtained information indicates the individual's belief (or another propositional attitude) independently of the individual's intention to disclose that attitude. But any such understanding would seem unstable, since someone can dissociate

[11] Moore, who has also defended the idea that neural information should be protected under intellectual property rights, finds that to identify the violation of an individual's privacy, "the burden of proof rests on those who would cross into private domains" [40]. In other words, it is incumbent on the intruder to demonstrate that he has accessed a person's mental privacy and must, therefore, have the proof that such access gave information *of* the person's mind.

themself from what one has thought or is thinking (to include their attitude to that thought). The consequence of insisting that such an *a priori* understanding is possible would be that others could attribute decoded mental content, including propositional attitudes, to someone, no matter their protestations to the contrary. The irrelevance of their protestations that the attitude attributed is not their own effectively denies that they have control over what is going on in their own mind. It would be as if an inquisitor demanded someone take up some belief but then refused to accept that anything the subject did or said could be evidence for their having assumed the required belief.

The symmetry assumed in Farahany's account is another instance of conflating the decoded output of someone's brain with that person's speech act, which speech act necessarily reveals an epistemic attitude of the speaker. By committing the ISF Farahany's account vitiates the individual's epistemic authority. This is because it does not grant the last word to the individual, since it licenses the intruder to take evoked utterances as decisive attestations of the individual's exclusive and secret belongings. In short, the peril of 'evoked utterances' is not that others can intercept our inner speech, but rather that others' assumptions about what we think *eliminate* our testimony.

To sum up: We initially saw that the mindreading-as-testimony view does not justify a Fifth Amendment protection against self-incrimination. Now, with Farahany's revised taxonomy of evidence, we see that it does not secure Fourth Amendment protection of one's mental privacy either. This is because Farahany runs together the genesis of evidence with its ownership as property, which errs by supposing the *apparent* symmetry of information content (namely, between originating the information content and owning it in the relevant sense) by contrast with cases of *actual* asymmetry (*i.e.*, the information that I entertain a thought and the information that I endorse it). In doing so, Farahany's account precludes individuals from disclaiming ownership of an evoked utterance when others insist on attributing it to them as their own. In this way, others are, despite one's disclaimings, licensed to infer one's endorsement of an evoked utterance. Consequently, Farahany shifts the epistemic authority from the witness to the intruder in determining the meaning of the testimonial statement.

With this demonstration of an intolerable consequence of falling into the ISF, we turn to a concern with our account. By disqualifying NSD as testimonial evidence, do we deny that users of communication neuroprosthetics can be witnesses?

4. NSD is hearsay evidence.

In this part, we want to address a tension between our capacity for self-expression and the decoding of one's inner speech. So far, we have denied that NSD can produce testimonial evidence because it differs in how we assess its trustworthiness compared to individuals' testimony. But if that is right, it seems we should also deny that users of neural speech prostheses, who have no other route to communicate, can qualify as witnesses in a legal setting. Does that in turn imply that 'neuroprosthetic communication' is an oxymoron, and that users of NSD prosthetics cannot speak for themselves?

We argue that assessing the testimony of users of NSD prosthetics should be epistemically different from an unmediated verbal statement. We are not denying that NSD can be life-changing, especially (as stated in the first part) for people with speech paralysis. But we stress that since neuroprosthetic communication is *mediated* speech, it is difficult to see whether users do speak *through* the NSD prosthesis, rather than the prosthesis "speaking" *on* their behalf. This is not simply a verbal difference but a matter of what grounds the distinction between testimony and hearsay.

We shall therefore show how NSD prosthetics relate to the distinction between testimonial and hearsay evidence by outlining (1) the distinction between speaking through and on one's behalf; (2) the question of epistemic authority with NSD prosthetics; and (3) the analogy between neuroprosthetic communication and hearsay. This will lead finally to a legal characterization of NSD evidence.

4.1. Speaking through and on one's behalf

A central observation in the first part was that the evidence for a speaker's commitment to a thought is her verbal assertion of that thought. Here, 'evidence' should be understood in two ways. First, the meaning of stating something entails that the speaker endorses a belief. Second, in the act of articulating such a statement, the speaker, and not someone else, is committed to that belief.

From this observation, we see that one's testimony could not be mediated by a second party. As pointed out against Farahany's view, the police cannot take intercepted thought as Isabel's testimony because they did not acquire it *through* Isabel's speech act. And because of this, they cannot make Isabel accountable for the meaning or truth of the intercepted thought. Therefore, if others are excluded from testifying on one's behalf, it follows that to testify requires the ability to speak for oneself.

This ability includes first an ability to produce an *outcome* through one's testimony, namely by inducing a belief in an audience. As Fricker puts it: "When testimony is trustingly accepted by an individual, she acquires beliefs *through* it" [41]. Consider here Isabel declaring at a trial:

(9) I saw Pablo preparing the bomb.

If the jury takes Isabel's word for (9), then they would have gained a belief about the bomb attack planning through her testimony. Here, the central reason for trusting Isabel's testimony is because she can speak for herself. Thus, with a presumption of Isabel's sincerity, her ability to speak for herself implies that she can formulate what she believes.

This presupposes, secondly, that the *means* to speak for oneself must be unmediated. If others can learn something through Isabel's testimony, it is by presuming that she speaks 'through' her own voice. That is, Isabel makes no intermediating interpretation when converting what she intends to say in her head into vocalized words using her vocal cords[12]. For that same reason, we say that we speak through the telephone rather than that the phone speaks on our behalf[13].

However, if in our absence *others* testify on our behalf, we have what is called *hearsay*. In law, hearsay evidence is generally inadmissible because the epistemic authority of the statement is unclear [42]. Suppose here someone else, say Sergio, reports on Isabel's behalf:

---

[12] An objection following from this observation is that speech coming from the mouth of someone is not *ipso facto* unmediated as we suppose. Consider the situation where someone testifies in a language in which their competence is limited, e.g. a second language. Sometimes the words of a second language emerge from the mouth of the witness, but we may treat them as mediated nonetheless because we know that they are intermediated by a self-conscious process of translation from native language to second language. If I think someone is translating from French to English in their head, so to say, then I might be on guard for false friends like 'sensible' which mean different things in the two languages. This may be relevant in my consideration of the person's testimony. This objection, however, relies on a misunderstanding of the ordinary sense of language processing. Following Wittgenstein, we do not have an inner, Augustinian, personal language from which we translate when we speak a language in which we are fully competent (by which we understand a competence that mostly operates unselfconsciously).

[13] This last point can perhaps be made clearer if we observe that a person in the *act* of speaking through a telephone, or a microphone could not immediately become a listener of her own voice. Contrast this with a client whose lawyer speaks on her behalf. The client could witness her own statement while it is conveyed by her lawyer; the client's statement we would say is being made with *another's* voice.

(10) Isabel said, "I saw Pablo preparing the bomb."

In this scenario, Sergio, who is unrelated to the crime in question, presents the truth of his testimony as being founded upon Isabel's eyewitness account. However, it seems unacceptable to treat (10) as testimony as we lack evidence of the *witness's* commitment to the assertion attributed to her. Therefore, we could not rely on Isabel's sincerity by just taking Sergio's word for it. Since Sergio is not speaking for himself but for Isabel, we cannot use Isabel's trustworthiness to support Sergio's assertion[14].

We contend that NSD prosthetics 'speak' on behalf of people rather than people speaking through it. The distinction turns on the gap between the user's communicative intention and what the neuroprosthesis infers. To show this, we base our argument on Maslen's and Rainey's discussions of the issue of executory control in neuroprosthetic communication.

4.2. Neuroprosthetic testimonies

Maslen and Rainey remark that the gap between the user's speech intention and the decoded output grows as the NSD system becomes more sophisticated in predicting the user's brain signals. As a result, the software may 'smooth out' one's intended speech by producing an utterance "functionally equivalent to a speaker's intentions, though not necessarily exactly what they wanted to say" (see [43], p. 9). Such modifications may go unnoticed in daily communication, but in a trial, we will want to know who has the epistemic authority over the NSD 'statement'. This comes down to ascertaining whether we can justifiably treat the NSD output as testimonial evidence.

However, this gap between the user's communicative intention and the decoded speech output is already a sound basis for maintaining that NSD speaks *on behalf of* the user of the NSD rather than the user speaking *through* it. If that is right, an audience will not obtain a belief *through* the user's communicative intention but rather through the NSD *on behalf of* the user. Indeed, Rainey *et al.* find that NSD prosthetics may *replace* rather than restore the user's speech [44].

We maintain therefore that since the evidence of the user's commitment to the decoded statement is not made evident in the NSD output, the NSD output appears to be a second-hand or derivative account of the person's inner speech, making it like hearsay rather than testimony.

We take this characterization from the fact that in law hearsay includes *someone* as well as *something*, for example documents, speaking on one's behalf[15]. This idea also has support in practice: Chandler *et al.* describe a case where the complainant, having cerebral palsy, gave testimony through her speech synthesizer. The testimony, however, was deemed hearsay and was therefore annulled [30].

4.3. When the machine speaks on our behalf

Now we see why neuroprosthetic communication should be legally characterized as hearsay evidence. If it is not accepted that NSD prosthetics allow users to speak for themselves, strictly speaking, then we should not treat the NSD output as testimony. Just as with Sergio's testimony about Isabel, we do not have evidence for the user's commitment to the speech attributed to her, *i.e.* the condition for assessing the user's sincerity is unavailable.

[14] Here, it is notable that Sergio is accountable for his saying that Isabel said what she did, but not accountable for the content of what Isabel said—if Isabel said it but was lying, Isabel, not Sergio, is accountable.

[15] Tapper 2004, p. 62: "The contents of a document may be incorporated in the evidence of a witness who testifies to having entered into a written contract, and the court may be referred to them because they contain admissible hearsay statements" [45].

Recall here Roskies' remark above about Tang *et al.*'s findings: "'I just jumped out [of the car]' was decoded as 'I had to push her out of the car'" The "enormous difference" in law Roskies mentions is obviously between a deponent stating something otherwise irreproachable and confessing to a potential crime. But it also raises the question of whom to blame if the decoded output proves inaccurate and the user has no other means to communicate?

Consider lying in hearsay: If it were proven that Sergio lied in avowing (10) on Isabel's behalf, we could not thereby blame Isabel for Sergio's lie. Likewise, if we found the NSD reporting a factually false statement, we cannot discover *through further information mediated by the NSD* whether the blame for the falsehood lies with an inaccuracy in the neural speech decoder or the user's dishonesty. In unmediated statements, if there is lying there is no mystery about who is lying. By contrast, mediated speech like NSD produces uncertainty about the source of the falsehood. With this uncertainty, no appeal can be made to the epistemic authority of the user regarding the accuracy of.

Correctly characterizing NSD as hearsay evidence does not, however, exclude NSD from being a means of communication for people physically unable to articulate speech. Indeed, questions of testimonial sincerity and decoding accuracy are mostly of legal concern rather than being an obstacle in ordinary daily life[16]. Nonetheless, the discussion does highlight two important facts. First, no matter how accurate NSD becomes, it does not enable an evaluation of someone's sincerity. Sincerity and testimony do not wait on technical advancements in neuroscience. Instead, testifying concerns recognizing someone's authority over their speech and the expectation of truthfulness consequent on the recognition of that authority. In that sense testifying is essentially a *social act* because it depends on *other people*, including legal and social institutions, to recognize that authority we have over the meaning and truth of our own words[17]. Indeed, a further shortcoming of hearsay is that it cannot be cross-examined, since strictly speaking, it was never someone's testimony in the witness box[18].

Second, advances in neurotechnology may well enable speech-impaired people to communicate, but it does not follow from the restoration of that ability that it can also restore their *freedom* of speech (*i.e.* the ability to speak for themselves), neither in the negative sense (freedom from testifying against oneself), nor in a positive sense (freedom to offer one's word as evidence in court). While we have argued that NSD should be characterized as hearsay evidence, that does not prevent it from being potentially useful in court[19]. In cases where the NSD prosthesis is the only option to communicate, someone could be offered the opportunity to endorse what the NSD prosthesis produces. This would be

[16] To be sure, the evaluation of neuroprosthetic speech is not the same as considering a person's attestation transferred through an interpreter in a trial. If the defendant only speaks Lingala, and those on the jury only speak Swahili, then the problem is not the deficiency of *speech* but of mutual *understanding*. In that case, the interpreter's role is provisory since the jury could come to learn Lingala and thereby evaluate the defendant's sincerity. The jury could record the defendant's testimony in Lingala, and, subsequently, come to understand what the defendant intended to say and whether the interpreter had rendered an incorrect translation. Regarding neuroprosthetic speech, however, we presume that the user *cannot* communicate and can therefore only make herself known by relying on an inference-based interpretation of her inner speech. Thus, *there is no other way* to rely on the user's words than by what is provided by the neuroprosthetic speech outcome.

[17] Moran explains that if we commonly mistrust a person's words, we may still grant her the ability to use and understand language properly while not crediting her "with the authority to *tell* anyone anything at all" (see [29], p. 8). It is salient for small children, for example, that we do not grant them the ability to perform certain speech acts, such as promising or declaring an intention, and for certain adults to cast a vote or hire someone. Idem.

[18] One reason for the inadmissibility of hearsay is that it cannot be cross-examined, as argued in *Bruton v. United States* [46]. Thanks to an anonymous reviewer for this legal detail.

[19] As Ho [42] reasons, in a case where the jury has no testimonial evidence to consider, it is reasonable to rely on hearsay rather than *no* evidence.

an exceptional circumstance and function as a last resort. Where unmediated testimony is available, the role of the NSD outcome should be confined to hearsay.

5. Conclusion

Utterances are the guardrails of unbridled thought[20]. We can suppress our words but not our thinking. Children learn how to speak and what is appropriate to say when by first learning how to filter the stream of conscious thought for the thoughts they will give concrete expression.

One value of our article lies in highlighting the fundamental philosophical, as well as legal, difference between thought and thought's expression. As argued, we are usually responsible only for what we can control. Hence, we are responsible for our utterances but not for what we think. We may have wicked thoughts and might feel shame for having them, but neither makes us morally, much less legally, accountable. The same difference visible in responsibility is also at work in the difference between an utterance, which is necessarily a speech act, and a thought, which is not of itself a mental or practical act.

An act necessarily has an actor, or author, and is therefore *attributable* to that person. Accordingly, an utterance is attributable to the utterer. Are thoughts in the same sense attributable to a thinker?

A premise in our discussion is that no matter the thought a person has, she can always dissociate herself from it. The person may entertain the most horrible opinions and carefully formulate them in her head, but as long as she does not make these thoughts articulate, there is no basis, *a fortiori* no evidence, to attribute them to her as her own.

In fact, most of our thoughts do not occur to us as *our* thinking. If it were possible to transcribe one day's worth of a person's rumination, the subject may be very limited in her ability to identify which thoughts were "her" thoughts or even to locate any personal aspect of them. Someone's stream of thought rolls along continually without a label indicating *who* is thinking. It is an exception—*the* exception from which philosophy begins—when someone stops to attend to the process of thinking and its results. We might exclaim to ourselves, when entertaining a shocking idea, "Did *I* just think that?" In consequence, ascribing a thought to a subject, e.g. by saying "*I* have got an idea!", is an *interpretation* after the fact of a course of thinking.

So, instead of being convinced, as Descartes was, that beneath every thought lay a tacit 'I think', we should rather say, as Nietzsche says, '*it* thinks': "… a thought comes when "it" wants to and not when "I" wish, so that it's a falsification of the facts to say that the subject "I" is the condition of the predicate "think." It thinks: but that this "it" is precisely that old, celebrated "I" is, to put it mildly, only an assumption, an assertion, in no way an "immediate certainty" [47].

Lastly, the legitimacy of necessarily ascribing a personal aspect to speech, unlike thinking, is that speech is necessarily *manifested* to others. While we may think *about* someone, speech is addressed to an audience, and the particular intended meaning is theirs to grasp. The audience recognizes this meaning through the speaker's utterance and by accepting the speaker's authority holds them personally responsible for the meaning and truth of their utterance. Hence, testimony is not a something, the possession of which might occur in various ways. Testimony is never stolen, it must be given. Testifying is not a solitary act, like thinking out loud or eating. Rather, a part of the meaning of a testimonial act is

[20] We thank G. Ken Holman for this quote (personal communication).

not only the *information* intended to be conveyed but also the *manifestation* itself; what the speaker intends to *communicate* about this information[21].

If NSD cannot be characterized as testimony, it is because we do not have a basis for attributing to the speaker some decoded interval of inner speech. If a forced entry to someone's mind were possible, perhaps into someone with locked-in syndrome, what is retrieved is not *someone's* thought or its expression but *a* thought or *an* expression. To attribute *ownership* of the thought is to add a further claim about *what* the thought means and that, as Nietzsche said, is not an "immediate certainty."

This point brings a valuable lesson for future studies in neuroethics, notably freedom of thought and mental privacy. To protect those rights in the context of neurotechnology, the UN rapporteur suggested distinguishing between *forum internum* and *forum externum* where the former comprehends "a person's inner sanctum (mind) where mental faculties are developed, exercised and defined" and the latter "the external realm" [50]. So if copyright and more generally the whole idea of *ownership* are not fine-grained enough to protect our mental privacy, we must move past them. The central question for neuroethics is not how to identify the *bounds* of the *forum internum* and when our "inner sanctum" has been transgressed, but on what basis and in what way this or that entity in the *forum internum* can be attributed to *me*. That is to say, the violation of mental privacy does not occur through entry to the mind or the brain, but through usurping the authority to say of some mental state what our relationship to it such that we may be accountable for it.

We hope that this article will serve as one step toward a new, *epistemological* approach to protecting mental privacy. This approach is concerned less to protect sub-personal *information* in the mind and more to reinforce the authority of the *person* to decide the meaning in and truths derivable from that information.

Declaration of generative AI and AI-assisted technologies

The authors did not use generative AI or AI-assisted technologies in the writing of this manuscript.

Acknowledgments

This work is an outgrowth of Tristan B. Borresen's master's dissertation, completed in 2024 at the University of Edinburgh in the philosophy of cognitive science under the supervision of David Levy. Tristan thanks Dr. Alistair Isaac and Neil Bramley's Computational Cognitive Science Lab Group for discussions around the topic of the dissertation. A special thanks to Professor Jennifer Chandler, G. Ken Holman, Nicolas Navarre, and Gustavo Cilleruelo Calderón for their inputs in revising the manuscript. Finally, thanks to Abel Wajnerman Paz for encouraging the publication of this essay.

Authors' contribution

Conceptualization, Tristan Baujault Borresen and David Levy; methodology, Tristan Baujault Borresen; validation, David Levy; formal analysis, Tristan Baujault Borresen; investigation, Tristan Baujault Borresen; resources, Tristan Baujault Borresen; writing—original draft preparation, Tristan Baujault Borresen; writing—review and editing, Tristan Baujault Borresen and David Levy; supervision, David

[21] Scholars after Grice have called this the *communicative intention*. See this point made explicit in e.g. [48] and [49].

Levy; project administration, Tristan Baujault Borresen. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Wittgenstein L. *Philosophical Investigations*, 4th ed. Oxford: Wiley-Blackwell, 2009.

[2] Bauby JD. *The Diving Bell and the Butterfly*, 1st ed. New York: Random House, 1997. pp. 132.

[3] Tang J, LeBel A, Jain S, Huth AG. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat. Neurosci.* 2023, 26(5):858–866.

[4] Littlejohn KT, Cho CJ, Liu JR, Silva AB, Yu B, *et al.* A streaming brain-to-voice neuroprosthesis to restore naturalistic communication. *Nat. Neurosci.* 2025, 28(4):902–912.

[5] Lavazza A, Giorgi R. Philosophical foundation of the right to mental integrity in the age of neurotechnologies. *Neuroethics* 2023, 16(1):10.

[6] Ligthart S, Bublitz C, Douglas T, Forsberg L, Meynen G. Rethinking the right to freedom of thought: a multidisciplinary analysis. *Hum. Rights Law Rev.* 2022, 22(4):ngac028.

[7] Ligthart S. Towards a human right to psychological continuity? Reflections on the rights to personal identity, self-determination, and personal integrity. *Eur. Conv. Hum. Rights Law Rev.* 2024, 5(2):199–229.

[8] Farahany NA. *The Battle for Your Brain: Defending the Right to Think Freely in the Age of Neurotechnology*, 1st ed. New York: St. Martin's Press, 2023.

[9] Haynes JD. Brain reading: decoding mental states from brain activity in humans. In *I Know What You're Thinking: Brain Imaging and Mental Privacy*, Oxford: Oxford University Press, 2011. p. 29.

[10] Gilbert F, Russo I. Mind-reading in AI and neurotechnology: evaluating claims, hype, and ethical implications for neurorights. *AI Ethics* 2024, 4(3):855–872.

[11] Pardo MS. Self-incrimination and the epistemology of testimony. *Cardozo Law Rev.* 2008, 30(3):1023–1046.

[12] Pardo MS. Testimony. *Tulane Law Rev.* 2007, 82(1):986845.

[13] Lackey J. Testimonial knowledge and transmission. *Philos. Q.* 1999, 49(197):471–490.

[14] Lackey J. *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press, 2008.

[15] Veas JE. Brain-reading technologies and the right against self-incrimination: a challenge for the distinction between testimonial and real evidence. *Ger. Law J.* 2025, 26(4):683–701.

[16] Brajt CR. Abridging the fifth amendment: compelled decryption, passwords, & biometrics. *Fordham Intell. Prop. Media Ent. Law J.* 2021, 31(4):1154–1198.

[17] Veas JE. Smartphone decryption via forced fingerprinting and the right against self-incrimination: the German federal court (BGH) addresses the problem. *Comput. Law Secur. Rev.* 2025, 59:106215.

[18] Ligthart SLTJ. Coercive neuroimaging, criminal law, and privacy: a European perspective. *J. Law Biosci.* 2019, 6(1):289–309.

[19] Lokaneeta J. *The Truth Machines: Policing, Violence and Scientific Interrogations in India*. Ann Arbor: University of Michigan Press, 2020.
[20] Nair S, Weisse K, Chandler J, Bansal P, Chandler JA. Neurotechnology on trial: a systematic review of brain fingerprinting in Indian criminal cases from 2002 to 2023. In *Proceedings of Neuroethics 2025 (Annual Meeting of the International Neuroethics Society)*, Munich, Germany, April 23–25, 2025.
[21] Chandler JA. Knowing the mind from brain data: the challenge of prediction and the fairness of relying on objective data about the mind. *Hast. Cent. Rep*. 2025, 55(4):15–23.
[22] Andriolo RJ. Rogers v. Richmond, 365 U.S. 534 (1961). *Mont. Law Rev.* 1962, 23(2):7.
[23] U.S. Supreme Court. Schmerber v. California, 384 U.S. 757 (1966). 1966. Available: https://supreme.justia.com/cases/federal/us/384/757/ (accessed on 21 May 2024).
[24] Murphy ER, Greely HT. What will be the limits of neuroscience-based mindreading in the law? In *Oxford Handbook of Neuroethics*. Oxford: Oxford University Press, 2011. pp. 635–654.
[25] Brown T, Murphy E. Through a scanner darkly: functional neuroimaging as evidence of a criminal defendant's past mental states. *Stanf. Law Rev.* 2010, 62(4):1119–1208.
[26] Council of Europe Treaty Series. Convention for the Protection of Human Rights and Fundamental Freedoms. 1960, pp. 1–16. Available: https://treaties.un.org/doc/publication/unts/volume%20213/volume-213-i-2889-english.pdf (accessed on 7 June 2024).
[27] Laudan L. The presumption of innocence: material or probatory? *Leg. Theory* 2005, 11(4):333–361.
[28] Grice H. Meaning. *Philos. Rev.* 1957, 66(3):377–388.
[29] Moran R. *The Exchange of Words: Speech, Testimony, and Intersubjectivity.* New York: Oxford University Press, 2018.
[30] Chandler JA, Van der Loos KI, Boehnke S, Beaudry JS, Buchman DZ, *et al.* Brain computer interfaces and communication disabilities: ethical, legal, and social aspects of decoding speech from the brain. *Front. Hum. Neurosci.* 2022, 16:841035.
[31] Roskies AL. Mindreading and Privacy. In *The Cognitive Neurosciences*, 5th ed. Cambridge: The MIT Press, 2014.
[32] Etkin A, Wager TD. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* 2007, 164(10):1476–1488.
[33] Uncapher MR, Boyd-Meredith JT, Chow TE, Rissman J, Wagner AD. Goal-directed modulation of neural memory patterns: implications for fMRI-based memory detection. *J. Neurosci.* 2015, 35(22):8531–8545.
[34] Reardon S. Mind-reading machines are here: is it time to worry? *Nature* 2023, 617(7960):236.
[35] Klein SR. Lies, omissions, and concealment: the Golden Rule in law enforcement and the federal Criminal Code. *Tex. Tech Law Rev.* 2007, 39(4):1321–1354.
[36] Bunn GC. *The Truth Machine. A Social History of the Lie Detector*. Baltimore: Johns Hopkins University Press, 2012.
[37] Hampshire S. *Freedom of the Individual*. Princeton: Princeton University Press, 2016.
[38] Farahany NA. Searching secrets. *Univ. Pa. Law Rev.* 2012, 160(5):1239–1308.
[39] Farahany NA. Incriminating thoughts. *Stanf. Law Rev.* 2012, 64(2):351–408.
[40] Moore AD. Privacy, neuroscience, and neuro-surveillance. *Res. Publica* 2017, 23(2):159–177.

[41] Fricker E. Testimony: knowing through being told. In *Handbook of epistemology*. Dordrecht: Springer, 2004. pp. 109–130.
[42] Ho HL. A theory of hearsay. *Oxf. J. Leg. Stud.* 1999, 19(3):403–420.
[43] Rainey S. A steadying hand: ascribing speech acts to users of predictive speech assistive technologies. *J. Law Med.* 2018, 26(1):44–53.
[44] Rainey S, Maslen H, Mégevand P, Arnal LH, Fourneret E, *et al.* Neuroprosthetic speech: the ethical significance of accuracy, control and pragmatics. *Cambridge Q. Healthcare Ethics* 2019, 28(4):657–670.
[45] C. Tapper, *Cross and Tapper on Evidence*, 10th ed. London: LexisNexis Butterworths, 2004.
[46] U.S. Supreme Court. Bruton v. United States, 391 U.S. 123 (1968). 1968. Available: https://supreme.justia.com/cases/federal/us/391/123/ (accessed on 19 January 2026).
[47] Nietzsche F. *Beyond Good and Evil*, 1st ed. New York: Random House, 1966.
[48] Sperber D, Wilson D. *La pertinence: communication et cognition*. Paris: Les Editions de Minuit, 1989.
[49] Recanati F. On defining communicative intentions. *Mind Lang.* 1986, 1(3):213–242.
[50] Shaheed A. Interim report of the special rapporteur on freedom of religion or belief, report on the freedom of thought. 2021. Available: https://www.ohchr.org/en/documents/thematic-reports/a76380-interim-report-special-rapporteur-freedom-religion-or-belief (accessed on 27 January 2026).