# Understanding stability in cognitive neuroscience through Hacking's lens[1]

Jacqueline Sullivan

*Abstract*: Ian Hacking instigated a revolution in 20th century philosophy of science by putting experiments ("interventions") at the top of a philosophical agenda that historically had focused nearly exclusively on representations ("theories"). In this paper, I focus on a set of conceptual tools Hacking (1992) put forward to understand how laboratory sciences become stable and to explain what such stability meant for the prospects of unity of science and kind discovery in experimental science. I first use Hacking's tools to understand sources of instability and disunity in rodent behavioral neuroscience. I then use them to understand recent grass-roots collaborative initiatives aimed at establishing stability in this research area and tease out some implications for unity of science and kind creation and discovery in cognitive neuroscience.

*Keywords*: cognition; laboratory science; natural kinds; stability; unity of science

A primary aim of cognitive neuroscience is to understand the neural underpinnings of human cognition. Cognitive neuroscience roughly may be divided into two branches, one which focuses on humans and human clinical populations, and the other, which focuses on non-human animals (e.g., non-human primates, rodents). In this paper, I will be concerned primarily with rodent behavioral neuroscience. Intervention experiments in rodents are crucial for identifying the neural mechanisms that give rise to cognition in humans; rodents afford the possibility of using state-of-the-art techniques to alter genetic, molecular, or circuit-level activity and determine the impact of such manipulations on behavioral performance in tasks designed to assess human relevant cognitive functions. Consider a cognitive function like decision-making, in which an organism has to make a choice between two available actions. A rodent behavioral researcher may design a task to assess decision-making in mice and artificially alter the activity of a population of neurons (e.g., dopamine neu-

[1] The author would like to thank Matteo Vagelli and Marica Setaro for helpful comments on an earlier draft of this paper.


ISSN (print) 2281-8618-ETS doi: 10.4454/philinq.v9i1.346

rons in the medial striatum) *in vivo* as mice perform the task in order to assess the impact of this manipulation on the mouse's performance. The same basic approach may be used to investigate a range of cognitive functions including: *working memory*, *cognitive flexibility*, *attention*, *motivation* and *response inhibition*, to name only a handful. Insofar as rodent behavioral research occurs in laboratories and involves the use of "apparatus used in isolation to interfere" (1992: 34) and, as I will show, to "create new phenomena", it may be regarded as constituting a laboratory science in Hacking's sense.[2]

During the past two decades, a picture has emerged in philosophy of science with respect to how areas of neuroscience directed at understanding the neural underpinnings of cognition, like rodent behavioral neuroscience, make progress (e.g., Bechtel 2008; Craver 2007; Picinnini & Craver 2011). The basic idea is that neuroscientists seek multi-level mechanistic explanations that describe the physical entities/components (e.g., neurons, neural circuits) and activities/processes (e.g., neuronal firing, neurotransmitter release) that bring phenomena of interest (e.g., reward-based learning) about. To take a widely cited example from the philosophical literature, activation of N-methyl-D aspartate receptors in area CA1 of the rat hippocampus is one component in the description of the multi-level mechanism of spatial memory (e.g., Craver 2007). Progress in neuroscience on the mechanistic view occurs as findings from experiments being undertaken in the same and different areas of neuroscience are "seamlessly integrated" into descriptions of multi-level mechanisms of cognitive phenomena (e.g., Picinnini & Craver 2011).

This view of progress in the mind-brain sciences has arisen primarily from the perspective of the philosophy of explanation, in the absence of careful evaluation of the precise kind of knowledge that individual neuroscientific experiments and research studies yield and absent an analysis of how results from different studies actually fit together. In the last two decades of the 20th century, however, Hacking (e.g. 1983; 1991; 1992) urged philosophers of science to relinquish their exclusive focus on "representations" (i.e., theories, explanations) and turn their attention to experiments ("interventions") – those processes by which phenomena are produced or "created" in scientific labo-

[2] Hacking (1992: 37) is not interested in "research at the frontiers of inquiry", which "can be as unstable as you please". Rodent behavioral neuroscience, insofar as it employs cutting-edge tools, is thus not an example of a stable science. My aim in this paper, however, is to use Hacking's tools to identify features of rodent behavioral neuroscience that investigators themselves believe jeopardize the production of stable knowledge – knowledge that they regard as necessary for progress in their field. I aim to show that Hacking's tools may be used to characterize the kind of stable knowledge this area of science lacks and that some scientists working in this area seek. As I see it, Hacking's descriptive conceptual tools may be prescriptively used to understand how to stabilize laboratory sciences (even if he did not intend them to be used in this way).

ratories. In this paper, I use conceptual tools that Hacking put forward to evaluate experimental practices in rodent behavioral neuroscience. I focus primarily on a set of conceptual tools that Hacking (1991; 1992) put forward to understand the stability of the laboratory sciences, which he used to tease out implications for the prospects of unity of science and kind discovery in experimental science.

I begin, in Section 1, by identifying some preliminary features of rodent behavioral neuroscience. The aim is to provide enough detail that Hacking's taxonomic framework for understanding stability in experimental science may later be applied. I then consider Hacking's (1991) claim that disunity is not a temporary feature of science, but, indeed, a permanent aspect of the scientific landscape, because, despite Thomas Kuhn's (1961) claims about revolutions in science, science does accumulate, and some parts do become stable – –a kind of local stability that is antithetical to global unity of science. I then describe the conceptual tools that Hacking (1992) introduced and used in order to understand how laboratory science becomes stable, and I use these tools for two purposes. First, I apply them to characterize the sources of current instability in rodent behavioral neuroscience. Second, I use them to illuminate strategies of stabilization currently being collaboratively implemented in the context of two grass-roots initiatives in this field. I conclude by teasing out implications pertaining to unity of neuroscience and the nature of the kinds that neuroscience, on Hacking's view, is likely to discover.

## 1. *Some preliminary observations about rodent behavioral research*

In neuroscience, rodent behavioral studies may be aimed at understanding species-specific cognition (e.g., mouse cognition), but rodents are more commonly used as models for humans – mammalian "stand-ins" that afford the possibility of combining tasks to assess cognition with cutting-edge visualization (e.g., fiber photometry) and intervention (e.g., optogenetics) technologies for intervening in molecular, cellular, and neural circuit activity to determine the impact of such interventions on cognitive functioning. Mouse models of neuropsychiatric, neurodegenerative and other brain disorders (e.g., concussion) also are regarded as crucial for identifying the neural mechanisms that underlie impairments in cognitive functions such as memory, attention and decision-making that accompany these disorders and developing effective therapeutic interventions to treat them.

One cognitive function that is crucial for navigating the world on a daily basis, and that is impaired in neurodegenerative diseases like Alzheimer's disease, is *pattern separation* – namely, the ability to distinguish memories from

each other, to separate one memory pattern from the next. Consider a simple illustrative example (Beckinschtein, Kent, Ooman, *et al.*, 2013). If you drive a car to school or the office, it is likely that you park your car in a different spot each day. Yet, you typically are able to remember where you parked your car in the parking lot each day, despite parking in a different spot from day to day. This is an instance of pattern separation.

One task that neuroscientists have used to study pattern separation in rodents is the Spontaneous Location Recognition (SLR) task. In one version of this task, a rodent is placed in an open circular arena and allowed to habituate to that environment. Three novel stimuli (different objects) are then introduced. In a simple version of the task, there is a sample phase in which two of the objects are placed 50° apart from each other and the third object is placed equidistant from each of the other two. The mouse is then placed into the arena to explore. In the choice phase of the task, which occurs 24 hours after the sample phase, two novel copies of the two objects that were placed 50° apart during the sample phase are presented. This time, however, one of the two objects is placed in a novel location (a location equidistant between its previous location and the "familiar" location of the other object).[3] Given that mice prefer novelty, a mouse that spends more time exploring the object in the novel location compared to the object in the familiar location divided by the total time it spends exploring is considered to have successfully "pattern separated" – i.e., to have a memory that enables them to distinguish the object in the novel location from the object in the familiar one (See Beckinschtein, Kent, Ooman, *et al.* 2013).

Rodent behavioral neuroscience is an interdisciplinary area of science that brings together investigators hailing from a variety of different fields including: genetics, animal behavior, neurophysiology, biochemistry and computational neuroscience, to name only a handful. Tasks like SLR may be combined with a variety of different visualization and intervention techniques that allow the activity of molecules, cells, and circuits to be detected and manipulated to determine the impact on behavioral performance. Mouse models of neuropsychiatric and neurodegenerative disease and other brain disorders (e.g., concussion) may be used in order to identify disruptions in neural circuit activity that underlie impaired performance on such tasks. There are many different apparatuses (e.g., mazes, open fields, classical conditioning chambers, touchscreen operant chambers (described in section 4)) and tasks (e.g., SLR, contextual fear conditioning, social recognition, paired associates learning) that rodent

[3] As rodents have a keen sense of smell, new, identical objects are used in the choice phase to rule out the possibility that the rodents are using olfactory cues to perform the task.

behavioral researchers may use in combination with intervention techniques to investigate the neural bases of different kinds of cognitive functions.

With this brief introduction to rodent behavioral neuroscience, I turn now to Hacking's views about disunity of science.

## 2. *Disunified sciences*

In "Disunified Sciences" (1991), Hacking identifies and evaluates a set of eleven theses characteristic of logical positivist understandings of the unity of science. I want to briefly consider a relevant subset of these, given that certain aspects of unity of science conceived of by the logical positivists remain implicit in contemporary thinking about progress in areas of neuroscience like rodent behavioral neuroscience. Specifically, recent arguments for unification focus on explanatory integration, which involves the integration of data from multiple experiments into explanations of cognitive functions (e.g., Craver 2007; Craver, Piccinini 2011). Yet, such integration is only possible if the constructs designating cognitive functions under which data from different experiments are being integrated are stable. As I will argue later in the paper, there are good reasons to think they are not.

Among the unity theses that Hacking (1991: 41) considers are two "metaphysical theses": (1) there is a single world, and it contains diverse kinds of phenomena that are (2) "interconnected". The epistemic aim of science is to understand this single world and science offers the best method for attaining such understanding. The logical positivists expressed confidence in the idea that "there is one right fundamental system of classifying everything" (taxonomic thesis), that will be expressed in a single scientific language (e.g., physics) (linguistic thesis) that identifies the stable regularities and tracks so-called "natural kinds" (Hacking 1991: 41). They believed that science gradually approximates towards this one right system by means of intertheoretic reduction (reductionist thesis)-namely, the establishment of bridge laws, as terms in reducing and reduced theories are connected (connectability), and reducing theories come to explain all the phenomena originally explained by the reduced theory (derivability). In the process, unity of science is achieved, as "many facts" are brought "under the wing of one intellectual structure" (Hacking 1991: 41). Moreover, the achievement of unity was not only descriptively accurate with respect to the history of science, but also an "on-going trend" in the heyday of logical positivism (Oppenheim, Putnam 1958).

Hacking aims to demonstrate that none of these unity theses are applicable to contemporary science. I am particularly interested in his arguments against intertheoretic reduction and the discovery of a single system of scientific clas-

sification that tracks natural kinds. Hacking offers two different lines of argument here. First, he notes the sheer diversity in language and methods that we find in contemporary science and the difficulties that heads of academic departments face in trying to unify different areas of science that may generally be classified as, for example, "biological", within a single "super-department" (Hacking 1991: 43). He points out that even Oppenheim and Putnam, who advocated for unity via theory reduction, and described it as an ongoing trend in science, themselves acknowledged certain "incompatible trends" in science that were antithetical to unity. Hacking also points to how "overspecialized" science has become to the extent that "in a quite straightforward sense there is no common language of science, and [. . .] as a matter of practicability, there could not be" (Hacking, 1991: 44). Yet, Hacking, agreeing with philosopher of science Patrick Suppes, does not regard "the irreducible pluralism of languages of science" as an obstacle to "the continued growth of science" (Suppes 1984: 121 as quoted in Hacking 1991: 44).

Hacking's second related strategy for establishing the disunity of science is to argue that much of contemporary science, rather than moving towards theoretical unity, becomes stable within a restricted domain. He relies in part on examples from physics to support this claim. For example, he points to scientists like Sheldon Glashow and Werner Heisenberg who have described Newtonian mechanics and classical quantum mechanics as theories that are not universally true, but "valid in [their] domain" (Hacking 1991: 48). Hacking notes that "the idea of a closed theory with its domain at once suggests disunity: different domains governed by different theories" (Hacking 1991: 48) rather than theory displacement or theory reduction. Hacking insists that from the vantage point of philosophy of scientific experimentation, we encounter a similar kind of stability when we look at laboratory science; "[it] is stable" he claims, "not because there is a domain of experiment, given by nature itself, to which certain theories are true" but "because there is a mutual maturing of types of apparatus, phenomena and theory" (Hacking 1991: 49). Such stability results in disunity, in part, because each laboratory science constitutes its own domain in which "bodies of knowledge" are not discarded but rather "supplement[ed] with new kinds of instruments" (Hacking 1991: 49).

At the end of "Disunified Sciences", Hacking emphasizes the need for a more detailed set of conceptual tools to analyze laboratory sciences and to understand how "experimental stability" emerges. He provides one such set of conceptual tools in "The Self-Vindication of the Laboratory Sciences" (1992), which is the focus of the next section.

## 3. *Hacking's lens: the view from the philosophy of experiment*

Hacking (1992) acknowledges that he regards his thesis about the stability of laboratory sciences as "an extension of [Pierre] Duhem's [coherentist] doctrine, that a theory [shown to be] inconsistent with an observation can always be saved by modifying an auxiliary hypothesis", even a hypothesis about the working of the very instrument used to test the theory (1992: 30). However, according to Hacking, Duhem's framework for understanding what happens in experimental contexts is inadequate because it focuses only on how stability is achieved as representations – "theory and auxiliary hypothesis" are "adjusted to each other". Hacking believes philosophers require a richer and more diverse set of tools for understanding experimentation and the stability of experimental science that can accommodate how much of the knowledge generated by the laboratory sciences is stable and how the devices and practices become permanent fixtures of the scientific landscape. To this end, he puts forward a "taxonomy of elements of experiment which [he claims] are mutually adjusted" or brought into coherence so as "to produce the self-vindicating character of laboratory science" (Hacking 1992: 32). He divides these elements into three categories that are intentionally broad so that each captures a wide range of items: (1) *ideas*, (2) *things* and (3) *marks*. In the rest of this section, I will consider each of these in turn.

Although Hacking sought to shift emphasis in philosophy of science away from theories and towards experiments, he did recognize the role that "representations" play in experimental contexts. The category of "Ideas" includes those empirical questions about a phenomenon of interest that an experiment is designed to answer. For example, is activation of a specific population of neurons necessary for spatial memory or visual associative learning? Questions may also be directed at the merits and failings of large-scale scientific theories, which is particularly common in areas of science like physics, but uncommon in areas like rodent behavioral neuroscience. Background knowledge or background beliefs on which an investigator relies, which may be neither systematized nor made explicit also fall into Hacking's *ideas* category. Background beliefs could range from an investigator's understanding of a concept such as *spatial memory*, to her understanding of how a given intervention or visualization technique (e.g., optogenetics) works, to her assumptions about potential confounds to be controlled for during an experiment (e.g., feeding times for a rat subject when successful task performance requires hunger as motivation). "Ideas" also include high-level "systematic theories" about the subject matter under study and "topical hypotheses" that are local to experimental contexts and connect together theoretical ideas with the implementation of those ideas

in the laboratory in a way that is revisable. A final element within the category of ideas involves the understanding on the part of the investigator as to the nature and structure of the apparatus (e.g., task analysis) or tools that are used to produce data and how those tools actually work.

Hacking's second category, "things" – includes all of the material elements involved in an experiment such as: the targets of investigation (e.g., mice and rats, cells, molecules, synapses) and the instruments or apparatuses (e.g., optogenetic techniques) used "to alter or interfere" with those targets (Hacking 1992: 46). The instruments that serve a productive function, for Hacking, insofar as they are used to "create phenomena" (e.g., Hacking 1983; 1992) differ from those instruments that are used to detect the effects of the intervention – to "determine or measure the result of the interference or modification of the target" (Hacking 1992: 47). The broader category of "*tools*" consists of "all the humble things upon which the experimenter must rely" in order to run the experiment – for example, microtomes for slicing tissue samples, artificial cerebrospinal fluid for preserving brain tissue samples, or the computer equipment and software for running a given cognitive task. Finally, *Data generators* are the parts of the experiment that generate the data (Hacking 1992: 48), such as movement tracking devices and reaction-time software–all of the programs that record data, including scientists recording data by hand.

Hacking's final category, "marks and the manipulation of marks" is intended to include the outputs of experiments – the data – as well as those processes to which the data are subject. In order to be interpretable, data must be reduced and analyzed statistically. Yet, Hacking remarks that it is important to remember that choice of data reduction, data enhancement and data analysis techniques are often influenced by "ideas" on the part of investigators including background knowledge and theoretical commitments. The final interpretation of the data is also done in light of the researcher or research team's background knowledge, understanding of how the apparatus and other tools used in the experiment work and, where relevant, high-level theory. Hacking claims that an important part at this stage of the laboratory work is an "estimation of systematic error, which requires explicit knowledge of the theory of the apparatus –and which has been too little studied by philosophers of science" (Hacking 1992: 49). Since the publication of Hacking's paper, a number of philosophers of science have sought to fill this gap (e.g. Mayo 1991; 1996; Sullivan 2018; Schickore 2005; 2019).

According to Hacking, the stability of a laboratory science is gradually established as these 15 elements falling into the broader categories of "thoughts, actions, materials, and marks" are "mutually adjusted to each other" and "what meshes (Kuhn's word) is at most a network of theories, models, approxima-

tions, together with understandings of the workings of our instruments and apparatus" (Hacking 1992: 30). Laboratory sciences become self-vindicating on Hacking's picture, insofar as eventually, "any test of theory is against apparatus that has evolved in conjunction with it – and in conjunction with modes of data analysis" (1992: 30).

Importantly, laboratory scientists have to engage in strategies of stabilization that bring these different elements into consilience. Although Hacking does not acknowledge it explicitly, laboratory sciences do not consist of a single laboratory running experiments in isolation, but investigators – research teams – running experiments in many different laboratories. The stability of experimental science that Hacking describes is thus not something that comes about in a single laboratory, but rather, across many different laboratories having investigators who share *thoughts*, *actions, materials*, *marks and strategies for manipulating marks* in common and who are collaboratively united in bringing these elements into productive symbiosis.

As I aim to show in the next two sections, Hacking's taxonomy of elements and views about the stability of laboratory science may be used as a foil for understanding why instability may occur in some laboratory sciences, not merely due to the fact that these sciences are on the cutting-edge, but also that researchers in the field may be engaged in practices that effectively destabilize the field insofar as their actions are not directed at bringing these elements into consilience. In such instances of instability, we may anticipate a lack of conceptual, methodological and explanatory unity within these fields. Also, in light of Hacking's framework, the possibility that a given laboratory science may stabilize in any number of ways depending upon who the actors are, and what ideas, actions, materials and marks they aim to bring into consilience is consistent with local unity, but as Hacking (1991) indicates, global disunity.

In the next section (Section 4), I use Hacking's framework to identify those aspects of experimental practice in rodent behavioral neuroscience that have served to promote the instability of the field and have been a barrier to the production of stable knowledge pertaining to the neural underpinnings of rodent cognition. The kind of instability that we encounter here is consistent with what might be regarded as counterproductive disunity. Yet, if Hacking is correct, there is such a thing as productive disunity – and it correlates with areas of science implementing strategies to arrive at stable knowledge – strategies that simultaneously result in the creation of phenomena, and the development of specialized languages and methods and associated practices that co-evolve and become "self-vindicating".

## 4. *Rodent behavioral neuroscience through Hacking's lens*

In Section 1, I briefly described some basic features of the structure of experiments in rodent behavioral neuroscience. I now want to elaborate on the structure of research in this field and evaluate it by way of Hacking's framework of "ideas, things, and marks".

First, consider Hacking's concept of "ideas", a category which includes empirical questions about phenomena of interest, high-level theories, background assumptions, topical hypothesis that relate theories to observations made in experimental contexts and beliefs about how a given experimental apparatus or tool works. As I mentioned in Section 1, researchers working in rodent behavioral neuroscience hail from a variety of different research traditions and theoretical backgrounds (e.g., animal psychology, neurophysiology) and have different technical expertise (e.g., expertise in assessing animal behavior or skill using *in vivo* circuit techniques). Given such differences, they do not necessarily agree about how to define terms typically used to designate cognitive functions (e.g., attention, working memory and motivation) and each field "contributes a distinctive vocabulary of terms and acronyms, all embedded to some degree or another in zeitgeists and conceptual frameworks" (Roediger, Dudai, and Fitzpatrick 2007: 1).

Although we do not encounter high-level theories in rodent behavioral neuroscience, researchers do have background assumptions about phenomena of interest, assumptions about what kinds of apparatus and tools are appropriate for addressing their empirical questions, theoretical understandings that inform the development of cognitive tasks and the use of intervention techniques as well as their understanding of how the tasks and tools they use actually work. Yet, differences in theoretical backgrounds, training and technical expertise across the field correlate with differences across researchers with respect to all of the different kinds of "ideas" that Hacking itemizes.

We encounter similar diversity with respect to Hacking's category of "things"; a number of different tasks may be used to study the "same" cognitive function, and not only do investigators differ with respect to what they regard as the most appropriate task or apparatus, but even when they use the same tool to investigate the same function, it is not uncommon for them to vary overall features of the task (e.g., stimuli, intertrial intervals) slightly (e.g. Sullivan 2009). Researchers also have different intuitions with respect to which tasks are most *appropriate* for measuring which functions and are granted the freedom to use those tasks and task parameters they deem most suitable for achieving their investigative aims, just so long as they provide good reasons for their choices from the perspective of peer review.

Differences in training also may impact the design and implementation of rodent behavioral experiments. For example, an expert in rodent behavior may be privy to aspects of an experimental design that may impact the behavioral performance of a mouse in a cognitive task (e.g., over-handling of the animal during different phases of the experiment) and confound the establishment of causal relationships between neural activity and behavior. They thus may modify aspects of the experimental protocol or specific task parameters with the aim of eliminating these confounds. In contrast, a researcher who is an expert in using neurophysiological techniques may be concerned with a different set of potential confounds having to do with consequences downstream of a pharmacological intervention. Such potential differences in epistemic standards that correlate with differences in expertise may thus exist. However, it is widely recognized that such methodological differences may result in differences in findings across laboratories purportedly investigating the same cognitive function (See for example, Crabbe, Wahlsten, Dudek 1999; Graybeal, Bachu, Mozhui *et al.* 2014; Sullivan 2009). This means that findings from multiple different research studies purportedly investigating mechanisms of the same phenomena cannot readily be integrated into unified explanations of common phenomena. And yet, discovering the neural mechanisms of cognition is not something that can take place in a single lab or in the context of a single research study. It requires contributions from many laboratories, not only to produce piecemeal findings about components of the neural mechanisms that give rise to a given cognitive function, but also to reproduce findings across laboratories (Beraldo, Palmer, Memar *et al.* 2019; Button, Ioannidis, Mokrysz, *et al.* 2013).

With respect to Hacking's category of "marks", researchers working in different laboratories also may use a variety of different tools for collecting, analyzing, and interpreting data, and employ different strategies to probe for and reduce error. Choices about which data analysis tools to use, what kinds of errors to probe and control for also vary with respect to one's training and technical expertise. An additional issue is that experiments in rodent cognitive neuroscience combine tools for assessing cognition with state-of-the-art visualization and/or intervention technologies. Yet, the error characteristics, especially of newer intervention and visualization technologies (e.g., optogenetics (Sullivan 2018), may not yet be known. A final and related issue is the lack of emphasis on the development of behavioral experiments that carefully individuate psychological functions involved in task performance and insure that the criterion of construct validity – that a given cognitive task actually measures the cognitive function it is intended to measure – is met prior to moving to experiments directed at identifying the neural underpinnings of these

functions (e.g. Krakauer, Ghazanfar, Gomez-Marin, *et al.* 2016; Niv 2020). There are thus epistemic blind spots in rodent behavioral neuroscience that are obstacles to the field advancing an understanding of the neural underpinnings of psychological functions.

Given the aforementioned observations, there is no sense in which the relationship between "ideas, materials, marks and [the] manipulation of marks" that we encounter in contemporary rodent behavioral neuroscience is stable, nor is the field on a trajectory towards stability. Yet, instability of the kind we find here is regarded by some neuroscientists themselves (i.e., those that I have cited in this section) as a barrier to progress in their field. Particularly in translational areas of cognitive neuroscience, in which the aim is to develop effective therapeutic interventions to treat neuropsychiatric and neurodegenerative disease-related cognitive impairments, the importance of reproducibility and the gradual coordinated accumulation of stable knowledge is regarded as essential for progress. In recent years, large-scale and smaller-scale collaborative grass roots initiatives have emerged with an eye towards stability of the kind Hacking describes. I turn now to analysis of these initiatives.

## 5. *Recent developments in rodent behavioral neuroscience through Hacking's lens*

In the first two decades of the 21st century, several large-scale initiatives were established in order to accelerate the discovery of novel therapeutic interventions to treat cognitive impairments in neuropsychiatric and neurodegenerative disease. Representative examples include the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRICS) initiative (e.g. Carter and Barch 2007; Moore, Geyer, *et al.* 2013), NEWMeds (e.g. Stensbøl and Kapur 2015), and the US National Institute of Mental Health's Research Domain Criteria Project (NIMH RDoC) (e.g., Insel, Cuthbert, Garvey *et al.*, 2010; Cuthbert & Kozack 2013). Each of these initiatives have brought together rodent behavioral neuroscientists, clinical researchers, cognitive neuroscientists working with humans and/or animal models, systems neuroscientists and members of the pharmaceutical industry with the aims of (1) developing more representative mouse models of neurodegenerative and neuropsychiatric diseases ("things"), (2) identifying a set of collaboratively agreed-upon psychological constructs corresponding to functions regarded as impaired in these diseases, ("ideas") (3) improving tools for assessing cognition in humans and mice ("things" and "manipulation of marks"), and (4) increasing the similarity of tools used for the behavioral assessment of cognitive functions across researchers and species ("things").

One way to understand the aims of these initiatives is to develop *stable* knowledge about the neural underpinnings of cognition and disruptions in neural circuitry that underlie these impairments in order to identify those circuits that may be targeted for therapeutic intervention. The measures that researchers involved in these initiatives regard as essential to these goals, are to develop a shared set of theoretical constructs (e.g., cognitive control, working memory) and types of apparatus/tasks (e.g., the Jitter orientation visual integration task (JOVI)) that are to be standardized across researchers working with human subjects and animal models, as well as a shared set of materials (e.g., apparatus, tasks, mouse models of disease) that are to be used in the drive to identify novel targets for therapeutic intervention. As Hacking claims, data interpretation relies on an investigator's background assumptions and theoretical commitments. Insofar as investigators involved in these initiatives are committed to a discreet set of theoretical constructs and general definitions of those constructs, the hope is that there will be some degree of consensus in how to interpret the data arising out of human and animal research. Thus, these initiatives are at least in theory aiming for coherence among Hacking-like elements – concepts, materials, and marks – that are disunified in cognitive neuroscience more generally.

These large-scale government supported research initiatives are on-going, however, to date, they have not produced stable knowledge or major advances in our understanding of cognition and cognitive dysfunction. While a number of reasons may be cited – clearly this is research on the cutting-edge and it is still early days – but one feature that such initiatives lack is an infrastructure to facilitate the stabilization of "ideas, things, and marks" across research groups and laboratories. It is one thing to point to changes that need to be made to experimental practice to facilitate progress and the production of stable knowledge and another thing for researchers to collaboratively implement these stabilization strategies across laboratories to achieve these goals.

The recent development of more grass-roots collaborative initiatives in rodent behavioral neuroscience (e.g., Beraldo, Palmer, Memar, *et al.* 2019; Dumont, Salewski, Beraldo, *et al.* 2020; Sullivan *et al.* 2020) and systems and computational neuroscience (with a focus on rodent behavioral research) (e.g., International Brain Laboratory 2017; Wool 2020) to accelerate discovery in these fields is suggestive that some researchers believe that achieving stability with respect to "ideas, things and marks" requires an unprecedented level of coordination across labs and research groups and an infrastructure similar to that found in other areas of science that have achieved stability historically, including physics and genomics (International Brain Laboratory 2017; Beraldo, Palmer, Memar *et al.* 2019). My aim in the rest of this section is to briefly evaluate these two grass-roots initiatives through Hacking's lens.

The first such initiative I want to consider has emerged around a novel platform, the *Mouse Translational Research Accelerator Platform (MouseTRAP)* (Sullivan *et al.* 2020). Spearheaded by researchers at Western University, MouseTRAP is centered on a touchscreen cognitive testing system for rodents, the Bussey-Saksida touchscreen system (e.g., Bussey, Muir, Robbins 1994; Bussey, Holmes, Lyon, *et al.* 2012; Bussey, Rothblat, Saksida 2001). The system consists of an operant chamber with a touchscreen upon which visual stimuli are presented. Rodents are trained and tested on different cognitive tasks using these visual stimuli and are required to respond directly to the visual stimuli with nose-pokes. Correct choices are rewarded with a drop of strawberry milkshake or a food pellet. There are currently over 20 different rodent touchscreen-based tasks for assessing cognitive functions in rodents ranging from working memory to cognitive flexibility to decision-making. The tasks are fully automated, ensuring the accuracy of task parameters and measures, and infrared beams and video tracking devices are used to monitor an animal's behavior while it performs in the apparatus. These features make the testing system and associated tasks readily standardizable across laboratories, allowing researchers all over the globe to use the same apparatus, stimuli, task parameters, appetitive rewards and data production and data analysis techniques.

In order to increase the reproducibility of rodent behavioral research and in response to increasing demand for the technology, the Bussey-Saksida touchscreen system was commercialized in 2009. Bussey, Saksida and colleagues published three invited papers in *Nature Protocols* (e.g., Horner, Heath, Hvoslef-Eide, *et al.*) with step-by-step instructions on how to prepare animals for training in the apparatus, how to pretrain and train the animals and how to analyze the behavioral data. As of December 2020, over 300 different research groups in more than 200 research institutes in at least 26 countries are using the touchscreen technology (Dumont, Salewski, Beraldo 2020). In 2018, two novel Open Science platforms were established to facilitate pre-publication knowledge-sharing (touchscreencognition.org) and data-sharing (mousebytes.ca) among members of the rodent touchscreen community. A primary aim of these Open Science platforms is "to create a community of scientists who share common methodology and are united in the goals of increasing methodological transparency and improving the reliability and reproducibility of research findings" (Sullivan, Dumont, Memar *et al.* 2020: 10).

If we consider MouseTRAP from the vantage point of Hacking's taxonomy of elements of experimental science, it possesses those features that lend themselves to the development of stable science – efforts are in fact being made to ensure that researchers share a common methodology for conducting research into the neural underpinnings of cognition, that they share ideas – for exam-

ple, empirical questions directed at specific phenomena (e.g., cognitive functions and impairments), topical hypotheses that relate specific understandings of those phenomena to what is observed in the laboratory, an understanding of how the apparatus works in the collection and production of data. They also share "things" in common – the targets of investigation (e.g., rodents, mouse models of disease), how to prepare those targets (as specified in the published protocols, and standardized operating procedures that are available on touchscreencognition.org), the touchscreen operant chamber itself and the tools (e.g., video-tracking devices) used to collect data. Those researchers who elect to use the methodology also share "marks and the manipulation of marks" – techniques of data assessment and analysis in common, and they are also at liberty to take advantage of Open Science platforms that allow them to share their knowledge, input and visualize their data and integrate and compare their data with data from other laboratories using the same methodology. MouseTRAP is suggestive of the fundamental role that scientists themselves must play to collaboratively produce stable science as Hacking conceives of it.[4]

Another notable collaborative grass-roots initiative is the International Brain Laboratory (IBL 2017; Wool 2020). It consists of ~80 researchers from 22 experimental and theoretical laboratories across the globe who are collaboratively aiming to identify the neural basis of decision-making. These researchers are using a standardized "steering-wheel task for head-fixed mice" in order to identify those brain areas that are involved in "decisions" made on the basis of "visual perception" and "history of reward" (IBL 2017: 1213). Using the same behavioral task across 22 laboratories, researchers in each laboratory will "record from many different brain areas" during task performance "using multiple recording modalities to build up a dense dataset of activity measurements during the task" (IBL 2017: 1213). These datasets will then be analyzed using computational techniques in order to understand how multiple brain regions interact during this task. IBL was developed because of the observed success of "team science" in other areas of science including physics and genomics. Moreover, "a critical IBL mandate is to ensure that theory and experiment converge at the ground level, and perpetually throughout [the] scientific process" (Wool, International Brain Laboratory 2020: 105).

IBL emphasizes the importance of bringing Hacking's elements of stability into a kind of consilience. The community seeks to ensure that members share theoretical and background knowledge, the same physical materials and

[4] I have referred to such collaboration as "coordinated pluralism" (2018). Knorr Cetina's (1999) concept of "epistemic culture" and Ankeny and Leonelli's concept of "repertoire", also may be used to shed interesting and important light on how stability or stable knowledge are collaboratively achieved in science.

tools and the same data production and data analysis techniques. They even emphasize the importance of "stabiliz[ing] large-scale collaborative science in traditional academia" in order to achieve the goal of "understanding the neural computations that support decision-making" (IBL 2018: 1213).

One way to conceive of these grass-roots initiatives is that they regard the accumulation of knowledge of the neural underpinnings of cognition to require what I described in Section 3 as "productive disunity". Such disunity involves the collaborative breaking off of smaller groups of investigators from how practice in a given area of science is traditionally done, in instances in which sticking with tradition involves "counterproductive disunity" that is antithetical to progress. It is an interesting question whether laboratory sciences like those Hacking (1992) uses as a basis for understanding stability began with small-scale collaborative revolutions much like these ones.

## 6. *Conclusion*

I want to end by teasing out some implications of my analysis for the unity of neuroscience and say something briefly, from the perspective of Hacking's lens, about the kinds we are liable to encounter in rodent behavioral neuroscience if such grass-roots initiatives are successful.

First, it is relevant to note that the experimental apparatuses at the heart of both of these initiatives satisfy Hacking's condition that laboratory sciences "create new phenomena". Nowhere in the world (as far as I know), except in laboratories that use rodent operant touchscreens, do we encounter rodents interacting with and engaging in cognitive tasks with computer touchscreens. Similarly, we do not encounter head-fixed mice out and about in the world turning steering wheels in response to visual stimuli. The kinds of cognitive functions under study using these apparatuses are created in laboratories. This does not make them any less real, but it is important to recognize the precise type of workmanship that goes into creating them (e.g. Boyd 2000). Moreover, if these small-scale initiatives are ultimately successful, they may yield what might be dubbed "coordinated kinds" (Mattu and Sullivan in press) – the result of the concerted alignment of conceptual and methodological practices across discrete research groups with respect to "ideas, things, and marks". To the extent that different such research groups emerge in cognitive neuroscience and are successful, organizing their practices around discrete sets of concepts, apparatus, tools, and data, we might imagine a plurality of discrete taxonomies of cognitive kinds that are stable but isolated from each other – a kind of "promiscuous realism" (Dupre 1993).[5]

[5] Thanks to Muhammad Ali Khalidi for this characterization – an idea to be worked out on an-

Second, insofar as the creation of these phenomena and investigations into their mechanisms are to be collaboratively subserved by small groups of researchers, and if such collaborations are successful in bringing about a kind of local stability – the kind of findings such research groups make about neural mechanisms are likely to be domain-specific – specific, for example, to those "ideas, things, and marks" that these groups collaboratively bring into consilience to achieve stable knowledge. This is consistent with Hacking's (1991) idea that successful stability is consistent with disunity – that it actually requires disunity – it requires a kind of isolation of a domain from factors that are antithetical to its stability.

On a final note, Hacking would likely be skeptical that these collaborative initiatives, even if they can yield stable knowledge about the neural mechanisms of cognition in rodents, will ultimately shed light on the mechanisms of human cognition, because "human kinds" are "unstable" in ways that make them unamenable to experimental control (e.g. Hacking 1995; 1999). Partially for reasons of space, evaluating and responding to such skeptical concerns will have to be saved for another occasion.

Jacqueline Sullivan
jsulli29@uwo.ca
University of Western Ontario

## *References*

Ankeny, Rachel, Sabina Leonelli, 2016, "Repertoires: A post-Kuhnian perspective on scientific change and collaborative research", in *Studies in History and Philosophy of Science*, 60: 18-28.

Beraldo, Flavio, Daniel Palmer, Sara Memar, *et al.*, 2019, "MouseBytes. An open-access high-throughput pipeline and database for rodent touchscreen-based cognitive assessment", *eLife*, 8: e49630.

Bekinschtein, Pedro, Brianne Kent, Charlotte Oomen *et al.*, 2013, "BDNF in the Dentate Gyrus is required for consolidation of "Pattern-Separated" memories", in *Cell Reports*, 5: 759-768.

Bechtel, William, 2008, *Mental Mechanism: Philosophical Perspectives on Cognitive* Neuroscience, Routledge, New York.

Boyd, Richard, 2000, "Kinds as the workmanship of men: realism, constructivism, and natural kinds", in Julian Nida-Rümelin, ed., *Rationalität, Realismus, Revision*, De Gruyter, Berlin: 52-89.

Bussey, Timothy, Andrew Holmes, Louisa Lyon, *et al.*, 2012, "New translational assays

other occasion.

for preclinical modelling of cognition in schizophrenia: The touchscreen testing method for mice and rats", in *Neuropharmacology*, 62, 3: 1191-1203.

Bussey, Timothy, Janice Muir, Tim Robbins, 1994, "A novel automated touchscreen procedure for assessing learning in the rat using computer graphic stimuli", in *Neuroscience Research Communications*, 15, 2: 103-110.

Bussey, Timothy, Lisa Saksida, Lawrence Rothblat, 2001, "Discrimination of computer-graphic stimuli by mice: A method for the behavioral characterization of transgenic and gene-knockout models", in *Behavioral Neuroscience*, 115, 4: 957-960.

Button, Katherine, John Ioannidis, Claire Mokrysz, *et al.*, 2013, "Power failure: Why small sample size undermines the reliability of neuroscience", in *Nature Reviews Neuroscience*, 14: 365-376.

Carter, Cameron, Deanna Barch, 2007, "Cognitive neuroscience-based approaches to measuring and improving treatment effects on cognition in schizophrenia: the CNTRICS initiative", in *Schizophrenia Bulletin*, 33, 5: 1131-1137.

Craver, Carl, 2007, *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford University Press, Oxford.

Crabbe, John, Douglas Wahlsten, Bruce Dudek, 1999, "Genetics of mouse behavior: interactions with laboratory environment", in *Science*, 284: 1670-1672.

Cuthbert, Bruce, Michael Kozack, 2013, "Constructing constructs for psychopathology: the NIMH research domain criteria", *Journal of Abnormal Psychology* 122(3): 928-937.

Duhem, Pierre, [1906] 1954, *The Aim and Structure of Physical Theory*, Princeton University Press, Princeton.

Dumont, Julie, Ryan Salewski, Flavio Beraldo, 2020, "Critical mass: the rise of a touchscreen technology community for rodent cognitive testing", in *Genes, Brain and Behavior*: e12650.

Dupré, John, 1993, *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*, Harvard University Press, Cambridge MA.

Graybeal, Carolyn, Minusa Bachu, Khyobeni Mozhui, *et al.*, 2014, "Strains and stressors: An analysis of touchscreen learning in genetically diverse mouse strains", in *PLOS ONE* 9: e87745.

Hacking, Ian, 1999, "Making up people", in Mario Biagioli, ed., *The Science Studies Reader*, Routledge, New York: 161-171.

Hacking, Ian, 1995, "The looping effects of human kinds", in Daniel Sberber *et al.*, eds., *Causal Cognition: A Multidisciplinary Debate*, Clarendon Press, Oxford: 351-383.

Hacking, Ian, 1992, "The self-vindication of the laboratory sciences", in Andrew Pickering, ed., *Science as Practice and Culture*, University of Chicago Press, Chicago: 29-64.

Hacking, Ian, 1991, "Disunified sciences", in R.J. Elvee, ed., *The End of Science?*, University of America Press, Lanham.

Hacking, Ian, 1983, *Representing and Intervening*, Cambridge University Press, Cambridge.

Horner, Alexa, Christopher Heath, Martha Hvoslef-Eide, *et al.*, "The touchscreen operant platform for testing learning and memory in rats and mice", in *Nature Protocols*, 8, 10: 1961-1984.

Insel, Thomas, Bruce Cuthbert, Marjorie Garvey, *et al.*, 2010, "Research domain criteria (rdoc): toward a new classification framework for research on mental disorders", *American Journal of Psychiatry*, 167, 7: 748-751.

International Brain Laboratory, 2017, "An international laboratory for systems and computational neuroscience", *Neuron*, 96, 6: 1213-1218.

Knorr Cetina, Karin, 1999, *Epistemic Cultures: How Science Makes Knowledge*, Harvard University Press, Cambridge MA.

Kuhn, Thomas, 1962, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.

Kraukauer, John, Asif Ghazanfar, Alex Gomez-Marin, *et al.*, 2017, "Neuroscience needs behavior: correcting a reductionist bias", in *Neuron*, 93, 3: 480-490.

Mattu, Jaipreet and Jacqueline Sullivan, (in press), "Classification, kinds, taxonomic stability and conceptual change", in *Agression and Violent Behavior.*

Mayo, Deborah, 1996, *Error and the Growth of Experimental Knowledge*, University of Chicago Press, Chicago.

Mayo, Deborah, 1991, "Novel evidence and severe tests", in *Philosophy of Science*, 58, 4: 523-552.

Moore, Holly, Marjorie Geyer, Cameron Carter, *et al.*, 2013, "Harnessing cognitive neuroscience to develop new treatments for improving cognition in schizophrenia: CNTRICS selected cognitive paradigms for animal models", in *Neuroscience Biobehavioral Reviews*, 9: Pt B:2087-2091.

Nagel, Ernest, 1961, *The Structure of Science: Problems in the Logic of Scientific Explanation*, Harcourt, Brace & World, New York.

Niv, Yael, 2020, "The primacy of behavioral research for understanding the brain", *PsyArXiv*: October 22; DOI: 10.31234/osf.io/y8mxe.

Oppenheim, Paul and Hilary Putnam, 1958, "Unity of science as a working hypothesis", in *Minnesota Studies in the Philosophy of Science*, 2: 3-36.

Piccinini, Gualtiero, and Carl Craver, 2011, "Integrating psychology and neuroscience: Functional analysis as mechanism sketches", in *Synthese*, 183, 3: 283-311.

Roediger III, Henry, Yadin Dudai, Susan Fitzpatrick, 2007, *Science of Memory: Concepts*, Oxford University Press, Oxford.

Schickore, Jutta, 2005, "Through thousands of errors we reach the truth – but how? On the epistemic roles of error in scientific practice", in *Studies in History and Philosophy of Science Part A*, 36, 3: 539-556.

Schikore, Jutta, 2020, "The structure and function of experimental control in the life sciences", in *Philosophy of Science*, 86, 2: 203-218.

Stensbøl T.B., Kapur S., 2015, NEWMEDS special issue commentary *Psychopharmacology* (Berl), 232(21-22): 3849-3851, DOI: 10.1007/s00213-015-4083-y

Sullivan, Jacqueline, Julie Dumont, Sara Memar, *et al*., 2020, "New frontiers in translational research: touchscreens, open science and the mouse translational research accelerator platform", in *Genes, Brain and Behavior*: e12705.

Sullivan, Jacqueline, 2018, "Optogenetics, pluralism and progress", in *Philosophy of Science*, 85, 5: 1090-1101.

Sullivan, Jacqueline, 2017, "Coordinated pluralism as a means to facilitate integrative taxonomies of cognition", in *Philosophical Explorations*, 20, 2: 129-145.

Sullivan, Jacqueline, 2016, "Construct stabilization and the unity of neuroscience", in *Philosophy of Science*, 83, 5: 662-673.

Sullivan, Jacqueline, 2014, "Stabilizing mental disorders: prospects and problems" in Harold Kincaid and Jacqueline Sullivan, eds., *Classifying Psychology: Mental Kinds and Natural Kinds*, MIT Press, Cambridge MA.

Sullivan, Jacqueline, 2009, "The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-reductionist Models of the Unity of Neuroscience", in *Synthese*, 167: 511-539.

Suppes, Patrick, 1984, *Probabilistic Metaphysics*, Blackwell Publishers, Oxford.

Wool, Lauren, International Brain Laboratory, 2020, "Knowledge across networks: how to build a global neuroscience collaboration", in *Current Opinion in Neurobiology*, 65: 100-107.