Free Will and Necker's Cube: Reason, Language and Top-Down Control in cognitive neuroscience

GRANT GILLETT AND SAM C. LIU

Abstract
The debates about human free will are traditionally the concern of metaphysics but neuroscientists have recently entered the field arguing that acts of the will are determined by brain events themselves causal products of other events. We examine that claim through the example of free or voluntary switch of perception in relation to the Necker cube. When I am asked to see the cube in one way, I decide whether I will follow the command (or do as I am asked) using skills that reason and language give to me and change my brain states accordingly. The voluntary shift of perspective in seeing the Necker cube this way or that exemplifies the top-down control exercised by a human being on the basis of the role of language and meaning in their activity. It also indicates the lived story that is at the centre of each human consciousness. In the third part of this essay, three arguments are used to undermine metaphysical objections to the very idea of top-down self control.

1. Introduction

Debates about human free will have traditionally been the domain of the metaphysician but recently neuroscientists have entered the field with experimental findings about the brain activity accompanying human voluntary action. In fact neuroscientists have a longstanding interest in the topic: Hughlings Jackson, the father of English neurology, remarked of the human brain, 'if the highest centres were perfectly automatic, there would be no such thing as a 'voluntary' operation, all being organised'.[1] He realised that human activity

This paper resulted from research supported by Francoise Baylis, Canada research Chair in Neuroethics at the University of Halifax. This work is also supported by the Canadian Institutes of Health Research grant NNF 80045 States of Mind: Emerging Issues in Neuroethics

[1] J. Hughlings Jackson, 'The Croonian lectures on evolution and dissolution of the nervous system', *The British Medical Journal* (1884), 705.

doi:10.1017/S003181911100057X © The Royal Institute of Philosophy, 2012

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

had the form it did because the arrangement of the brain was designed to over-ride automatic and (relatively) mechanistic responses such that 'the progress of evolution is from the most to the least automatic… those which are concerned during one's present thinking and doings'.[2] Hughlings Jackson was aware that stereotypy of the type found in mechanistic explanations was only found in the lower reaches of the nervous system and that speech had a crucial role in the adaptation of human beings to their complex and cooperatively constructed world.[3] In this respect he prefigures Wegner, a contemporary critic of free will who remarks that what he means by voluntary is that which a person 'can do when asked'(Wegner, 2004; Dennett, 2007). In more technical terms the myriad cues and connections and the rule-following skills that underpin meaning and understanding disrupt both predictability and a purely functional explanation of higher order cognitive processes and the control of action (in terms of physical conditions or extension rather than sense). Intentional behaviour requires, as Dennett argues, a different order of explanation, naturalistic but attuned to the context of adaptation of the whole organism – or human ethology.[4] We will argue that Wegner is more right in relating voluntary behaviour to 'what a person can do when asked.'

1.1 Human free will and brain events

Determinist accounts of human psychology argue that free will is an illusion reflecting neural processing to do with proprioception and the explanation of our bodily activity.[5,6] All our cognitive functions (and in particular those that bring about our actions) are, according to the theory, the result of 'mechanisms – the rattling off of genetic, unconscious, neural, cognitive emotional social and yet other chains that lead … to the things people do'. The view seems to be supported by the observation that unconscious neural events

[2] Ibid., 704–705.

[3] Franz, E., and Gillett, G., Hughlings Jackson's evolutionary neurology: a unifying frameworkfor cognitive neuroscience *Brain*, 2011, **134**(10):3114–3120.

[4] Rom Harré and Grant Gillett, *The Discursive Mind* (SAGE Publications, 1994).

[5] Daniel M. Wegner, *The Illusion of Conscious Will* (Cambridge (Mass.): MIT Press, 2002).

[6] Guy Claxton, 'Whodunnit? Unpicking the "Seems" of Free Will', *Journal of Cousciousness Studies* **6** (1999), 8–9, 99–113.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

precede the formation of conscious intentions to act.[7,8] The thesis that events in our brains generate our activity dictating what we do independently of our conscious decisions currently has widespread currency and rests, metaphysically on the bottom up determination of mental events by brain events.[9,10]

To defend a robust thesis that we do in fact exercise conscious self control and therefore have free will is to claim that we, as human beings holistically construed in terms of our conscious interactions with others, have control over our brain function in meaningful ways not fully explained by what goes on in the brain.

We could defend that thesis using a Kantian conception of transcendental free will: our free will indicates a source of acts and inner determinations not arising from the causal transactions governing physical events.[11] Kant suggested that the origin of an action cannot be understood if we focus on the brain alone because we need to appeal to reasons and their reason-giving force to explain what is done by creatures able to heed those reasons and determine whether or not to act on them. In that sense, free will would make us different from mere animals because we respond to prescriptive norms (imperatives or conceptions of laws) and must therefore explain what we do in terms of 'intelligible causes' in such a way 'that we cannot get beyond them' and must think of an agent,[12] as possessing 'a faculty through which an empirical series of events first begins'.[13] Kant argues that we see efficacy of reason in the imperatives which in matters of conduct we 'impose as rules on our

[7] Wegner, D. & Wheatley, T. Apparent mental causation *American Psychologist* 1999, 480–492.

[8] Benjamin Libet, 'Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action', *The Behavioral and Brain Science* **8** (1985), 529–566, note 5.

[9] Jaegwon Kim, *Essays in the Metaphysics of Mind* (Oxford University Press, 2010).

[10] Gillett, G. 2011 review of Kim's metaphysics of mind, *Philosophy* **87**, no. 339 (this issue).

[11] KantI.: (1789 [1929]) *The Critique of pure reason* (tr. N. Kemp Smith) London: Macmillan (references will be given by pagination in this edition, here: B574).

[12] (B585).

[13] (B581). In fact Locke and Strawson both argue that this is our basic 'model' for causation (Locke, J. (1689, [1975]) *An Essay concerning Human Understanding* (ed. P. Nidditch) Oxford: Clarendon. Strawson, P. (1985) Causation and explanation in Vermazen, B. & Hintikka, M. *Essays on Davidson: actions and events*. Oxford; Clarendon, 115–136.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

active powers'[14] (presaging the remark by Wegner and Dennett) and identifying a promising starting point to investigate the actions of the will, the role of speech in human thought, and brain function. Kant argues 'This active being would to this extent be independent and free of all the natural necessity present in the world of sense. Of it one would say quite correctly that it begins its effects in the world from itself without its action being in itself'.[15] We will interpret Kant's position through the following neuro-philosophical argument for human self-control and a top-down conception of freedom of the will and brain events.

1.1.1 Human actions can be explained in terms of the reasons and arguments being enacted in executing them.

1.1.2 When a person acts on the basis of being asked to do this or that the origin of the act is the conscious decision to accede to the request.

1.1.3 Conscious human decisions, whatever the **causal workings** of the brain at the time, sometimes configure brain events from top down to produce an action.

1.2 Clarifying the position

The position is clarified by contrasting human action with animal action. Animals respond to the environment according to their memories, their species specific tendencies, and their learning history. They act as a result of a confluence of those influences (as outlined by Wegner above) and neither enter into arguments about their reasons for action nor heed situation specific imperatives conveying rational or rhetorical considerations. They respond in ways, in principle predictable from their experiences and their dispositions (innate or learned). They do not direct or redirect their thinking in terms of the perspective of others who see the world differently and offer alternative framing conceptions. Animals cannot, for instance, devise or make use of metaphors in the way they react to situations. If their experiential history and the information from the current context conditions a certain orientation or response, animal behaviour is structured in accordance with that. Human beings are otherwise so that, as Hughlings Jackson notes, they use the power of words 'for thinking on novel and complex subjects, for ordering

[14] (B575).
[15] (B569).

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

images in new and complex relations'[16] such that consciousness and propositionalizing operate in the control of behaviour. Luria, the neuro-psychologist, remarks that the distinctively human organization of behaviour occurs 'with the close participation of speech'.[17]Human beings can view things from different perspectives and freely activate different viewpoints with the aid of speech to change their ways of thinking. Speech yields options for 'the organization of conscious activity' and 'higher forms of cognitive activity', and 'the most complex forms of human behaviour'.[18] Since humans are also animals (with a physical body, instincts and habitual ways of reacting), they may not use their voluntary powers to direct their cognition[19] but rather follow the confluence of natural laws. The issue then becomes whether speech and imperatives as we understand them do have the effectiveness that is posited.

Determinists think that since natural laws can explain most of human behaviour, they must explain what happens in the brain and causal paths are, in principle, traceable from the patterns of activity controlling action back to the internal and external stimuli controlling and directing them. But if those explanatory connections are not plausibly of the form required, that argument would fail (except as a dogma). We will argue that, focusing on higher mental functions, we must face the fact that human beings exert top-down control over their brain function evident in the ability to change perspectives on the Necker cube if asked to do so. We argue that human free will attunes voluntary activity, through learned skills of speech and language use. We will then offer a defusing argument against metaphysically basic brain mechanisms underpinning determinism and for a conception of top-down causation that makes free will (neuro-philosophically) plausible.

2. The Necker cube and free will

2.1 Bottom and top

In any given situation or set of stimulus conditions, all the information I receive is stored in my brain through biological processes

[16] J. Hughlings Jackson, 'On Affection of Speech from Disease of the Brain', *Brain* **1** (1878), 3, 323, 312.

[17] Luria, A.R. (1973). *The Working Brain*. (Harmondsworth: Penguin), 93.

[18] Luria, 1973, 79, 84, 94.

[19] John McCrone, 'A Bifold Model of Freewill', *Journal of Consciousness Studies* **6** (1999), 8–9, 241–259.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

involving Long Term Potentiation LTP and patterns of hippocampus based integration.[20] Without that happening, we cannot store memories and well developed cognitive schemata to enact in situations both old and new and the very idea of effective free will or conscious control of action is just a pipe dream (or Cartesian fantasy).

Our representations of situations can be informed by two main sources; passive or habitual sources and active sources that we search out or construct and then use in new ways. Passive sources can be further divided into three categories:

2.1.1 Randomness, unconscious initiated ideas and other associative responses,
2.1.2 Internal causal chains such as hormones for sexual desire, hunger, and sleepiness.
2.1.3 Environment stimuli, as when we are distracted from consciously directed voluntary activity byincidental stimuli.

Active sources arise from self directed information gathering and contribute to a reason-governed choice of cognitive frame or strategy. What is more, two specific bases: 1) communication from others; and 2) a decision to try and re-frame a problem can be used to direct, develop, and relocate our ideas and the way we set ourselves to act. We will argue that any information passively received potentially provokes an active possibility of responding on the basis of reason (or as the result of conversation – manifest if we do something when asked). The Necker cube shows the will – voluntary self-control on the basis of a conscious reason to act – directing brain processing.

2.2 Necker cube

Consider the Necker cube and the two ways of seeing it (see Figure 1).[21] I may see it in some preferred or automatic way or, if someone directs me to see it in a certain way (with either the top or the bottom corner coming out of the page), I can switch my point of view to comply with the request. Although I can only have one view point (one internal representation) at a time, most people can voluntarily switch viewpoints once they have mastered the skill or technique of doing so.[22] Each viewpoint manifests a holistic excitation

[20] James J. Knierim, 'Hippocampus and Memory: Can We Have Our Place and Fear It Too?', *Neuron* **37** (2003), 3, 372–374.
[21] Ludwig Wittgenstein, *Philosophical Investigation* (Trans. 4th ed. Blackwell Publishing Ltd, 2009), part 2.
[22] Wittgenstein, op. cit. note 20, part 2:xi.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

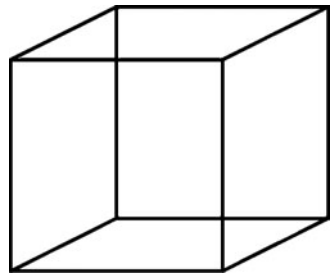


Figure 1. The Necker

state of my perceptual networks. Therefore I can voluntary control my brain states depending on the skills I have learnt to master. Notice that the direction of causation is exactly as Kant claims it is, from the imperative (freely obeyed, nobody makes me do it and nothing compels me to see the cube one way rather than another) to the brain state. This is top down causation in action in that a decision – 'I will do what I am asked to do' – determines the state of my brain, and therefore the content of my experience. It follows that we can freely, or voluntarily change our brain states and the way we see the cube (and other figures).

We can therefore construct a *Necker cube syllogism*, as follows.

2.2.1 The ability to voluntarily do this or that when asked is a clear example of self control or free will (voluntarily obeying an imperative).
2.2.2 My perception of a stimulus array is based on a specific brain state.
2.2.3 I can learn to see the Necker cube this way and then that when asked to.
2.2.4 I can voluntarily change my brain state on request.

2.3 Changes of attention

To be able to see the Necker cube in one way then another, a person must change his or her brain state. Given that consciousness, the control of behaviour, and attention are central topics in cognitive neuroscience and concern the allocation of cerebral resources,[23]

[23] Alan Allport, 'Attention and Control: Have We Been Asking the Wrong Questions? A Critical Review of Twenty-Five Years', David E. Meyer and Sylvan Kornblum (eds), *Attention and Performance XIV:*

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

focusing on attention (or aspect seeing) is a means of examining our role as voluntary agents – creatures who can do things when asked. This ability to construct and use internal representations (a specific excitation pattern or active neuronal associative assembly) at will is clearly a way of controlling neuro-cognitive function[24,25] such that we can change very rapidly between different excitation patterns (internal representations or vectors in an associative network) depending on interest and task orientation. A shift of attention, voluntary or elicited, by changing that activation pattern puts a scene in a different light and is part of many interesting perceptual phenomena: imagine looking at the picture of the Dalmatian dog in the shadows and being unable to see it as a dog until somebody points out what it is (Figure 2), perhaps by using certain cue statements – 'There is its head' and so on. In that case what was invisible (in one state of brain activation) becomes visible.[26]

We have a range of concepts used in constructing internal representations and we learn to re-represent situations during cognitive development so that our frames of cognitive processing can shift and expand our thinking.[27] For example, when I see the dog, I see it because I have the relevant concept. If I realised that the dog I was looking at was a wolf, I would deploy a different concept that may fit the same stimulus array. Notice again, that I must see it as one or the other and I can only have one internal representation in my mind at one time but that I can voluntarily (i.e. do so when asked) switch between states. And, when in a given representational state, I may not perceive other things that do not fit with it, for example in some dramatic attention blindness experiments, people cannot see a moon walking ape because their attention has been completely occupied by counting the number of passes of a ball.[28] When the

---

Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience (Cambridge, MA: MIT Press 1993), 183–218.

[24] J. Kevin O'Regan, Ronald A. Rensink and James J. Clark, 'Change-blindness as a result of "*mudsplashes*"', *Nature* **398** (1999), 34.

[25] Nilli Lavie, 'Distracted and Confused?: Selective Attention Under Load', *Trends in Cognitive Sciences* **9** (2005), 2, 75–82.

[26] Henry Gleitman, *Psychology* (3rd ed. New York: W.W. Norton & Co., 1994), 224.

[27] Annette Karmiloff-Smith, *Beyond Modularity: A Developmental Perspective on Cognitive Science* (Cambridge (Mass.): MIT Press, 1992).

[28] Daniel J. Simons and Christopher F. Chabris, 'Gorillas in our midst: Sustained Inattentional Blindness for Dynamic Events', *Perception* **28** (1999), 1059–1074.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press



Figure 2. Dalmation dog in the shadow

observers' attention is freed up from counting passes, they can search their visual field without being dominated by that task and see the moon-walking ape, particularly if they are told that it will enter the scene.

The same thing applies when we are thinking of something in a logical or mathematical way, we can block the internal representations we need and focus on the thing that we are trying to solve in a way that will not give us the answer. When we change our attention (a specific neuronal activation pattern) we 'click out of' the dominant brain process by an act of will (in a way that is analogous to the Necker cube case). In fact, having mastered a set of cognitive techniques, we can do these things as easily as we can raise an arm without really attending to the muscles involved.[29]

2.4 Rationality and language

We cultivate abilities to change our attention by learning to follow the rules that structure perception, selective attention, rationality and reason just as they do a game like chess.[30] I learn the rules through voluntary submission and application, and so I equip myself with abilities (to do the knight's move, to put the king in check, and so on). A computer can also 'learn' the rules, in the form of a set of

[29] Wittgenstein op cit. note 20, 612–627.

[30] Gillett, G. (1992) *Representation meaning and thought* (Oxford: Clarendon).

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

procedures governing a highly stereotyped domain and it can accumulate goal related strategies (in relation to which the intentional stance might be useful). But in taking the intentional stance, I treat the computer like an interest directed rule follower and not like a machine (as discussed by Winch[31]). But, unlike a human being engaged with human argument, imperative and all those things we do 'with the close participation of speech' terms like 'voluntary', 'heeding advice' and 'being in control of oneself' do not apply to the machine. When we impose rule-based constraints on our behaviour, we play the game of doing things when we are asked and we commit ourselves by eliminating some options and incorporating others into our behaviour so as to respond to a discursive context where 'freedom and resentment' have meaning.[32] In that context we are not causally compelled to move thus or so, but we may or may not do what is asked of us; at any stage we can 'spit the dummy' and go with a freely ranging trajectory of action/brain excitation.

It requires self control to master the rules and skills of any game but it gives me freedom to enact my reasoning and exercise my will in playing the game. That mastery allows me to direct my thinking and acting in ways that are effective in a segment of the human life world. We learn the rules and use them to regulate or organise our own creative activity as we do in the (relatively) unrestricted domain of the world at large (recall Hughlings Jackson and Luria). And skills, when mastered, may become nearly automatic so that I can deploy them smoothly without the intrusion of conscious step-by-step planning. Thus a result of learning from others is the ability to use the structure of language to evaluate realistic options and rehearse the consequences for each option so that, as Popper notes 'my hypotheses die in my place'.[33,34] In chess and life, I build my mastery through experience obtained from the interaction with others and thereby create in my brain pathways that result in effective strategies, some explicitly thought through and others, more intuitive, not following anything like conscious or pre-determined rules but rather innovative and creative. Chess

[31] Winch, P. (1958) *The Idea of a Social Science and its philosophy* (London: Routledge).

[32] P.F. Strawson, *Freedom and Resentment and Other Essays* (London: Methuen & Co Ltd, 1974).

[33] Daniel C. Dennett, *Freedom Evolves* (New York: Penguin Books, 2003).

[34] Op. cit. note 17, John McCrone.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

reflects, on a miniature and confined scale, the wilful control and shaping of brain function in the real world whereby I escape 'automatic activity' and become a voluntary actor.[35] Such learning gives a person more options in certain situations and may oppose unreflective habits or dispositions. For instance, if experience urges that I have 99% chance to be hurt if I step to help another and I am disposed not to, I can still brace myself for pain, go with the 1% chance and, for reasons that I have trained myself to find compelling, get stuck in. In a different case, I may approach a problem with one idea and be offered a different strategy. Despite my misgivings, if I follow my teacher's suggestion, my neuro-cognitive state changes and I might learn a technique or skill that generalizes to a range of situations. Therefore I extend my abilities through the exercise of my will (by doing something when asked or enjoined to) and now have options available stemming from my self control or ability to enact what I have been taught, and displace my habitual tendencies. In that way 'what I can do when asked' a criterion for voluntariness, underpins genuine freedom of the type Nietzsche indicates when he dismisses the question of determinism and free will and directs us to look only at strength of will – the ability to enact the result of thought in your life.[36,37]

To recap: voluntary aspect seeing, in cognitive neuroscientific terms, has the brain acting as a whole system imposing whole-part or top-down changes on neural assemblies through the engagement of a human being in a context of discourse. Through instruction and submission to prescriptive norms, and in part by freely spontaneous play, a person masters and voluntarily learns to mobilize combinations of thought and action that s/he has hit on, perhaps by chance.[38]

2.5 Attunement

The translation of dialogue or argument (discourse) into neural patterns able to modulate human activity and structure our interactions

[35] Op. cit. note 1, 705, Hughlings Jackson.

[36] Friedrich W. Nietzsche, 1886 *Beyond good and evil* (Trans. London: Penguin, 1886).

[37] Grant Gillett, 2008 *Subjectivity and being somebody* (Exeter: Imprint Academic).

[38] Op. cit. note 19, 5. Wittgenstein, L. (1972b). *The blue and brown books*. (Oxford: Blackwell). 5.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

with the world is, in part, the function of the brain and the reason why it works both from top down and from bottom up.[39,40,41]

The concept of free will, in our terms, is linked to work in cognitive science and philosophy of mind that identifies our abilities, voluntarily enacted from top down, to manipulate internal representations and therefore our adaptation to the environment.[42] The concepts *attunement* and *enactment* are the perceptual and motor sides of that adaptation. Attunement refers to our ability to perceptually track significant features of the environment but is also understood (in social cognition) where an individual attunes him or herself to the way that others act and react through the mirror neuron system.[43] Attunement therefore allows the person to enact patterns of perceptuo-motor activity that deal with problems by drawing on existing cognitive resources.[44,45] Given that the focusing of cognitive resources is monothematic and serial (the brain is to be in one dominant activation pattern at any given time), the skills mastered by the subject and the cognitive control that they potentiate are an important part of everyday cognition and action whereby the whole organism exhibits learned techniques (as in chess or social problem solving).

2.6 Top-down self control

Reasoning and our ability to heed and follow commands enable a human subject to see the different views of Necker cube and master other perceptual techniques. A human being is attuned to languages and the rules governing them so as to frame his or her behaviour in a variety of ways in diverse situations. That attunement and self regulation ('with the close participation of speech'[46]) both come into play

[39] A.R. Luria, *The Working Brain: An Introduction to Neuropsychology* (Basic Books, 1973).

[40] Op. cit. note 17, John McCron.

[41] Op. cit. note 32.

[42] James J. Gibson, *The Ecological Approach to Visual Perception* (Lawrence Erlbaum Associates, 1986).

[43] Vittorio Gallese, 'Intentional attunement: A Neurophysiological perpective on Social Cognition and Its Disruption in Autism', *Brain Research* **1079** (2006), 15–24.

[44] Karl E. Weick, 'Enacted Sensemaking in Crisis Situation', *Journal of Management Studies* **25** (1988), 4, 305–317.

[45] Jesse Prinz, 'Putting the Brakes on Enactive Perception', *Psyche* **12** (2006), 1, 1–19.

[46] In Luria's terms.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

in aspect seeing as in learning to see the Necker cube in different ways. To follow the commands of another (e.g. 'See the Necker cube in this way!') requires active understanding and the ability to change distributed patterns of brain activity and engages a human being as a meaning-oriented organism with the world in ways governed by rationality and the use of language through techniques of self-control (implicit in human rule-following).

When someone shows me a puzzle picture (such as the dalmation dog under the tree) and says 'See it thus!' you can ignore her command, nothing causes or compels you to obey. If you obey, you have experiences yielding new understandings built on cognitive skills involving the voluntary direction of consciousness. Subsequently you have more freedom to act in advantageous ways in the future because your skills of perception and self-control present you with more options rather than leaving you at the mercy of events and unable to respond as you choose.

The Necker cube shows that my brain states are under the control of my will and that our past experiences, whatever dispositions they may create, do not have power to force us to think or act in certain way.[47] When we freely make decisions, our past experiences, our memories, and the skills we have mastered may play a part in shaping our experiences and decisions at present time, but they do not determine them. They can 1) present me with options for addressing my current task, or 2) frame the task in a certain way. A teacher or even an acquaintance can, through conversation, enhance my cognitive skills so that I can draw on new ways of thinking when I consider what I ought to do, reason about the situation, and consider which strategy should guide me. Each move configures my ongoing brain activity and the ability to make use of our past experiences and memories in this non-automatic way is what Hughlings Jackson and common usage refer to as free will or voluntary action. It represents an adaptation of the whole organism to an ethological (here discursive) norm (or imperative) and is understood within that context (the domain or space of reasons[48]) it is thus, in the requisite way, not best approached through a nexus of physical laws despite not being random or a result of blind chance.

[47] Jean Paul Sartre, *Being and Nothingness* (Trans. H. Barnes London: Methuen & Co., 1958).

[48] As portrayed by John McDowell (*Mind and world* Cambridge (Mass.): Harvard University Press, 1994) or Wilfred Sellars (1997, *Empiricism and the philosophy of Mind*, (Cambridge (Mass.) Harvard University Press).

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

The directing force of reason acts, in conformity with physical laws and their regularity but it plays a distinctive role in human adaptation that make me able to achieve an imagined goal. We are living in this world and part of the physical nexus, as Kant notes, but free will and reason allow us to originate actions, including those voluntary shifts in the ongoing function of our brains. The results then play out according to chains of physical causation but the voluntary change of perception associated with the Necker cube is just one example where shifting the focus of thought changes the patterns of activity in the brain in ways arising from without the scope of neuroscience. In other similar cases I might use my associative network, structured in ways that I am not in control of to do similar things (e.g. thinking of a duck might trigger me to think of quacks, ponds and weeds but thinking of a rabbit might trigger thoughts of burrows, grass and *Watership Down*). The associations reflect physical laws governing neural connectivity but I, the holistic organism with reasoning and self control functions intact, can determine which set of associations I develop at a given time; for instance, decide to keep reading this paper or go mind-wandering.

Free will allows a person: 1) to *actively* think of a viewpoint, 2) to change viewpoints (*to relocate their thinking*); 3) to develop a given view point in depth (*to develop their thinking*); and 4) to choose which viewpoints to follow (*to direct their thinking*). In each case brain activity is channelled in ways adopted for various reasons many of which arise in conversation and argument (e.g. because I am asked). A person's decision (perhaps influenced by habit or attitude) is variably responsive to reason according to the person's self control, current state of mind and body, and strength of will. Thus we may let ourselves down in various ways explored in discussions of akrasia. But an automatic or mechanistic link between stimuli and action do not suffice to explain what is happening and we do need concepts like 'weakness of the will' (which may include being swept along by emotion and so on). A person's ability to choose according to one's all things considered best judgments arises from self development, listening to others, adopting certain attitudes, and framing our experiences in various ways so as to defeat the 'falling into' of a weak will.[49]

The exercise of will can be so quick as to be apparently automatic. When I see the Necker cube first this way and then that, I actively

[49] As Aristotle (1925) *Nichomachean Ethics* (tr. D. Ross) (Oxford: University Press) and Donald Davidson (1980) *Essays on Actions and Events* (Oxford: Clarendon), both argue.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

search – almost automatically – through past experiences to find ways of thinking to match with what I am seeing but I do so through the mastery of techniques that take me from stimulus arrays to meaningful schemata or concepts,[50] as is revealed when a person has difficulty in seeing the Necker cube in more than one way.

One could object that when someone is asked to change their perceptual approach, or maybe just decides to do so, this demonstrates our power over certain brain functions but that fact does not, in itself, defeat the determinist who can reply that the request itself is an example of determining causation or, in the case of my own decision, something else in my psyche, itself caused, that in turn caused my change of perspective. So even if these changes in perception are not automatic or sub-conscious, they are not (or may not be) undetermined. We have, however, argued that the skills involved in perceptual and cognitive switches of the type discussed here result from what Luria calls 'the close participation of speech' and the mastery of the rule governed performance involved in linguistic behaviour. This behaviour follows connections of rhetoric and logic not physical causation and so to explain its effects and the regularities that occur in it is to invoke a different order of things from that of physical energy transactions. For this reason the (deterministic) web of event causation is not the structure at the heart of the metaphysics in question.[51]

2.7 Free will is used in diverse cognitive functions

The will is used to recombine cognitive events freely in our mind on almost any basis, e.g. we can combine stairs and a piano keyboard to create keyboard stairs, we can put stairs together with apples and pairs (according to phonetic rhyme) to create rhyming slang (which may be quite poetic as when 'wife' becomes 'trouble and strife'). This is a pervasive feature of human creativity – new combinations of images, the capacity to direct and re-direct ideas and to actively re-enter, relocate, reconnect and selectively activate neuro-cognitive assemblies (given the masteries required).

Metaphor, cognitive blending, and categorization fuel innovative cognition so that Hughlings Jackson, thinking neurologically, refers to consciousness as an amalgam of will, memory, reason and

[50] 1953, Op. cit. note 20, part 2, section 11.

[51] (Unless one accepts the contentious metaphysical theses we undermine in part 3.)

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

emotion.[52] Attention is the most obvious technique of self control under the direction of the will and we use it every time we select from a perceptual array just the features that support the discernment of something we are looking for. When we do so, we engage creatively and in an exploratory way, mobilising different cognitive schemata so as to select from various rule-following abilities those needed to construct useful experiential structures that can be enacted in the analysis of experience and in problem solving.[53]

We choose concepts to re-frame puzzling information from the environment[54] (e.g. *a cube with its top left hand corner coming out of the page, a rabbit*) when we categorize things. We also mobilize different metaphors to blend information and reconfigure data gleaned from experience (therefore we can see a beam of light as a wave or a stream of particles) or create modified concepts (normal family based on a single breeding pair, extended family, reconstituted family). Thinking is not, to use Hughlings Jackson's term 'automatic', and close links between perception, imagination, and problem solving provide subjects with cognitive systems as rich as ours. Thinking of this type is not always driven by new information (although it may respond to incidental associations) so that the process is not determined from bottom up (as in the Necker's cube, the duck/rabbit or the old woman/young woman figures).[55] We have the freedom to think in ways related to but not given by the stimulus arrays we engage with and therefore we can potentially decide what aspects to see, and what interpretations, attitudes, and theories to generate from our perceptual data.

But here we might object to talk of our will deploying concepts to reframe puzzling information but, for example, in the dog case (Figure 2), try as I might, it may be that I can't see the dog; and when I finally see it, it jumps out at me, as something almost forced on me and then I cannot not see it. Sometimes switching views on the Necker cube and the duck-rabbit seems rather like that, and maybe even despite my efforts, which may have impeded things by my trying too hard and perhaps focusing on the wrong things. We should, however, notice that the perceptual skills involved are cognitively manipulable to a certain extent but neither completely cognitively penetrable nor completely cognitively malleable. That is like

[52] Op. cit. note 1, 704.

[53] Grand Gillett, 'Freedom of the will and mental content', *Ratio* **6** (1993), 89–107.

[54] Anne Treisman, 'The Binding Problem' *Current opinion in neurobiology* **6** (1996), 2, 171–178.

[55] Op. cit. note 23. Gleitman.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

any voluntary activity relying on partially automatized subconscious mechanisms and processes – a strong will sufficient to direct one's activity does not mean that one always succeeds nor that the result is totally within one's control, only that, for reasons that may be intelligible to me alone, I set in train the events that dynamically and skilfully engage me with my environment and, in that respect, the explanatory 'buck' cannot be passed beyond my reasons – it stops there (to paraphrase Kant[56]).

3. Metaphysical obstacles to cognitive neuroscientific sense

We are left with three outstanding metaphysical problems.

3.1 How is it that a general theory of physical causation does not imply a deterministic theory of brain function?
3.2 Why should we resist mechanistic causation as a nomological or singular causal relation between brain events?
3.3 Why does nomological causation not apply to mental events and yet principled mental explanation is possible?

3.1 Determinism and the brain

Robert Hanna[57] picks up some similar strands of thought within a Kantian analysis arguing that we are open ended biological systems that are self organising as loci of operations on the environment obeying laws of dynamic systems theory. We have also invoked the discursive interactions that make one a conscious rule-follower. Propositional function and cognitive frames are used to construct a situation and potentiate actions that creatively respond to it in a voluntary self-controlled way. Human thinking, giving rise to decision and action, is not deterministic in the way required by metaphysical arguments against free will because it originates with my making a situation intelligible to myself in rule-governed ways (formulated within the domain of reason) and therefore in a context where I do some things because I am asked or instructed to. Thus -

Argument from biological systems theory:

3.1.1 Unfree will is an inference from mechanistic determination.

[56] As in 1.1 above.
[57] Robert Hanna, 'Freedom, Teleology, and Rational Causation', *Kant Yearbook* **1** (2009), 99–143.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

3.1.2 Mechanistic determination is not the process at work in open-ended Biological Systems in Dynamic Relationships with their Contexts (BSDRC).
3.1.3 The will is a way of referring to the operation of a human being as a BSDRC.
3.1.4 Human beings as agents do not have unfree will.

Hanna argues that human information is stored in a dynamic way that does not determine a specific direction of causal determination as it would in a closed nomological system. By gathering information and knowledge, a brain develops a more dynamic configuration that inherently presents indefinitely many possible solutions to a problem (bounded only by imagination and connections available through the semantic field used to structure cognitive frames). The brain as a biological system in dynamic relationship with its context interacts with that context to create a steady state mode of function rather than homeostatic reversion to a stable end point [thus the BSDRC]). Such a dynamic system does not yield free will in itself but does mean that I have become a better instrument for the exercise of free will in so far as I am truly responsive to the imperatives of reason and not merely the demands of biology or regression to a dispositional mean.

Animals have dynamic biological systems responsive in this way to their biological purposes, but not free will because they cannot actively control their thinking in the light of reason and (on the basis of argument or counterfactuals). Their passive/passionate responses to environmental stimuli do not engage with metaphor, imagination, rule-following abilities and so on. This even if they have more options in certain situations because of superior natural capacities (e.g cheetahs can run faster than us), and they may generalize responses from an habitual domain (as in chimpanzees' tool use) but their responses are biologically limited by a mix of phylogeny and ontogeny. They see in ways they are accustomed to (as do we on many occasions) and cannot actively think of the same thing in the huge diversity of ways available to a human being (think of a toddler's imaginative play). Only human beings actively control and direct their minds 'with the close participation of speech' (Luria) and actively choose from the options that they bring to their attention in that way. For that reason, as Aristotle remarks, animals see what is true and thinking, which can be false 'is found in no animal in which there is not also reason'.[58]

[58] As in *De Anima* (tr D.W. Hamlyn) Bk III.3 [427b12] Oxford: University Press. Wittgenstein also remarks 'thought can also be of what is not the case'(PI 950).

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

3.2 Mechanistic causation and brain events

One of us has discussed this at length elsewhere.[59,60,61] We might briefly rehearse just two arguments relating to singular causal connections and metaphysical indivduation.

3.2.1 The no independent existence argument.

This argument notes that within the seamless web of brain excitation there are neuronal events that precede a voluntary movement but there is no discrete event comprising the action nor one comprising the neural substrate of an intention to act (see 3.3.1–3.3.4 for Libet's version of this claim). Therefore there is no discrete brain event that is both unconscious and that causes the conscious action because there are no independent existents to serve as cause and effect.[62] In fact Dennett also attacks this view in that it requires us to locate the internal causes of mental events in particular places in the brain at particular identifiable times (evident to the mental subject).[63]

3.2.2 The metaphysical status of the relevant brain events.

There is a troubling naivety of 'the belief in vertical fidelity' in metaphysics. The thought is that an explanatory type in the world of mind or meaning has a corresponding individuable type evident to neuroscience. Consider the mental type associated with *boojum*: to identify a relevant set of neural states would require that we map uses of the term *boojum* onto tracts of brain activity. But the set of neural events generated is unlikely to have any principled individuating criteria sufficient for designating it a distinct neuroscientific type. Thus, as a metaphysical species, it does not have its own specifications but depends for its individuation on the level of the subject as engaged with the world of mind and meaning. That the neuroscientific conditions are neither necessary nor sufficient to identify the events or

[59] Op. cit. note 35, G. Gillett 2008.

[60] Grant Gillett, 'Free Will and Events in the Brain', *Journal of Mind and Behaviour* **22** (2001), 3, 287–310.

[61] Grant Gillett, 'Intention, autonomy and brain events', *Bioethics* **23** (2009), 6, 330–339.

[62] Michael Morris, 'Causes of Behaviour', *The Philosophical Quarterly* **36** (1986), 143, 123–144.

[63] Op. cit. note 28, Dennett 2003.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

states concerned is evident in a simple possibility: if, having delineated a set of brain states claimed to be definitive of mental state M', a new subject who qualified as being in M' – for instance as understanding the term *boojum*, did not have a neural state fitting the parameters (quite probably unspecifiable if the argument above is correct) then the duscursive ascription of M' is decisive and the brain-based specification has to follow it. Therefore the individuating conditions of the putative neuroscientific entity depend on a nexus of mental relations rather than on anything generated by neuroscience (or suited to subvene M').[64]

3.3 Causation, mental events, and brain events

The Necker's cube and other examples of the voluntary control of brain states indicate that 'top down causation' is a reality, *pace* some theorists.[65] Firstly, the idea of causality of that type seems to be indispensible in making sense of what is going on at the neuro-level (as in the human ability to switch perspectives at will and the change in brain states that must be correlated with it). Secondly we must heed the fact that we select from the neurophysiological or brain imaging record states of interest only because events at the personal or communicative level pickout the tract of interest. Therefore the agent level is indispensible in identifying the actual entities (should there be such) that count as physical states of a given type (the neuroscientific states that constitute under-standing the term *boojum* for instance). Dependence of the subvenient level on types individuated at the personal or whole subject/agent level, together with the failure of individuation of suitable cause or effect events at the neuroscientific level, seem to be secure metaphysical nails in the lid of the coffin of any supervenience thesis. Thirdly, the states that a brain gets into may only be able to be explained by tracing the history of that brain as the brain of a person located in the human life world at a certain point such that certain adaptations have shaped its neural networks – e.g. learning Swahili, then being removed to New York, then being trained as a microsoft graphics designer for 'Sim' games, and so on. But if neural or brain states are, in principle, only explicable by appealing to such things as meaningful human relationships, being

[64] Op. cit. note 9, Gillett 2011 review of Kim's metaphysics of mind, *Philosophy* **87**, no. 339 (this issue).

[65] Op. cit. note 8, Kim 2010 *The metaphysics of mind* (Oxford University Press).

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

exposed to cultural icons (such as 'the Hunting of the Snark') and causation is a reflection and articulation of the metaphysical dependence of a state of affairs on those structures and events that explain it, then the metaphysics of mind required for neuro-determinism embeds a serious mistake.

What is more we might notice that the intentional explanation of action often depends on actual mental states that have dynamic content (e.g. what I see before me here and now is a fridge containing beer when I last looked). Such states (often referred to as dynamic mental states with *de re* or world-involving content in their individuation) are necessarily singular in their intentional object specifications even though instances of demonstratively context dependent senses.[66] Thus the particularities of their explanatory role cannot be filled by a general specification of the objects concerned apt for nomological (or projectable) laws. Therfeore, if we cannot appeal to singular causation because of a metaphysical failure to generate a motivated identification of events to serve as cause and effect, nor to nomological causation because of the lack of nomological types with the relevant explanatory content, we seem to lack any account of causation that will do any deterministic work for us in the explanation of human action and the way is clear for the non-reductive top-down causation that we see in the Necker's cube example.

We should lastly, in the spirit of naturalistic metaphysics, revisit the Libet arguments that allegedly undermine our belief in the conscious will as the origin of action.[67] Dennett, as noted, undermines the naïve view that there is a place and time in the brain where the extensionless point of conscious willing might occur.[68,69] One of us has also attacked Libet's model and its assumptions, arguing that the claim that there is a brain event prior to action, which is unconscious and only reported after the action is initiated, embeds the following false premises.[70]

3.3.1 That an action is a discrete bodily movement (M_B) is false in that an action is a dynamic pattern creating an effect in

[66] Gareth Evans, *The Varieties of Reference* (Oxford and New York: Oxford University Press, 1982).
[67] Op. cit. note 6. LIbet.
[68] Daniel C. Dennett, *Consciousness Explained* (Little, Brown and Co., 1991).
[69] Op. cit. note 28, Dennett 2003.
[70] Op. cit. note 35, Gillett 2008.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press

the environment (an operant) with an environmental target and not a pattern of motor events as its basis.[71]

3.3.2 That there is a mental event which denotes the actual physical cause of the act fails due to the arguments above about the individuation of brain events and states.

3.3.3 That one can fix the time of a mental event on the basis of its reportability is false because reportability itself is the coming together of a series of cogitations in which the subject is self monitoring and also initiating an action[72] and such mergings indeterminately unfold in the neurocognitive domain rather than being datable physical happenings.

3.3.4 That the detectable brain event is the cause of the act rather than being a reflection of preparatory moves or neural events involved in acting with intent is a hangover of the inner impetus model of action that one of us has attacked at length and in many places.[73]

We would therefore conclude that there are no metaphysical obstacles to accepting, as the case of Necker's cube seems plainly to show, that human beings are organisms that perform voluntary actions in the sense of things we can do when asked as a result of conscious decisions to act.[74,75] Therefore human beings have all that they need in terms of freedom of the will and no findings about brain states have done anything to undermine that claim.

Bioethics Centre, University of Otago
Neuroscience Department, University of Otago
grant.gillett@otaho.ac.nz; liuch471@student.otago.ac.nz

[71] Op. cit. note 34, Luria 1973.
[72] Op. cit. note 28, Dennett 2003.
[73] Gillett, G. (2001) Intention and agency in Ngaire Naffine, Rosemary J. Owens, and John Matthew Williams, *Intention in law and Philosophy* (Burlington: Ashgate), 57–69.
[74] Op. cit. note 4. just check the note numbers to see they still apply.
[75] Op. cit. note 28, Dennett 2003.

https://doi.org/10.1017/S003181911100057X Published online by Cambridge University Press