HIT and Brain Reward Function: A Case of Mistaken Identity (Theory)

Cory Wright (cory.wright@zoho.com) Department of Philosophy, California State University Long Beach, 1250 Bellflower Boulevard, Long Beach CA 90840–2408 USA

Matteo Colombo (m.colombo@uvt.nl) Tilburg Center for Logic, Ethics, & Philosophy of Science, Tilburg University, PO Box 90153, 5000 LE Tilburg NL

Alexander Beard (alexander.beard@zoho.com) Department of Philosophy, California State University Long Beach, 1250 Bellflower Boulevard, Long Beach CA 90840–2408 USA

Abstract. This paper employs a case study from the history of neuroscience—brain reward function—to scrutinize the inductive argument for the so-called 'Heuristic Identity Theory' (HIT). The case fails to support HIT, illustrating why other case studies previously thought to provide empirical support for HIT also fold under scrutiny. After distinguishing two different ways of understanding the types of identity claims presupposed by HIT and considering other conceptual problems, we conclude that HIT is not an alternative to the traditional identity theory so much as a relabeling of previously discussed strategies for mechanistic discovery.

Keywords: dopamine, explanation, identity, localization, mechanism, reward

1. Introduction

The central thesis of the mind/brain identity theory is that every type of mental state or process is identical with some type of brain state or process. Early proponents of the theory intended to rebut two general claims: that mental states and processes are something other than brain states and processes, and that mental states are merely correlated with brain states.[1] Later proponents of the theory intended to enshrine a few further theoretical commitments: that kinds of mental states and processes are reducible without remainder to certain kinds of states and processes in the brain; that the identity relation involved in mind/brain identity statements is necessary; and that mind/brain identity statements are neither analytic nor a priori.

Sadly, the identity theory fell on hard times. From the early 1960s to the late 1980s, it was beleaguered by numerous philosophical challenges, including concerns over multiple realizability, violations of Leibniz's law and the modal logic of identity statements, and the apparent incorrigibility of introspective reports.[2] Since the early 1990s, examinations of case studies and actual scientific practice have become equally important prongs in both challenges to, and defenses of, philosophical theses about mind/brain relationships (Bickle 2003); but unfortunately, traditional identity theorists have had a paucity of cases of psychoneural identities to celebrate.

These challenges have created the conditions for an alternative to the traditional identity theory, dubbed the 'Heuristic Identity Theory' (HIT). The traditional theory

[1] See Place (1956); Feigl (1958); Smart (1959); Armstrong (1968). Note that identity theorists have often interpreted uses of the term *brain states* as being 'just a placeholder for whatever the eventual working parts of brains—in fact, of nervous systems, perhaps embodied nervous systems—turn out to be' (Polger 2009: 3).

[2] For defenses of the identity theory on conceptual and empirical grounds, see, e.g., Enç (1983); Shapiro (2000); Sufka & Lynch (2000); Polger (2004, 2009); Couch (2004); Shapiro & Polger (2012).

supposes that scientists work by generating psychoneural correlations and then accumulating evidence for type-identities to explain them. According to HIT, this supposition puts the cart before the horse. Psychoneural identities are not discovered after a period of protracted scientific research; instead, scientists intrepidly hypothesize them at the outset of inquiry and then use those hypotheses as a discovery heuristic for driving further research.

Interest in HIT has coincided with other favored philosophical trends, including the recrimination of reductionist themes in the psychological sciences, the rise of explanatory pluralism, and the mechanistic approach to explanation. Embracing pluralism and mechanistic explanation in particular (e.g., Bechtel & Wright 2009), advocates of HIT have appealed to a few cases from neuroscientific practice to support the claim that 'hypothetical identities […] regularly serve as the critical premises in explanatory proposals that inaugurate new lines of scientific investigation' (McCauley 2012: 192).

This paper examines this claim, asking two questions. Has HIT received sufficiently convincing empirical support? Is HIT a genuine alternative to the traditional identity theory?

To answer these questions, we shall proceed as follows. In §2, we survey HIT's basic commitments. In §3, we detail the history of electrophysiological and neuropharmacological research on brain reward function (BRF), which provides an exemplary case study against which to test the predictions and commitments of HIT. Unexpectedly, this case failed to support HIT, motivating closer scrutiny of previously invoked cases. In §4, we argue that, on a standard conception of identity, these previous cases also fail to provide convincing empirical support for HIT; on a broader conception of identity, existing case studies do support HIT, but not as a version of identity theory, as advocates present it.

Philosophical attempts to detail the claims of HIT have only supported a broadly mechanistic approach to mechanistic explanation and mechanism discovery, which often posits hypotheses about the localization of psychological functions in the brain, or hypotheses about the causal production of psychological phenomena by certain neural mechanisms. While scientists often advance such hypotheses as heuristics for discovering mechanisms (Bechtel & Richardson 1993/2010), it is a mistake to confuse these relations for the identity relation. As it is not a theory about identity, HIT is not any kind of mind/brain identity theory. In particular, HIT is not a new alternative identity theory so much as an epicycle on what has been called the 'new mechanical philosophy' in cognitive neuroscience. We consider various additional conceptual problems in §§5–6, and conclude in §7.

2. Heuristic Identity Theory

2.1. ORIGINATION

HIT seems to have originated from remarks by William Wimsatt (1976a: 227–229), who suggested that explanation—not reduction—is the proper context in which to understand the role of psychoneural identities in science. For Wimsatt, identities are not the endpoint or goal of an intertheoretic reduction, and they do not serve as a regulative ideal against which intertheoretical relations are formally judged. Rather, they are 'tools' used primarily to ferret out errors, and thus play a merely enabling role upstream in the process of refining explanations: '[i]dentity claims […] provide probes of potentially unlimited sensitivity and depth for pinpointing sources of explanatory failures' (1976a: 227). Wimsatt's remarks were later reformulated, en

passant, by Robert McCauley:

> Instead of identities being assigned late in the game to those coextensionalities which prove persistently recalcitrant to explanation, they are often proposed relatively early, initiating wholly new lines of research. When in doubt (many scientists) assume the truth of a proposed identity until empirical research clearly indicates otherwise. The postulation of identities is a research tool for extending the explanatory range of theories. They are not proposed as the grounds for justifying eliminative moves in microreductions (even if, after the fact, they may be cited as such). (1981: 225)

Unfortunately, HIT went dark after these two opening salvos, but was resurrected in a series of papers by McCauley & William Bechtel.[3]

From this body of work emerged the central thesis of HIT: identity statements are discovery heuristics.[4] Bechtel & McCauley put the thesis this way: 'claims between psychological processes and neural mechanisms [sic] are advanced as heuristics that serve to guide further research' (1999: 71). The thought is that identity statements play a heuristic role in the initial development and guidance of research in multiple fields and at multiple levels of analysis: 'identity claims are made early in a research program and serve as heuristic for further research' (Bechtel 2002: 236).

The process of advancing heuristic identities has multiple phases, which McCauley & Bechtel (2001: 751) described as follows. Initially, identity statements involve discrepancies. These discrepancies prompt further research at various levels of analysis to ascertain which proposals should prevail or in which directions to proceed. Next, this further research yields more precise hypotheses about the systems and patterns engaged, provoking new speculations at multiple explanatory levels. Then, speculations imply new ways of orchestrating familiar facts and theories within levels, and suggest new avenues of research. Finally, some of these avenues of research produce new cross-scientific conflicts, which likely begin this cycle anew.

So, in positing identities between mental states or processes and neural states or processes, scientists can apply knowledge of neural mechanisms to guide the development of models of mental states and processes at higher levels of analysis; and then they can justifiably rely on knowledge of mental states and processes to search for lower-level neural mechanisms. As McCauley & Bechtel put it, 'what we learn about an entity or process under one description should apply to it under its other descriptions' (2001: 753), which Bechtel repeats elsewhere: '[… u]nlike [the] traditional identity theory, the focus is on using the differences between what is known about the processes under each description as a discovery heuristic to revise the other' (2008a: 990). For advocates of HIT, what ultimately matters is the explosive amount of science that can be generated by positing identities. If psychoneural identities play this role—fecundity—in scientific practice, then HIT would rest vindicated in a way that eschews the metaphysical issues raised by traditional versions of identity theory.

2.2. THREE OVERARCHING COMMITMENTS

The thesis that identity statements are discovery heuristics is analyzable into at least

[3] Bechtel & McCauley (1999); McCauley & Bechtel (2001); Bechtel (2002, 2008a,b); McCauley (2012); see also Schouten & Looren de Jong (2001); Looren de Jong (2006); Bechtel & Hamilton (2007).

[4] See McCauley & Bechtel (2001: 753) and Bechtel (2008b: 71); see also Bechtel & McCauley (1999: 67, 71); Bechtel & Hamilton (2007: 414).

three successive commitments. Firstly, identity statements are 'thoroughly hypothetical' statements about psychoneural relations (Bechtel & McCauley 1999: 67, 71). Secondly, these hypothetical identity statements are posited at the outset of a period of scientific research. As Bechtel & McCauley wrote, 'scientists adopt [identity statements] as hypotheses in the course of empirical investigation to guide subsequent inquiry, rather than settling on them merely as the results of such inquiry' (1999: 67; see also Bechtel 2002: 236), and again, '[i]dentity claims (e.g., that water is H_2O or pain is C-fiber firing) are […] made at the outset of investigation, often on the basis of a limited number of correlations' (Bechtel 2008b: 70). Thirdly, these inaugural statements are then used to facilitate self-correcting research at multiple levels of analysis and explanation. As Bechtel wrote, '[t]he one thing that is necessary in an overly simple initial hypothesis is that it suggest a line of research that can reveal its own inadequacies and hence point to profitable revisions' (2002: 235).

These three commitments imply several interesting features. One is that, being thoroughly hypothetical statements about psychoneural relations, these identities have truth-values (and so are truth-apt). Another is that these thoroughly hypothetical truth-evaluable statements are empirical. A third feature is that these thoroughly hypothetical truth-evaluable empirical statements are expected to be descriptively inadequate (typically, either false or else not true, relative to which logics are employed). Advocates of HIT have been clear and consistent in characterizing identities as statements that are expressly introduced to generate research that may heuristically expose its own errors and lead to improvements in our epistemic situation. Indeed, because their revision and refinement is a central objective of subsequent scientific research, these identities are intended to be disposable—i.e., the proverbial 'ladder' that is eventually thrown away after stimulating new research—unlike scientific truths, which are usually precious and hard-won.

2.3. AN INDUCTION OVER CASES

Advocates present HIT as a general strategy of discovery that is commonly employed during the early stages of inquiry, particularly in cognitive neuroscience (Bechtel & McCauley 1999: 67). The justification for HIT is generated by the method of induction over cases: advocates attempt to document many actual cases in a variety of scientific fields, where identity statements are made at the outset of research and subsequently play a heuristic role in driving future research before eventually being discarded. Consequently, if HIT were correct and faithful to scientific practice, we would expect an array of supporting historical episodes throughout the sciences.

Bechtel & McCauley (1999) and McCauley & Bechtel (2001) made their initial case for HIT by appealing to early research on vision, and Bechtel's (2008b) account of memory encoding was the only other main documented test case of heuristic psychoneural identities for almost a decade. Another case study pertaining to human face recognition was recently proposed in an attempt to strengthen this induction (Burnston et al. 2011; McCauley 2012; cf. Kievit et al. 2011).

The basic form of each case study is threefold: a report of the onset of a research period, in which an alleged identity statement bridged research at different levels of analysis and explanation, followed by evidence that the identity statement provided the impetus for increasingly sophisticated research and new scientific discoveries at different levels; and finally, in order to show that identity statements are heuristic and disposable, a demonstration that the identity statement was later discarded once

the task of enabling was discharged.

One problem confronting HIT is that two or three (or even four) cases make for a weak induction, and does not support the claim that HIT is 'a common discovery strategy in science' (Bechtel 2008b: 71). Accordingly, we turn to the history of electrophysiological and psychopharmacological research on brain reward function (BRF) in the following section. With a clearly demarcated period of inquiry, research at multiple levels of analysis connecting psychological and neural states, a variety of human and non-human animal studies, rapid scientific discovery at multiple levels, and clear parallels with the cases HIT advocates have already appealed to, we initially hypothesized that BRF research would be fertile ground for psychoneural identity claims, and would yield additional empirical support for HIT.

3. Heuristic Identity Theory (Or Back to the Past): Locating Identity Claims in BRF Research

3.1. ELECTROPHYSIOLOGY: THE FIRST STREAM (1950–1956)

3.1.1. *Serendipitous discoveries*

In 1950, Robert Heath led a Tulane University medical team in exploring whether electrical stimulation could effectively treat schizophrenia. By the summer of 1952, Heath's team had operated on 24 (conscious) patients, stimulating various regions—including the caudate, hypothalamus, anterior thalamus, and septal area—while each patient reported their subjective experiences. 'Alertness' was commonly mentioned, and patients also reported experiencing anxiety, confusion, and lucidity. But in the four cases involving septal stimulation, patients described their experiences as 'jovial', 'euphoric', and 'pleasant' (Heath 1954). Unfortunately, Heath was primed to focus on observations regarding the relation between stimulation and alertness, and overlooked the significance of the serendipitous discovery that septal stimulation was associated with pleasure (Baumeister 2006: 97).

While Heath's team was performing their surgical experiments at Tulane, Peter Milner was at McGill University conducting his doctoral work under Donald Hebb's supervision; Seth Sharpless, a philosophy student from Chicago, was working likewise in Herbert Jasper's nearby laboratory. With Sharpless, Milner ran pilot experiments on electrical stimulation of rats' reticular activating system; but they were only able to demonstrate that the stimulation was aversive (rather than motivating, as they had hoped). In the fall of 1953, James Olds arrived at McGill to begin his postdoctoral fellowship, and Hebb had Milner teach Olds some basic neuroanatomy along with their intracranial self-stimulation (ICSS) paradigm. Olds's initial attempt to plant an electrode in the reticular activating system unexpectedly resulted in one rat who behaved as if it were being rewarded by stimulation. X-rays confirmed that Olds's implantation had missed its mark, and that the electrode tip had instead terminated near the septal area.

Further ICSS experiments allowed Olds & Milner to demonstrate that rats will work hard to repeatedly operate a manipulandum for direct electrical stimulation to the midbrain. In one case, a rat repeatedly bar-pressed 7,500 times in 12 hours (1954: 425). In some cases, response rates rose nearly to 5,000 times in an hour; and in another case, one rat bar-pressed up to 2,000 times per hour for 24 consecutive hours (1956b: 114–115). The most robust and reliable ICSS-elicited behavior was generated from the median forebrain bundle (MFB) in the mesolimbic system, which Olds (1956b)—in his exposé of their discovery—called the brain's 'pleasure center'.

Together, these two serendipitous discoveries—one pursued, one not—mark the outset of over 60 years of inquiry into brain reward function. Olds & Milner's (1954) demonstration that stimulation could be highly rewarding (or pleasurable or motivating or reinforcing—they knew not which) immediately instigated hundreds of further studies attempting to replicate and interpret their findings. Because they provide an operationally-defined measure of reward function, ICSS experiments quickly became a dominant investigative approach to plumbing the relationship between states involving neural activation of the MFB and the hedonic state of psychological pleasure. Within just a few years, experimenters showed that the neural circuitries subserving reward function could be homologously activated in a wide range of non-human animals beyond rats (e.g., snails, fish, rabbits, cats, guinea pigs, dogs, horses, and various non-human primates), suggesting a conservation of evolutionarily-basic mechanisms across reward-dependent neural state types. Later experiments also confirmed that electrical stimulation of specific brain regions has the same impact on motivation as other natural rewards, like food or water for hungry or thirsty animals (Morrison 1955; Brady et al. 1957; Trowill et al. 1969).

3.1.2. *Identity crisis*

The discovery of the brain's 'pleasure center' was a major episode in the history of behavioral neuroscience. HIT implies that we should expect hypothetical psychoneural identity claims at the outset of inquiry—i.e., statements of the form *x is y* or *x is identical to y,* where '*x*' and '*y*' are the types of mental and neural relata named in statements of identity or equality (e.g., '*pleasure = MFB activation*'). A close examination of this initial period of research, however, reveals no such statements.

Olds & Milner (1954) emphasized three 'reward' regions: the septal area, cingulate cortex, and mammillothalamic tract. Their observations led them to make three main classes of claims. First were claims of correlation. For instance, in reviewing the previous decade of ICSS work in rodents, Olds & Olds concluded that '[…] the correlation of positive reinforcement with the anatomic system based on the medial forebrain bundle appears to be valid' (1963: 271; see also Olds 1956a: 281). In human patients, Heath & Gallant observed, '[i]n 14 of our 54 patients, we were able to correlate some types of reproducible electrical activity in specific areas of the brain with patterns of thought activity […] The outstanding feature of the data obtained with our methods is a correlation between physical activity in specific parts of the old olfactory structures of the brain and emotional behavior' (1964: 86, 104). And in his literature review on ICSS and BRF, Philip Zeigler responded to the complaint that the greatest gap in physiological psychology is the absence of any convincing physiological correlate of learning and motivation by saying that '[e]lectrical stimulation of the brain is among the most promising of the available techniques for establishing such correlates' (1957: 363).

Second were localization claims: e.g., '[w]e are led to speculate that a system of structures previously attributed to the rhinencephalon may provide the locus for the reward phenomenon', and again, '[t]he brain does seem to have definite loci of pleasure and pain, and we shall review here the experiments which have led to this conclusion. […] Our present tentative conclusion is that emotional and motivational mechanisms can indeed be localized in the brain' (Olds & Milner 1954: 425; Olds 1956b: 105, 116; see also Bursten & Delgado 1958). Zeigler, encapsulating these first two kinds of claims, lamented that '[o]n the basis of the studies reported thus far, only one anatomical generalization may be made. All the structures involved in this phenomenon may be included in or related to the so-called 'limbic system'' (1957: 363, 373).

Third were claims about the causal and functional roles of certain neural

processes: e.g., 'electrical stimulation in the septal area has an effect which is apparently equivalent to that of a conventional primary reward […]' (Olds & Milner 1954: 421); and e.g., '[… w]e have perhaps located a system within the brain whose peculiar function is to produce a rewarding effect on behavior', and again, '[…] there is an area in the basomedial forebrain at about the level where the islands of calleja are most pronounced where electrical stimulation has extremely great rewarding effects' (1954: 426; Olds, 1956c: 511). Other scientists building on Olds & Milner's initial discovery took them to have 'demonstrated a rewarding effect produced by electric stimulation of some areas of the brain' (Sidman et al. 1955: 830; see also Olds 1958), although a few were less assertive. Zeigler, encapsulating these latter two kinds of claims, cautioned that '[i]t has not yet been conclusively demonstrated that we are dealing with 'pleasure centers' or 'a system within the brain whose function it is to produce a rewarding effect upon behavior'' (1957: 373). Even still, the use of causal-mechanical *production*-language of the form *activity in brain region x produces or elicits effect y* turns out to be a persistent trend in early ICSS literature. But to interpret this *production*-language in terms of heuristic identities would be to commit the infamous 'Betty Crocker fallacy' in which non-causal relations like identity or constitution are confused for causal ones (Churchland 1998). To their credit, advocates of HIT have typically avoided this fallacy (e.g., Wimsatt 1976a,b; Craver & Bechtel 2007); but continuing to do so requires restricting their basis of textual evidence so that statements about the causal production of mental states and processes are not confused with identity statements. Prima facie, production is one relation, identity another.

These three types of relational claims—i.e., correlation, localization, and causal/functional role ascription—are no more consistent with HIT than with its competitors; traditional identity theorists and functionalists, for example, could accommodate them all. More problematically, whereas these three types of relational claims are ubiquitous in the scientific literature, relational claims about any kind of identity (type/type, token/token, etc.) are altogether absent from all major scientific reports comprising the early ICSS literature. Heath's work did not include them. Olds's papers did not include them; neither did Milner's, nor their joint papers, nor those that they wrote with neuroscientists Brenda Milner or Marianne Olds. Identity claims are also absent from the Masters thesis of Olds's main student at the time, Rolfe Morrison, and they do not appear in Sharpless's work with Jasper.

A fourth class of claims were those of a merely predicational nature: e.g., '[there are] subcortical areas in the human brain in which brief electrical stimulation appears to have rewarding or reinforcing properties' (Bishop et al. 1963: 396). But in all such cases, statements of predication, function and production, location, or correlation fail to amount to anything resembling the identity claims, and so fail to vindicate HIT. Since the ICSS literature provides a canonical case study for HIT and HIT implies that identity statements are posited at the outset of research, the lack of identities in the literature is unsettling.

3.2. NEUROPHARMACOLOGY: THE SECOND STREAM (1957–1976)

3.2.1. *Dopamine*

Just when the electrophysiology of BRF was taking shape, so too was the neurochemistry underlying reward. The mid-1950s saw early suggestions concerning the physiological significance of dopamine (DA), along with further discovery of DA within the brain (Blaschko 1957; Montagu 1957; Carlsson et al. 1958). As investigations into the occurrence of DA neurons continued, areas with high concentrations of DA (specifically the corpus striatum) were shown to contain less

noradrenaline (NA) than expected (Bertler & Rosengren 1959). Some took the distribution of DA in the brain to indicate a function unique to this previously-overlooked neurotransmitter (Carlsson 1959). Advancements in histochemical methods allowed researchers to trace the brain pathways of catecholamine neurons within a decade, leading to the discovery of a nigro-neostriatal DA neuron system, and DA pathways projecting from the substantia nigra pars compacta and the ventral tegmental area (Dahlstrom & Fuxe 1965; Anden et al. 1966; Hornykiewicz 1966).

Neuropharmacological research proved to be integral to the formation of catecholaminergic theories of reward. In one of the earliest attempts to bridge neuropharmacological and electrophysiological ICSS research, Olds et al. (1956) showed that reserpine and other catecholamine-blocking tranquilizers diminished reward functionality in hypothalamic- and septal-area electrode placements. Other early experiments investigated the effects of tranquilizing agents on rates of self-stimulation, and, following histochemically-verified sites of NA fibers, led to the hypothesis that reward was primarily mediated by NA (Olds & Travis 1960; Stein 1968; Margules 1969). The noradrenergic theory of reward was further bolstered by the finding that tendencies toward self-stimulation were abolished by inhibiting the conversion of DA to NA (Stein & Wise 1969; Wise & Stein 1969).

The 1970s, however, saw the role previously ascribed to NA eclipsed in favor of a dopaminergic theory of reward.[5] Facilitation of DA activity with DA agonists (e.g., amphetamine) have concomitant effects on ICSS response rates (Crow 1972), while administration of DA antagonists suppress self-stimulation (Franklin 1978). On the other hand, neither DA-β-hydroxylase inhibitors (e.g., disulfuram), which selectively block NA synthesis without perturbing DA, nor neurochemical 6-hydroxydopamine lesions of nuclei in the lower brain stem (e.g., A1, A2, A5, A7) and locus coeruleus (A6), which deplete forebrain NA, diminished self-stimulation behavior (Lippa et al. 1973; Clavier et al. 1976; Koob et al. 1976; Corbett et al. 1977). These and other findings led to the emergence of the DA theory of reward as the most promising and fruitful investigative direction.

The catecholaminergic theories of reward advanced in the late 1960s and 1970s would seem to be a good place to uncover some of the identity claims predicted by HIT. Identification of the psychological states of pleasure and reward—not simply with states of the MFB as its neuroanatomical locus—but with the activities or processes of neurotransmitters (e.g., NA and DA) throughout the brain's mesocorticolimbic systems, would square nicely with the standard HIT narrative.

While relatively simplistic hypotheses involving certain mental states and neural states were superseded by those involving NA, and later, DA activity, statements of identity relations remain absent from the scientific literature. With regard to earlier NA theories of reward, one sees the same classes of claims: production, correlation, localization, and functional role ascription—and occasional grounding-like claims. For instance, Stein stated that '[…] electrical stimulation of the MFB produces at least part of its rewarding […] effect by activating adrenergic synapses in the lateral hypothalamus, pre-optic area, limbic lobe, and neocortex' (1967: 139). Commenting on the significance of the observed release of NA into the rostral hypothalamus and amygdala by 'rewarding MFB stimulation' Stein & Wise concluded 'that the release of norepinephrine at these and other terminal sites of the MFB in the forebrain is responsible, at least in part, for the facilitation of behavior caused by rewarding stimulation' (1969: 197). Cutting to the chase, Margules reported that '[…] dorsal

[5] For an overview of this literature, see Hornykiewicz (1986, 2002); Milner (1991); Marsden (2006); Wise (2008).

tegmental reward has a noradrenergic basis' (1969: 34). These and hundreds of other similar statements again reveal that identity claims—heuristic or otherwise—are not posited at the outset of research. Rather than claims about the relation of identity between psychological states of pleasure, reward, motivation, or the like, on one hand, and the activity of a major neurotransmitter system on the other—as one would expect if HIT were correct, what one finds are attributions of a generic 'basis' or 'role' of various neurotransmitters in the reward phenomenon.[6]

In sum, examination of the history of BRF research reveals no heuristic identity statements at the outset of inquiry. In these early stages, scientists instead ventured claims about correlation, localization, and causal/functional roles. This result—contrary to our initial hypothesis—is both puzzling, since BRF research should be relatively ideal ground for HIT, and worrisome, since the argument for HIT is an induction over (presently few) cases. To be clear, no one thinks that a single disconfirming case or counterexample demonstrates that HIT is incorrect. But what the absence of identity statements does do is warrant doubt and further scrutiny.

4. Re-evaluating HIT

4.1. ESTABLISHING THE OUTSET OF INQUIRY

Can the absence of heuristic identity statements be explained away? One possibility would be to suggest either that they occurred much earlier than the 1950s, or else occurred much later. Neither of these goalpost-moving gambits is likely to work, however.

While some interesting research on pleasure and euphoria was conducted at the end of the 19th century, much of it was conceptual and most scholars demured from asserting identities. For instance, Henry Rutgers Marshall, whose algedonic continuum theory of pain/pleasure eschewed the identification of pleasures with either sensations or brain processes, is representative of both the trends and the knowledge at the time: '[t]here is not the faintest indication, to my knowledge, of the existence of a pleasure centre in the brain' (1892: 634). One obvious reason why is just that midbrain structures are difficult to access, and the methods of the day did not permit deep brain stimulation experiments. Moreover, from the beginning to mid-20th century, much of the relevant research focused on behavioral responses to external rewarding stimuli rather than conative or cognitive experience or neural localization.

As for the second stream, the story of DA does have an interesting prehistory, dating back to 1910–1911. Independently of each other, George Barger & Alfred Ewins (1910), working at the Wellcome Physiological Research Laboratory in London, and Mannich & Jacobson (1910), working at the University of Berlin's Institute of Pharmacology, first synthesized DA. With Barger, Henry Dale (1910) began characterizing its pharmacological effects in cats. A year later, the Polish chemist Casimir Funk (1911) first synthesized DOPA at the Lister Institute of Preventive Medicine in London. Torquato Torquati unwittingly isolated the precursor L-dopa, and his (1913) results were more robustly chemically

[6] In more recent literature in the 1990s and beyond, the computational functions of DA in relation to reward prediction error, incentive salience, and precision have become important research foci (see Colombo (2014) for an overview; see also Colombo & Wright 2017). While constructs like COMPUTATIONAL EQUIVALENCE and ISOMORPHISM are employed to characterize DA activity computationally, they differ from the concept IDENTITY and so cannot just be substituted for it. If this literature vindicates any philosophical frameworks, functionalism would likely come before HIT.

characterized by Markus Guggenheim (1913) at Hoffman LaRoche in Switzerland, who self-experimented with large doses of L-dopa. But despite this burst of research in numerous labs in various countries, there were no attempts to connect up these biochemical results with research about either neural or psychological states, and so no attempts to state psychoneural identities. Indeed, the realization that DA occurs naturally in the brain was still several decades off (Roe 1997).

While it has long been established that the 1950s marked the outset of research into BRF, it is also possible that advocates might instead claim that research into BRF in the 1950s was insufficiently developed to have warranted any scientists making any heuristic identity claims. That maneuver gives the appearance of playing a shell game, however, which shelters HIT from potentially problematic cases. Secondly, in pursuing it, advocates of HIT incur the burden of providing a general answer about the conditions on research maturation. More problematically still, the maneuver may backfire, since cases where identity claims are made only after research has sufficiently developed may be cases that better support the traditional identity theory over HIT.

4.2. CARTESIANISM TO THE RESCUE?

A more plausible response stems from the observation that both Cartesianism about animal minds and behaviorism were paradigmatic in psychology in the 1950s; and either one may have inhibited neuroscientists' willingness to theorize about psychological states or processes as the relata of identity relations.

While the cognitive revolution was in full swing by 1959–1960, this historical context is certainly important not to underlook. However, in so looking, observe that the literature published during this period of serendipitous discovery on non-human animal subjects was perfectly sanguine in ascribing mental states like pleasure, reward, motivation, and emotion. For instance, Brady's (1955) ICSS experiments focused on 'motivational–emotional factors'. For his part, Olds described his experiments with Milner in terms of experimentation on "higher feelings' such as love, fear, pain, and pleasure', and construed ICSS experiments on rats as a psychological method for detecting and measuring positive emotional behavior—pleasure and the satisfaction of specific 'wants' (1956b: 105, 107); moreover, commentators at the time interpreted these descriptions as such: e.g., 'Olds has proposed that the pleasure system works as a positive feedback mechanism […]' (Pribram 1960: 9).

Early 1960s studies of electrical stimulation on human subjects' mental states and processes were also no less sanguine. Not only did they lend support to Olds's initial hypotheses, but also they continued to resort to the language of 'production' and 'correlation'—not identity, nor even equality—to do so.[7] For instance, Heath stated that '[t]he findings reported herein are consistent with other data obtained in our depth electrode studies in man which indicate a correlation between activity of the septal region and the pleasure response' (1972: 15–16). And as Sem-Jacobsen reported, '[f]eeling of ease and relaxation, feeling of joy with smiling, and great satisfaction have been elicited from different areas. (1959: 414). Consequently, it's worth considering whether neuroscientists' willingness to experiment on and theorize about psychological states was inhibited by either Cartesianism about animal minds or behaviorism; but as an explanation for the lack of identity

[7] See, e.g., Sem-Jacobsen (1959); Delgado & Hamlin (1960); Bishop et al. (1963); Heath (1963); Heath & Gallant (1964).

statements at the outset of research, this one provides insufficient cover.

4.3. RETHINKING HIT'S OTHER CASE STUDIES

The absence of identity statements at the outset of inquiry into the nature of BRF, heuristic or otherwise, warrants scrutiny of the claim that HIT captures a widely-used research strategy, as well as a closer look back at the original cases put forward in support of HIT. Upon closer scrutiny, those original cases provide the support for HIT only under a very heterodox conceptualization of identity.

4.3.1. *pFC processing and memory*

One can be a mechanist without being a heuristic identity theorist, just as one can be a heuristic identity theorist without being a mechanist: these two theories are logically independent. So it is interesting that Bechtel (2008b: 78) describes Endel Tulving's neuroimaging research on episodic memory as an 'exemplar of a heuristic identity claim'. Exemplary in what sense? Bechtel's answer is this:

> Tulving and collaborators advanced the hemispheric encoding/retrieval asymmetry (HERA) hypothesis according to which the left pFC is more involved than the right in encoding information about novel events (and in retrieval of information from semantic memory), while the right pFC is more involved than the left in tasks involving retrieval of episodic memories. […] HERA is a bold hypothesis implying that different brain areas are responsible for the operations of encoding and retrieval of episodic memories. It constitutes an important step towards developing a functional decomposition into component operations of the overall mechanism presumed responsible for episodic memory. As is often the case, such initial steps do not settle the matter, but prepare the way for further research. (2008b: 77–78)

As Bechtel presented it, Tulving hypothesized that different areas of the pFC are differentially 'involved in' and 'responsible for' different mnemonic operations. Elsewhere, he suggested that the relation hypothesized by Tulving was the 'linking' relation (2008a: 990). But involvement is not identity; it is not even equality or equivalence. Mutatis mutandis for the linking relation: hypothesizing or even demonstrating that two relata are linked is to state or show little more than that they are co-related. Nor are the predicates 'being responsible for' and 'being identical with' even approximately coextensive, and neither does the former advance beyond a correlation in its implications. Consequently, HIT seems to have a bait-and-switch problem; for where the identity relation should be posited, as would be expected of an identity theorist, we are treated only to the much weaker 'linking' and 'involvement' relations between brain structures and mental operations.

4.3.2. *FFA processing and face perception*

Advocates have recruited another case to support HIT: research on face perception in relation to activity in the fusiform face area (FFA). But, here, too, the same problems recur.

Nancy Kanwisher and colleagues reported a three-part fMRI study that they took to warrant the claim that '[FFA] is selectively involved in the perception of faces'; they further suggested that disconfirmation of alternative hypotheses increases the probability that face perception and recognition is indeed localized in FFA: 'the elimination of these main alternative hypotheses provides compelling evidence that [FFA …] is specifically involved in the perception of faces' (1997: 4302, 4309).

After rehearsing these 'involvement'-claims, McCauley took the liberty of construing them as direct evidence for HIT:

> [a]fter providing additional negative experimental evidence against a few more of the prominent alternative hypotheses, Kanwisher, McDermott, and Chun identify activity in the FFA with the (psychological) task of face perception. Their protestation about 'compelling evidence' notwithstanding, they have advanced a theoretically provocative, cross-scientific, hypothetical identity. (2012: 199)

McCauley's misrepresentation of the scientists' claims is no isolated incident; advocates of HIT have presented a unified front: 'the identity claim, then, is that the process of face recognition is identical to the activity of the FFA' (Burnston et al. 2011: 111). Granted, statements such as 'face recognition processes are identical to FFA processes' or 'the activity of recognizing faces is identical to the activity of the FFA' are indeed identity statements. The trouble is that the statements were asserted, or even invented—not by Kanwisher and her colleagues—but by advocates of HIT themselves.[8]

For their part, the scientists neither explicitly asserted nor tacitly implied or committed themselves to any such identity claims. What they did conclude is the weaker claim that that FFA is selectively involved in the perception of faces. And not only have they yet to publicly endorse HIT, they have consistently articulated this weaker and more cautious conclusion for almost two decades. For example, in their review of the literature of face perception, Kanwisher & Yovel summarized their position as follows: '[i]n this chapter, we described the current state of knowledge about face-selective regions of cortex in humans and their role in face perception. Current evidence supports the hypothesis that the FFA is specifically involved in face perception per se' (2009: 853–854).[9]

4.4. CASTING A BROADER NET

The failure of BRF research to confirm HIT prompted a reappraisal of the cases previously invoked to support HIT; in turn, closer scrutiny of those cases reveals not only that identity statements do not occur in their respective scientific literatures, either, but also that advocates are doing substantial interpretative work in understanding the research.

HIT advocates have not articulated what they count as identities. But Bechtel & McCauley (1999) and McCauley & Bechtel (2001) do motivate HIT as one way of responding to objections against the traditional type-identity theory in the philosophy of mind—more on this in §5—and do emphasize that HIT embeds the principle of the indiscernibility of identicals (i.e., the converse of 'Leibniz's law'). So, while it is not impossible that they have a heterodox conceptualization in mind, the far simpler and more natural interpretation is just that they construe identity classically (i.e., as a dyadic equivalence relation with the usual features: reflexivity, symmetry, transitivity, bijectivity, etc.). But if identity is understood in this way, then HIT suffers from a lack of empirical support.

---

[8] Given that the reputation of HIT is staked on increased sensitivity to actual scientific practice, it's troubling that readers who are unfamiliar the Kanwisher paper are simply just forced to take these HIT claims at face value. If this reputation is bought at the price of ignoring rudimentary metaphysical issues about relations, then it is a price not worth paying.

[9] One reason why Kanwisher, Haxby, Yovel, and other researchers have not asserted the identity of FFA activity and the perceptual processing of faces is that FFA activity was always only part of the neuroanatomical story, and scientists knew this two decades ago. Current understanding suggests that the network is functionally partitioned into ventral and dorsal streams, which are constituted in part by face-selective areas in both the anterior and posterior superior temporal sulcus (aSTS-FA, pSTS-FA), the inferior frontal gyrus (IFG-F), and the anterior temporal and right occipital lobes (ATL-FA, OFA) (Duchaine & Yovel 2015). Because the mechanism(s) enjoy a complex localization through a network of neural areas and pathways, face perception is not identical with FFA activity.

Of course, since its advocates have not specified how they conceive of identity, perhaps what these results reveal is that HIT requires a broader conceptualization of identity—one with non-strict criteria, which encompasses causal and functional relations, plus relations of localization and correlation in addition to equivalence or equality.

On this broader understanding of identity, the cases previously reviewed would provide empirical support for HIT. But it also incurs problems. One is that HIT no longer appears as a genuine alternative to the type-type identity theory, so much as a misleading relabeling of mechanists' strategies for mechanism discovery. Relatedly, the reconceptualization is heterodox at best, and not obviously a theory of identity. And it would fail to distinguish between different kinds of hypotheses advanced during early stages of scientific research, which may play different heuristic roles in scientific discovery, explanation, and confirmation.

Ultimately, then, advocates appear caught in a dilemma: either HIT adequately describes scientific practice but not psychoneural identity, or else HIT adequately describes psychoneural identity but not scientific practice.[10] In the next two sections, we focus on further worries and argue that HIT suffers from conceptual confusion. Meantime, we conclude that the case-based inductive argument for HIT remains weak: at present, there is no chronicle of any cavalcade of heuristic identities.

5. HIT as Type-identity Theory?

5.1. CORRELATION OBJECTION

5.1.1. *Mapping, correspondence, identity*

HIT was motivated, in part, by the need to respond to the so-called 'correlation objection' to the traditional mind/brain identity theory. According to Bechtel & McCauley (1999: 67, 69), the objection is as follows. While neuroscientists can discern and isolate certain brain states or processes that are correlated or associated with certain mental states or processes, there is no possible evidence that could support advancing from correlations to identities that would not also apply, eo ipso, to the correlations themselves. So identity theorists cannot establish the actual identity of neural and psychological states, but only their regular correlation; statements about psychoneural correlations exhaust the factual content of hypothesized identity statements.

In response, Bechtel & McCauley averred that 'mapping at least some mental states (viz., many that figure in scientific psychology) one-to-one with physical states is a perfectly normal part of research in cognitive neuroscience' (1999: 67). And indeed it is. But how does this response address the correlation objection?

For a function $f: M \rightarrow P$ from the domain of (descriptions of) mental states or processes to the codomain of (descriptions of) physical states or processes to be one-to-one, every image of the codomain must be mapped to by at most one argument of the domain. Hence, if $f(x) = f(y)$ implies $x = y$ for all $x, y \in$ M, then indeed, every mental state or process can be injectively mapped with physical states or processes. The mapping need not also be onto; presumably, some physical states (e.g., total persistent vegetative coma) won't themselves backmap to mental states or processes. But neither will it be one-to-one unless every—not just at least some—image of the codomain gets mapped to by each argument x_i.

[10] We thank one referee for helping us frame the problem this way.

Genuine one-to-one mappings might also be insufficient if the kinds of mappings required to warrant type-type mind/brain identity statements have to be invertible. Relatedly, while an identity function $i(x) = x$ has the feature of being one-to-one, it also has the feature of being onto; since solely one-to-one functions don't have that feature, neither will they just be an identity function unless upgraded further. So what's needed is an identity relation, not Bechtel & McCauley's injective 'mapping of at least some mental states'.

Moreover, a mapping that is both one-to-one and onto will give us the correspondence relation. But the correspondence relation differs from identity. For example, in the truth literature, correspondence theorists contend that true propositions correspond to facts; among their arch-rivals are the identity theorists, most of whom take facts just to be true propositions. Just as we would err in confusing their respective theories, we ought to distinguish mind/brain correspondence theories from identity theories. Even if we further upgraded what Bechtel & McCauley had in mind to a bijective mapping, the correlation objection may still be gripping; for even if each mental state or process in the domain were uniquely mapped to and from each physical state or process in the codomain, it would follow that mental and physical states or processes are isomorphic but not yet that the mental and physical states or processes of each ordered pair are indistinguishable with respect to any/all of their features. Isomorphism is not automorphism.

5.1.2. *Establishing identities*

Bechtel & McCauley offer a second response to the correlation objection on behalf of type-identity theory: 'In a perversely Humean spirit, [objectors] set the bar impossibly high, requiring identity theorists to establish each identity claim's truth—in effect—beyond a shadow of a doubt. Discredited in philosophy of science, verificationism, oddly, enjoys new life in philosophy of mind.' (1999: 69).

First, the correlation objection neither entails nor is entailed by verificationism, whether in the philosophies of science or mind or semantics or elsewhere. These are orthogonal ideas. Second, whether the stronger case for any putative identities ultimately always collapses into the weaker case for correlations or correspondences does not pertain to the project of doubt-removal, whether construed in terms of the achievement of epistemic states like knowledge or the occurrence of psychological states like certainty and overconfidence. So the suggestions that the correlation objection either has given new life to verificationism, or implies a requirement to establish beyond a shadow of a doubt the truth of an identity statement, are largely immaterial. In this passage, Bechtel & McCauley have introduced a red herring; and so if there are good responses to the correlation objection, this is not one of them.

Still, one may suspect that spirited Humeans are afoot. Suppose that σ is an identity statement of the form '$Fa = Fb$'. Given that σ only if 'σ' is true, then for S to establish identity statements just is for S to establish truths about the actual identity of those neural and psychological states. Is this, as Bechtel & McCauley complain, a bar too high? Well, any such bar that can be raised by alethic ascent can also be lowered; for the familiar T-schemata sanction a logical equivalence between truth-apt identity statements and explicit ascriptions of truth to them. Given that 'σ' is true only if σ, then for S to establish truths about the actual identity of neural and psychological states just is for S to establish those identity statements. Consequently, and contra Bechtel & McCauley, the correlation objection does not necessarily require identity theorists to establish each identity claim's truth, much less beyond a shadow of a doubt. Rather, the objector can just as well be understood as just challenging identity theorists to establish any identity statement.

5.2. NECESSITY

With their Kripkean counterparts, advocates of HIT also take identity statements to be neither analytic nor a priori; but while synthetic and a posteriori, they deny that identities have their truth-values necessarily: they are neither necessarily true nor necessarily false. Instead, advocates take them to be possibly true but probably false. Of course, familiar arguments take this idea to be a mistake. Given the modal necessity of self-identity, it seems to follow that if any two individuals *a* and *b* are identical then necessarily they are identical:

$(a = b)$	[supposition of identity]
$(\Box(a = a)$	[necessity of self-identity]
$(\Box(a = a) \supset \Box(a = a))$	[redundancy]
$(\Box(a = a) \supset \Box(a = b))$	[substitution]
$(\Box(a = b)$	[MP]
$(a = b) \supset (\Box(a = b))$	[necessity of identity]

The argument supposes an identity, and then uses the necessity of self-identity to validly demonstrate that identities are necessary. If HIT advocates deny the conclusion, likely they must also deny the initial premise—the supposition of contingent identity. After all, to suppose a contingent identity between *a* and *b* is to suppose that '$a = b$' is true; but, pace HIT, '$a = b$' is contingently untrue and merely plays a heuristic role. Yet, if "'$a = b$' is a contingently untrue' implies that $a \neq b$, then the denial of '$a = b$' looks to be the denial that an identity holds—not the acceptance or advancement of a special kind of heuristic identity.[11] So the issue is not easy to waive off. More problematically, if all identities are necessary and heuristic 'identities' are not necessary, then neither are they identities.

5.3. HEURISTIC IDENTITY STATEMENTS AS PREMISES

Scientific explanation and prediction have often been reformulated argumentatively. Accordingly, Bechtel & McCauley take identity statements play the role of premises in explanatory arguments, and take the statements themselves to need no further explanation.[12] One reason why is that scientists cannot be said to come to explanatory knowledge of what is contingently untrue. Consequently, unless scientists infer them from false premises, heuristic identity statements cannot serve as the conclusions of good explanatory arguments. Advocates of HIT are happy to accept the latter disjunct: it is because heuristic identities cannot serve as the conclusions of good explanatory arguments that they must serve as their premises if they participate at all.

But neither is it clear that heuristic identities can serve as the premises of good

[11] To assert σ just is to present σ as being true. If identities are, ex hypothesi, disposable heuristics that are probably untrue, then scientists' utterances of identities are unlikely to have the illocutionary force of assertion, pace HIT. Acceptance and advancement are not normatively constrained in the same way.

[12] See Bechtel & McCauley (1999: 69); McCauley & Bechtel (2001: 737, 753, 756); McCauley (2007: 153); Bechtel (2008: 70); McCauley (2012: 192, 193). It is unclear whether HIT has anti-realist implications; but the presumption of identities being, ex hypothesi, false premises or input that inaugurate a period of scientific research severs the relationship between truth and success at the heart of the abductive inference for scientific realism.

explanatory arguments. Traditional identity theorists can conclude that *a* = *b* from the premises that mental state *a* = the state that has *F* and neural state *b* has *F*; or from the premises that mental state *a* is *F* and neural state *b* is *F*, but that only one state is *F*; or from others still. But whereas traditional identity theorists can generate valid reconstructions of explanatory arguments involving hypothetical identities, heuristic identity theorists cannot just transpose premises and conclusions without imputing invalid reasoning to scientists; for from the premise that *a* = *b*, it does not follow that mental state *a* = the state that has *F*, nor that neural state *b* has *F*, nor their conjunction (even if all true). Likewise, neither does it follow that mental state *a* is *F*, nor that neural state *b* is *F*, nor that only one state is *F*, nor their conjunction.

So, given that premises and conclusions cannot be simply transposed, the thought that hypothetical identity statements are not the conclusions of scientific research but the premises leaves unaddressed—unlike the traditional version—what sorts of conclusions HIT takes to be derived from those identities.

Worse, the claim that identities are premises also threatens to misrepresent the dispute between the traditional and heuristic versions. Various philosophers of mind (e.g., Hill 1991; Kim 1996; McLaughlin 2001) have argued that identity is the best explanation of strong correlations of properties but not vice-versa; but in that case, statements of correlations serve as derived conclusions while identity statements serve as premises in the traditional identity theory no less than the hypothetical identity claims of HIT. Moreover, traditional identity theorists can likewise accept that identities do not themselves require explanations. Yet, as Polger (2004: 32, 37) argues, it is because identities are necessary that no further explanation of them is needed.

6. Does HIT Rest on Confusion?

6.1. EQUIVOCATION

In the caption of the 'locations of function' diagram in Olds's *Scientific American* paper, he wrote, '[t]he labels here identify the centers which correspond to those investigated in the rat' (1956b: 107). Obviously enough, 'identify' here means little more than 'name' or 'denote', and has nothing to do with the sense of identity relevant to any traditional version of identity theory. Yet, in re-examining HIT's celebrated case studies, notice that advocates frequently confuse two kinds of claims by equivocating over 'identity' and 'identity'-esque terms: claims about the 'identity' of states and claims about 'identifying' (or discovering) states. The former concerns a particular kind of relation, identity, while the latter concerns scientists' recognitional abilities to detect and characterize. The former is relevant to evaluating whether HIT is a genuine version of identity theory; the latter is a Trojan horse.

The equivocation works by first focusing on the scientific endeavor of 'identifying' phenomena—i.e., detecting a neural structure or discerning a mental operation—and then by exploiting the nominalization 'identification' to confuse the aforementioned senses of identity relation and detection/recognition abilities. To take one example, Bechtel self-describes the 'approach, which [he and McCauley] have dubbed heuristic identity theory', as 'requir[ing] not just identifying the brain areas associated with cognitive tasks but using such identifications as heuristic guides to understanding how the brain performs those tasks' (2008b: 50; see also Bechtel 2012: 4; McCauley 2012: 199). Here, HIT is described as requiring two things: firstly, that brain areas be discovered and recognized for what they are and

what they do. But again, recognition is one thing, identity something altogether different. Secondly, HIT is described as requiring that brain areas be 'associated with cognitive tasks'. But neither is identity mere association, on any analysis; claims about psychoneural associations or correlations are not intersubstitutable for psychoneural identities.

The confusion of, and equivocation on, these two kinds of claims is no isolated incident.[13] For instance, Bechtel goes on to describe 'recent research that is directed towards identifying the brain structures involved when people perform various memory tasks can play a heuristic role in identifying the mental operations involved' (2008b: 71), which ramifies throughout his discussion of 'ways in which identifying working parts of a neural mechanism can serve a heuristic role in identifying mental operations' (2008b: 76). Of course, it's perfectly reasonable to maintain that scientists who ascertain that some entity is what performs an activity (that π is ψ-ing, that ψ-ing occurs in π, etc.) have thereby performed an identifying act; but identifying acts in this sense of recognition and ascertainment are irrelevant to the debate with traditional identity theorists over whether mental states and processes are type-type identical to neural states and processes.[14] And this irrelevance prompts the worry that the whole of HIT rests on a mistake.

6.2. IDENTITIES AS MERE LOCALIZATIONS

Mechanists maintain that phenomena are scientifically explained in terms of mechanisms, but there is no consensus on how best to characterize the relationship between them. Different mechanists have focused on different kinds of relationships, including those picked out by the phrases *being responsible for*, *producing*, *underlying*, and *maintaining*. None of these relationships amount to a (heuristically-described) identity relationship.

The contemporary analysis of mechanisms is framed using structure versus function and entity versus activity distinctions, applied in both inter- and intralevel contexts.[15] By their very nature, mechanisms and mechanistic systems are both hierarchical and mereological, in at least the sense that they are constituted, inter alia, by part/whole relations between structures at different levels.[16] Construed functionally, a mechanistic activity ϕ-ing is composed of component activities ψ_1, …, ψ_n, each of which is the activity of at least one component entity π_i. Construed structurally, a mechanism Σ is composed of component entities π_1, …, π_n, each of which interacts with other such entities in virtue of performing at least one activity ψ_i. The component parts—whether activities or entities—occur at lower levels than

[13] Nor is it confined to cognitive neuroscience. The confusion between the identity relation, on one hand, and the event or activity of specifying one or more relata, on the other, was applied to cell biology and biochemistry too: 'cytologists using various microscopes, identified various organelles in the cell, whereas biochemists, preparing homogenates and using various assays, identified chemical reactions' (Bechtel & Hamilton 2007: 413).

[14] It also puts philosophers who are looking for an alternative to the standard psychoneural identity-cum-reduction accounts, and who are sympathetic to explanatory pluralism, in a bind. For example, Schouten & Looren de Jong tried to navigate the problem by artfully deploying phrases like 'identificatory connections between […] accounts of vision' (2001: 801) in order to clarify HIT without committing themselves to the mistake.

[15] See, e.g., Bechtel (2002, 2008b); Craver & Bechtel (2007); Schouten & Looren de Jong (2001); Wright & Bechtel (2007); Wright (2007, 2012).

[16] It is important to clarify that the constitutive relevance relation differs from the kinds of identity relations posited by identity theorists. Unlike identity, constitutive relevance relations obtain between a whole and its part, and mechanists employ them for distinguishing relevant from irrelevant factors in a mechanism (see Craver & Bechtel 2007).

do the composite systemic activities and entities constituted by them.

Localization and de-/composition are essential experimental and explanatory strategies in mechanistic science, and statements about them are ubiquitous. As Bechtel reminds, '[i]dentifying parts and operations requires decomposing a mechanism—taking it apart (actually, or in analysis) either structurally into its parts or functionally into its operations. One important part of this task is localizing operations in the parts that perform them.' (2012: 4). While this is an entirely à propos rehearsal of mechanists' doctrine, what advocates of HIT add is that statements about localization and identification are actually clandestine identity statements:

> [l]ocalization claims are identity claims between the parts characterized in terms of their physical constitutions and parts characterized in terms of the operations they are involved in.' (ibidem)

> [s]ince localization claims maintain that it is the same entity that constitutes a particular structure and has performs [sic] a specific operation, they are identity claims in the sense advanced by the mind/brain identity theory.' (Bechtel & Hamilton 2007: 413).

Granted, the normal course of mechanistic science may involve scientists ascertaining that some component activity or function is performed by some component entity or structural part. But it does not follow that localization statements just are identity statements. (The seat of the soul is the pineal gland, says the Cartesian. But localizing the soul thus hardly suffices to unmask our Cartesian as a psychoneural identity theorist.)[17]

First-order localization statements of the form 'π has the activity ψ' or 'the ψ-ing is done by π' have a basic subject-predicate structure. So, too, do second-order localization statements that predicate something to either possessive nominal (i.e., π's ψ-ing) with grammatical landmark/trajector structure or the subject term of a definite description (the ϕ-ing of Σ, etc.). But there is a yawning logical difference between statements of predication and attribution versus statements of identity, and mechanists have no good reason to go along with advocates of HIT in conflating the 'is' of predication with the 'is' of identity.

Moreover, in quantifying over a particular component part π_i as what has property F or what plays the functional role Fx, localization statements still express a simple predication of the form '$\exists x$ such that $Fx \wedge x = \pi_i$'. Yet, none of this implies that operations and activities just are the entities that perform them. To take one example, phasic bursting (ψ-ing) is a basic operation of VTA DA neurons (π) in the mesocorticolimbic system (Grace 1991). But they are cross-categorical classes of phenomena—neither identical nor equivalent—and many other kinds of cells besides VTA DA neurons in many different neural areas besides the mesocorticolimbic system also engage in this kind of signaling. So bursting is localizable to VTA DA neurons, but is not identical with any of them or the property of being one.

Consequently, there is no entailment from localization statements of the form 'π is the locus of ψ-ing' to identity statements of the form '$\pi = \psi$' (Place's assertion that 'consciousness is a brain process' did not mean merely that consciousness is localizable). So there is no good reason for either mechanists or traditional identity theorists to take aboard this additional non-sequitur, though the converse may

[17] To qualify, for Descartes the pineal gland was full of animal spirits and was the interfacing organ between soul and body. Hence, the pineal gland's relation to the soul was not intended to be one of identity.

hold.[18]

6.3. INTRALEVEL IDENTITIES: A DUBIOUS AMMENDMENT

As studied by psychiatrists and clinical psychologists, the mental state or process of anhedonia in schizophrenic patients is a phenomenon residing at a far higher set of levels than the opioid states or processes in their ventral pallida, as studied by neuropharmacologists. According to the traditional mind/brain identity theory, a psychoneural identity statement about the relation of anhedonic experiences and opioid states in the ventral pallida would be an interlevel statement relating personal- and subpersonal-level states.

Advocates of HIT were consistently clear that they, too, conceive of heuristic identities as cross-scientific interlevel statements. For instance, Bechtel & McCauley's (1999: 69; 2001: 752–753) response to the challenge of multiple realizability across species was that articulation of the problem often neglects the issue of 'grain' between levels. More explicitly still, McCauley & Bechtel (2001: 736–737) contended that HIT is expressly designed to accommodate the multilevel character of scientific progress, and that the heuristic identity statements they envision are ones that concern cross-scientific interlevel relationships. They wrote:

> [c]ross-scientific hypothetical identities are perfectly common means for abetting the study of some phenomenon at multiple levels of explanation […] and are heuristics of discovery that inspire multilevel programs of research. […] The theories at each level ascribe distinct properties to the entities and processes that the interlevel, hypothetical identities connect. (2001: 753)

Additionally, Looren de Jong (2006: 439) alluded to the coupling of biophysical research on dipole generators with the electrophysiology of event-related potentials to understand psycholinguistic aspects of speech comprehension and production. Even if allusion to this research does not improve the inductive argument for HIT, Looren de Jong was right to note that claims about the 'multilevel anarchistic level-hopping' seem empirically adequate.

Unfortunately, the misconstrual of localization claims as identity statements has led advocates of HIT to begin asserting that heuristic identities are ultimately intralevel affairs. For instance, Bechtel & Hamilton wrote, '[t]he identity theory is often construed as advancing a reduction of psychology to neuroscience, since neuroscience is at a lower level than psychology. From the point of view of mechanistic explanation, […] accounts of the part of the system and the operation it is performing are at the same level' (2007: 413). Burnston et al. make a similar point in the case of FFA: pace HIT, the psychological function of face recognition is eventually decomposed into more fine-grained subpersonal processes at lower levels, and some of those subpersonal processes are themselves eventually attributed to specific brain areas like FFA; 'the identity statements that end up succeeding are intralevel—they are between different descriptions of phenomena at the same level of complexity' (2011: 113).

True, some activities do occur at the same level as some entities: e.g., action potentials and neurons are activities and entities at the same level. To take another

[18] A common presumption is that the numerical identity of two things x and y explains their being co-located at a time; but co-location is not sufficient to infer that $x = y$. Moreover, recall that heuristic identities are characterized as thoroughly hypothetical empirical statements that are disposable and probably untrue; but the inference from truths about co-location to untrue heuristic identity statements is invalid.

example, 'oligodendrocytes generate myelin' predicates the relational property of myelin-generation of members of the class of oligodendrocytes. Generating myelin is an intralevel affair, in part because it is a diachronic causally productive activity that occurs at the same level that oligodendrocytes occupy. But such activities manifest the usual causal features: asymmetry, non-synchronicity, distinctness of cause and effect, etc. By contrast, identity is a non-causal, symmetrical, reflexive, synchronic relation, and commonly held to be an interlevel affair. Consequently, one of the reasons for thinking that the relevant identities are indeed psychoneural is that they may have features that preclude them from being intralevel.

Of course, with Burnston et al., we grant that descriptions of neural processes may eventually be predicated of specific descriptions of brain areas or parts, after many iterated rounds of decomposing psychological functions into exceedingly fine-grained low-level subpersonal processes. But these are poor grounds for claiming that heuristic identities are localizing intralevel predications. For one thing, the component operations of a neural mechanism are not picked out with psychological vocabulary (e.g., descriptions of myelin-generation are not highly decomposed psychological functions). More importantly, the point of HIT was not to posit successful or true intralevel localization statements about highly decomposed neural phenomena and their neural activities at the end of inquiry. Rather, identities were supposed to be heuristically disposable interlevel psychoneural identity statements, which are made at the outset of research on the basis of scant evidence, and which express the cross-scientific relationship between neural phenomena and undecomposed psychological phenomena across levels. Consequently, the treatment of localization statements as identity statements is doubly problematic if it leads to the claim that heuristic identities are intralevel statements posited near the end of inquiry.

7. Back to mechanistic explanation

What the history of BRF research illuminates is exactly what closer scrutiny of the showcase examples of visual processing, memory encoding, and face perception also reveal: discussion of mechanistic discovery and explanation which over-interprets identity-talk where either there is none or none is needed, and which bottoms out in claims about how de-/composition and localization statements are continually refined in light of new scientific discoveries. What the scientific literature does not reveal is evidence supporting HIT's commitments to heuristic psychoneural identity claims being posited at the outset of research and playing a heuristic role in guiding explanatory projects. Instead, advocates of HIT seem to be confusing identity claims for a pastiche of other claims involving function, causal production, reduction, localization, and property instantiation.

Whereas traditional metaphysics is often detached from the details of scientific practice—relying more on intuition, logic, and conceptual analysis—naturalized metaphysics directly serves it. HIT may be thought of as an exemplar of naturalized metaphysics, which delivers an empirically-grounded alternative to the traditional identity theory in philosophy of mind. Upon examination, however, HIT illustrates adverse aspects of both. Insufficient heed is paid to the textual evidence of the actual science, while logical, linguistic, and conceptual analysis is underutilized. If metaphysics and science are to interact fruitfully, we must combine a sophisticated understanding of scientific theory and practice with a metaphysician's patience for rarified constructs and fine distinctions along with their demand for punctiliousness.

HIT masquerades as an identity theory, a misnomer at best. But it fails to make

contact with the very theory it purports to displace, and it surrenders the competitive advantages that identity theorists previously worked hard to establish. As Polger observed, the idea of 'interim' hypothetical identities is entirely compatible with the thought that 'heuristic identities, in the course of theory building, [are] replaced by more sophisticated realization relations' (2004: 32; see also Schouten & Looren de Jong 2001: 802). For his part, Polger suggested that this compatibility implies that HIT offers no support to the identity theory that doesn't also apply to functionalism in philosophy of mind. Yet, Polger's concern generalizes beyond debates between identity theorists and functionalists; for any theoretical solution to the mind/body problem is compatible with the thought that scientific research begins with disposable heuristic identities, where 'identity' is understood broadly.

Examples like BRF research are more than capable of demonstrating the virtues of the mechanists' program in this regard. For instance, Pribram (1960: 10) presciently redescribed Olds's ICSS findings using a now-standard mechanistic approach, without resorting to talk of identities, heuristic or otherwise. This is what advocates of HIT ultimately want, anyway: evidence of mechanism discovery and mechanistic explanation, understood in terms of functional decomposition and localization. As for identity, that should be left to the identity theorists.

Acknowledgements

The authors thank Peter Fazekas, Raoul Gervais, and two anonymous referees for their close reading and written feedback on the paper. Tom Polger, John Bickle, Trey Boone, Mark Couch, Kevin Morris, Joseph McCaffrey, and other audience members at the 2016 SSPP provided additional feedback. Alexander Beard's research was supported by Cal State Long Beach's Office of Research and Special Programs.

References

Anden NE, Dahlstrom A, Fuxe K, Larsson K, Olson L, & Ungerstedt U (1966). Ascending monoamine neurons to the telencephalon and diencephalon. *Acta Physiologica Scandinavica*, 67: 313–326.

Armstrong DM (1968). *A Materialist Theory of Mind*. London: Routledge & Kegan Paul.

Barger G & Dale HH (1910). Chemical structure and sympathomimetic action of amines. *Journal of Physiology*, 41: 19–59.

Barger G & Ewins AJ (1910). Some phenolic derivatives of β-phenylethylamine. *Journal of the Chemical Society*, 97: 2253–2261.

Baumeister AA (2006). Serendipity and the cerebral localization of pleasure. *Journal of the History of the Neurosciences*, 15: 92–98.

Bechtel WP (2002). Decomposing the mind-brain: a long-term pursuit. *Brain and Mind*, 3: 229–242.

Bechtel WP (2008a). Mechanisms in cognitive psychology: what are the operations? *Philosophy of Science*, 75: 983–994.

Bechtel WP (2008b). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Routledge.

Bechtel WP (2012). Identity, reduction, and conserved mechanisms: perspectives from circadian rhythm research. In S Gozzano & C Hill (eds.), *New Perspectives on*

Type Identity: The Mental and the Physical (43–65). Cambridge: Cambridge University Press.

Bechtel, WP & Richardson, RC (1993/2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Second Edition. Cambridge, MA: MIT Press/Bradford Books.

Bechtel WP & Hamilton A (2007). Reduction, integration, and the unity of science: natural, behavioral, and social sciences and the humanities. In T Kuipers (ed.), *Philosophy of Science: Focal Issues* (377–430). New York: Elsevier.

Bechtel WP & McCauley R (1999). Heuristic identity theory (or back to the future): the mind-body problem against the background of research strategies in cognitive neuroscience. In M Hahn & S Stoness (eds.), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (67–72). Mahwah: Lawrence Erlbaum Associates.

Bechtel WP & Wright C (2009). What is psychological explanation? In P Calvo & J Symons (eds.), *Routledge Companion to the Philosophy of Psychology* (113–130). New York: Routledge.

Bertler A & Rosengren E (1959). Occurrence and distribution of dopamine in brain and other tissues. *Experientia*, 15: 10–11.

Bickle J (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Approach*. Dordrecht: Kluwer Academic Publishing.

Bishop MP, Elder ST, & Heath RG (1963). Intracranial self-stimulation in man. *Science*, 140: 394–396.

Blaschko H (1957). Metabolism and storage of biogenic amines. *Experientia*, 13: 9–12.

Brady JV (1955). Motivational–emotional factors and intracranial self-stimulation. *American Psychologist*, 10: 396.

Brady JV, Boren JJ, Conrad D, & Sidman M (1957). The effect of food and water deprivation upon intracranial self-stimulation. *Journal of Comparative Physiological Psychology*, 50: 134–137.

Bursten B & Delgado JMR (1958). Positive reinforcement induced by intracerebral stimulation in the monkey. *Journal of Comparative and Physiological Psychology*, 51: 6–10.

Burnston D, Sheredos B, & Bechtel WP (2011). HIT on the psychometric approach. *Psychological Inquiry*, 22: 108–114.

Carlsson A (1959). The occurrence, distribution and physiological role of catecholamines in the nervous system. *Pharmacological Reviews*, 11: 490–493.

Carlsson A, Lindqvist M, Magnusson T, & Waldeck B (1958). On the presence of 3-hydroxytyramine in brain. *Science*, 127: 471.

Churchland PM (1998). Betty Crocker's theory of consciousness. In PM Churchland & PS Churchland, *On the Contrary: Critical Essays, 1987–1997* (113–122). Cambridge: MIT Press.

Clavier RM, Fibiger HC, & Phillips AG (1976). Evidence that self-stimulation of the region of the locus coeruleus in rats does not depend upon noradrenergic projections to telencephalon. *Brain Research*, 113: 71–81.

Colombo M (2014). Deep and beautiful: the reward prediction error hypothesis of dopamine. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 45: 57–67.

Colombo M & Wright C (2017). Explanatory pluralism: an unrewarding prediction error for free energy theorists. *Brain and Cognition*, 112: 3–12.

Corbett D, Skelton RW, & Wise RA (1977). Dorsal noradrenergic bundle lesions fail to disrupt self-stimulation from the region of locus coeruleus. *Brain Research*, 133: 37–44.

Couch M (2004). Discussion: a defense of Bechtel & Mundale. *Philosophy of Science*, 71: 198–204.

Craver CF & Bechtel WP (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22: 547–563.

Crow TJ (1972). Intracranial self-stimulation with electrodes in the region of the locus coeruleus. *Brain Research*, 36: 275–287.

Dahlstrom A & Fuxe K (1965). Evidence for the existence of monoamine neurons in the central nervous system. *Acta Physiologica Scandinavica*, 64: 1–85.

Delgado JMR & Hamlin H (1960). Spontaneous and evoked electrical seizures in animals and in humans. In ER Ramey (ed.), *Electrical Studies on the Unanesthetized Brain* (133–158). New York: Hoeber.

Duchaine B & Yovel G (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, 1: 393–416.

Enç B (1983). In defense of the identity theory. *Journal of Philosophy*, 80: 279–298.

Feigl H (1958). The 'mental' and the 'physical'. In H Feigl, M Scriven, & G Maxwell (eds.), *Minnesota Studies in the Philosophy of Science, vol. 2* (370–497). Minneapolis: University of Minnesota Press.

Franklin KBJ (1978). Catecholamines and self-stimulation: reward and performance effects dissociated. *Pharmacology, Biochemistry, and Behavior*, 9: 813–820.

Funk C (1911). Synthesis of dl-3:4-Dihydroxyphenylalanine. *Journal of the Chemical Society*, 99: 554–557.

Grace AA (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*, 41: 1–24.

Guggenheim M (1913). Dioxyphenylalanin, eine neue aminosäure aus vicia faba. *Zeitschrift für Physiologische Chemie*, 88: 276–284.

Heath RG (1954). *Studies in Schizophrenia: A Multidisciplinary Approach to Mind-Brain Relationships*. Cambridge: Harvard University Press.

Heath RG (1963). Electrical self-stimulation of the brain in man. *American Journal of Psychiatry*, 120: 571–577.

Heath RG (1972). Pleasure and brain activity in man: deep and surface electroencephalograms during orgasm. *Journal of Nervous and Mental Disease*, 154: 3–18.

Heath RG & Gallant DM (1964). Activity of the human brain during emotional thought. In RG Heath (ed.), *The Role of Pleasure in Behavior* (83–106). New York: Hoeber.

Hill C (1991). *Sensations*. Cambridge: Cambridge University Press.

Hornykiewicz O (1966). Dopamine (3-hydroxytyramine) and brain function. *Pharmacological Reviews*, 18: 925–964.

Hornykiewicz O (1986). A quarter century of brain dopamine research. In G Woodruff, J Poat, & P Roberts (eds.), *Dopaminergic Systems and their Regulation* (3–18). London: Macmillan.

Hornykiewicz O (2002). Brain dopamine: a historical perspective In G Di Chiara (ed.), *Dopamine in the CNS* (1–22). Berlin: Springer-Verlag.

Kanwisher N, McDermott J, & Chun MM (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17: 4302–4311.

Kanwisher N & Yovel G (2009). Cortical specialization for face perception in humans. In JT Cacioppo & GG Berntson (eds.), *Handbook of Neuroscience for the Behavioral Sciences* (841–858). Hoboken: J. Wiley & Sons.

Kievit RA, Romeijn J-W, Waldorp LJ, Wicherts JM, Wicherts HS, & Borsboom D (2011). Modeling mind and matter: reductionism and psychological

measurement in cognitive neuroscience. *Psychological Inquiry*, 22: 139–157.
Kim J (1996). *Philosophy of Mind*. Boulder: Westview Press.
Koob GF, Balcom J, & Meyerhoff JL (1976). Increases in intracranial self-stimulation in the posterior hypothalamus following unilateral lesions in the locus coeruleus. *Brain Research*, 101: 554–560.
Lippa AS, Autelman SM, Fisher AE, & Caufield DR (1973). Neurochemical mediation of reward: significant role of dopamine? *Pharmacological Biochemistry*, 1: 23–28.
Looren de Jong, H (2006). Explicating pluralism: where the mind to molecule pathway gets off track—reply to Bickle. *Synthese*, 151: 435–443.
Mannich C & Jacobson W (1910). Über oxyphenylalkylamine und dioxyphenylalkylam– ine. *Berichte der Deutschen Chemischen Gesellschaft*, 43: 189–197.
Margules DL (1969). Noradrenergic rather than serotonergic basis for reward in the dorsal tegmentum. *Journal of Comparative and Physiological Psychology*, 67: 32–35.
Marsden C (2006). Dopamine: the rewarding years. *British Journal of Pharmacology*, 147: S136–S144.
Marshall HR (1892). Pleasure-pain and sensation. *Philosophical Review*, 1: 625–648.
McCauley R (1981). Hypothetical identities and ontological economizing: comments on Causey's program for the unity of science. *Philosophy of Science*, 48: 218–227.
McCauley R (1996). Explanatory pluralism and the coevolution of theories in science. In his (ed.), *The Churchlands and their Critics* (17–47). Oxford: Blackwell.
McCauley R (2007). Enriching philosophical models of cross-scientific relations: incorporating diachronic theories. In MKD Schouten & H Looren de Jong (eds.), *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience, and Reduction* (199–223). Oxford: Blackwell.
McCauley R (2012). About face: philosophical naturalism, the heuristic identity theory, and recent findings about prosopagnosia. In S Gozzano & C Hill (eds.), *New Perspectives on Type Identity: The Mental and the Physical* (186–206). Cambridge: Cambridge University Press.
McCauley R & Bechtel WP (2001). Explanatory pluralism and heuristic identity theory. *Theory and Psychology*, 11: 736–760.
McLaughlin B (2001). In defense of new wave materialism: a response to Horgan & Tienson. In C Gillett & B Loewer (eds.), *Physicalism and Its Discontents* (319–330). Cambridge: Cambridge University Press.
Milner P (1991). Brain-stimulation reward: a review. *Canadian Journal of Psychology*, 45: 1–36.
Montagu KA (1957). Catechol compounds in rat tissues and in brains of different animals. *Nature*, 180: 244–245.
Morrison GR (1955). The relation of rewarding intracranial stimulation to biological drive. Masters Thesis, McGill University.
Olds J (1956a). A preliminary mapping of electrical reinforcing effects in the rat brain. *Journal of Comparative and Physiological Psychology*, 49: 281–285.
Olds J (1956b). Pleasure centers in the brain. *Scientific American*, 195: 105–116.
Olds J (1956c). Runway and maze behavior controlled by basomedial forebrain stimulation in the rat. *Journal of Comparative and Physiological Psychology*, 49: 507–512.
Olds J (1958). Self-stimulation of the brain. *Science*, 127: 315–324.
Olds J, Killam KF & Bach-y-Rita P (1956). Self-stimulation of the brain used as a screening method for tranquilizing drugs. *Science*, 124: 265–266.
Olds J & Milner P (1954). Positive reinforcement produced by electrical stimulation

of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, 47: 419–427.

Olds ME & Olds J (1963). Approach-avoidance analysis of rat diencephalon. *Journal of Comparative Neurology*, 120: 259–295.

Olds J & Travis RP (1960). Effects of chlorpromazine, meprobamate, pentobarbital and morphine on self-stimulation. *Journal of Pharmacology and Experimental Therapeutics*, 128: 397–404.

Place UT (1956). Is consciousness a brain process? *British Journal of Psychology*, 47: 44–50.

Polger T (2004). *Natural Minds*. Cambridge: MIT Press.

Polger T (2009). Identity theories. *Philosophy Compass*, 4: 1–13.

Pribram K (1960). A review of theory in physiological psychology. *Annual Review of Psychology*, 11: 1–40.

Roe D (1997). From DOPA to Parkinson's disease: the early history of dopamine research. *Journal of the History of the Neurosciences*, 6: 291–301.

Sem-Jacobsen CW (1959). Depth-electrographic observations in psychotic patients: a system related to emotion and behavior. *Acta Psychiatrica Scandinavica (Suppl.)*, 34: 412–416.

Schouten MKD & Looren de Jong H (2001). Pluralism and heuristic identification: some explorations in behavioral genetics. *Theory and Psychology*, 11: 796–807.

Shapiro L (2000). Multiple realizations. *Journal of Philosophy*, 97: 635–654.

Shapiro L & Polger T (2012). Identity, variability, and multiple realization in the special sciences. In C Hill & S Gozzano (eds.), *New Perspectives on Type Identity* (264–287). Cambridge: Cambridge University Press.

Sidman M, Brady JV, Boren JJ, Conrad DG, & Schulman A (1955). Reward schedules and behavior maintained by intracranial self-stimulation. *Science*, 122: 830–831.

Smart JJC (1959). Sensations and brain processes. *Philosophical Review*, 68: 141–156.

Stein L (1967). Psychopharmacological substrates of mental depression. In S Garattini & M Dukes (eds.), *Antidepressant Drugs* (130–140). Amsterdam: Expecta Medica Foundation.

Stein L (1968). Chemistry of reward and punishment. In DH Efron (ed.), *Psychopharmacology, A Review of Progress: 1957–1967*. (105–123). Washington DC: US Government Printing Office.

Stein L & Wise CD (1969). Release of norepinephrine from hypothalamus and amygdala by rewarding medial forebrain bundle stimulation and amphetamine. *Journal of Comparative and Physiological Psychology*, 67: 189–198.

Stellar E (1957). Physiological psychology. *Annual Review of Psychology*, 8: 415–436.

Sufka K & Lynch MP (2000). Sensations and pain processes. *Philosophical Psychology*, 13: 299–311.

Torquati T (1913). Sulla presenza di una sostanza azotata nei germogli del semi di 'vicia faba'. *Archivio di Farmacologia Sperimentale e Scienze Affine*, 15: 213–223.

Trowill JA, Panksepp J, & Gandelman R (1969). An incentive model of rewarding brain stimulation. *Psychological Review*, 76: 264–281.

Wimsatt W (1976a). Reductionsm, levels of organization and the mind-body problem. In G Globus, I Savodnik, & G Maxwell (eds.), *Consciousness and the Brain* (199–267). New York: Plenum.

Wimsatt W (1976b). Reductive explanation: a functional account. In AC Michalos, CA Hooker, G Pearce, & RS Cohen (eds.), *Boston Studies in the Philosophy of Science, vol. 30: PSA 1974* (671–710). Dordrecht: Reidel.

Wise CD & Stein L (1969). Facilitation of brain self-stimulation by central administration of norepinephrine. *Science*, 163: 299–301.

Wise RA (2008). Dopamine and reward: the anhedonia hypothesis 30 years on. *Neurotoxicity Research*, 14: 169–183.

Wright C (2007). Is psychological explanation going extinct? In MKD Schouten & H Looren de Jong (eds.), *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience, and Reduction* (249–274). Oxford: Blackwell.

Wright C (2012). Mechanistic explanation without the ontic conception. *European Journal of Philosophy of Science*, 2: 375–394.

Wright C (2015). The ontic conception of scientific explanation. *Studies in the History and Philosophy of Science*, 54: 20–30.

Wright C & Bechtel WP (2007). Mechanisms and psychological explanation. In P Thagard (ed.), *Handbook of philosophy of psychology and cognitive science* (39–79). New York: Elsevier.

Zeigler HP (1957). Electrical stimulation of the brain and the psychophysiology of learning and motivation. *Psychological Bulletin*, 54: 363–382.