From “Intrusive Thinking: From Molecules to Free Will,” edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

5

Convergent Experimental Systems for Dissecting the Neurobiology of Intrusive Thought

A Road Map

Shannon L. Gourley, Antonello Bonci, Michael R. Bruchas, Shelly B. Flagel, Suzanne N. Haber, Peter W. Kalivas, Amy L. Milton, Paul E. M. Phillips, Marina R. Picciotto, and Jeremy K. Seamans

Abstract

Nonhuman experimental systems (also known as model organisms) are critical for understanding the neurobiology of intrusive thought. These model systems allow for the ability to manipulate specific neurocircuits, neurotransmitters, neuromodulators, and physiological and intracellular signaling events associated with behavioral markers that may be linked to intrusive thought. They permit unparalleled control over the external and genetic environments in ways and to degrees that are not possible in humans. Intrusive thought is an emergent property of multiple systems: emotional, cognitive, motor, and autonomic/somatic. In an animal model, one can ask specific questions about these systems and how they may be linked to, permit, or suppress intrusions. For example, how are specific connections, neuromodulators, or cell types involved in each of these systems, and how do they help form or maintain behaviors consistent with intrusive thought? Are positive versus negative valences unbalanced? Are common systems

Group photos (top left to bottom right) Shannon Gourley, Antonello Bonci, Suzanne Haber, Peter Kalivas, Amy Milton, Michael Bruchas, Shelly Flagel, Paul Phillips, Shannon Gourley, Jeremy Seamans, Antonello Bonci, Suzanne Haber, Marina Picciotto, Paul Phillips, Peter Kalivas, Amy Milton, Jeremy Seamans, Marina Picciotto, Shannon Gourley and Antonello Bonci, Shelly Flagel, Michael Bruchas

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

hijacked by intrusive thought, agnostic to the valence or content of the thought? Resolving these issues could be transformative for the treatment of several neuropsychiatric illnesses that are commonly characterized by intrusive thought. This chapter presents a road map for studying the neural mechanisms underlying intrusive thought using nonhuman experimental systems.

Introduction

Understanding the neurobiology of intrusive thought requires unfettered and unrestricted access to the brain. Thus, one turns to nonhuman experimental systems (also known as model organisms) because they allow admission to the brain as well as unparalleled control over the external and genetic environments, employing technical and experimental strategies that are not possible in humans. This access permits dissecting, quantifying, and manipulating specific neurocircuits, neurotransmitters, neuromodulators, and physiological and intracellular signaling events associated with behaviors. Once we develop strategies to infer intrusive thought in nonhuman experimental systems, several goals can be pursued, such as the identification of neurocircuits which are, or are not, associated with intrusive thought. Are distinct subcircuits, neuromodulators, or cell types involved in forming or maintaining intrusive thoughts with positive versus negative valence? Are common systems hijacked by intrusive thought, agnostic to the valence or content of the thought? Resolving these issues could be transformative in developing treatments for several neuropsychiatric illnesses that contain intrusive thinking as a pathogenic endophenotype. As discussed at greater lengths at other points in this volume, illnesses include common disorders such as drug addiction, posttraumatic stress disorder (PTSD), depression, and obsessive-compulsive disorder (OCD).

To develop a road map for studying intrusive thought in nonhuman experimental systems, our discussion begins by defining intrusive thought in the context of biological frameworks for the research laboratory. Next, we focus on conceptualizing intrusive thought as an emergent property of multiple systems. This leads us to formulate a road map for investigating intrusive thought in the future. Finally, we conclude by exploring the point from which we started and analyzing where we still need to go.

Defining Intrusive Thought in Biological Frameworks for the Research Laboratory

Our first goal is to set forth principles by which we can capture aspects of intrusions within the domain of experimental systems. Intrusive thought has been defined as *unwanted*, *unintended*, *conscious mental events* lacking control (Clark 2005). The aspects of this definition that we are best able

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

to capture, operationally and quantifiably, in an experimental subject are neurobehavioral events that occur ectopically (e.g., out of their appropriate contexts), recurrence, resistance to change, and induction of arousal. Such events are insensitive to modification by external stimuli that would typically redirect neural or behavioral activity, and these intrusions often interrupt adaptive behaviors.

Our aim is to develop a framework for how we might understand the neurobiology of intrusive thought using nonhuman experimental systems. A primary challenge is the inability of our subjects to express their thoughts, so to speak, which forces us to interpret their behavior as a surrogate measure of thoughts. To address this issue, we describe in Table 5.1 key concepts that should optimize any given approach. These concepts include construct, predictive, and face validities. Construct validity refers to the degree to which a given experimental strategy accurately measures what it is meant to be measuring. Predictive validity refers to the extent a strategy can make accurate predictions about the human condition. For instance, if a drug has anxiolytic properties in humans, it should have anxiolytic properties in a valid task of anxiety-like behavior in a rodent or nonhuman primate. Finally, face validity refers to the degree to which a given strategy reflects what it is attempting to model. We might ask, "Does this approach seem like it will measure intrusive thought?" Of course, reliability and reproducibility are also key considerations. In addition, we highlight the notion of antecedents, which, in this chapter, refers to factors that predispose an organism to, or directly triggers, an intrusive thought.

Table 5.1 Validities and considerations in designing research strategies.

Construct validity	The interpretability, meaningfulness, or explanatory power of a given model; the degree to which a test measures what it claims to be measuring: How well does it capture the underlying constructs?
Predictive validity	The ability of a model to lead to accurate predictions about the human phenomenon: How well does a procedure identify pharmacological agents tested in model organisms that have therapeutic value in humans?
Face validity	The extent to which a test is subjectively viewed as reflecting the concept it intends to measure: Does it seem like it is really going to measure intrusive thought?
Reliability	Stability and consistency with which a variable of interest can be measured; phenomenon is readily reproduced under similar circumstances
Antecedents	The extent to which conditions in the model recapitulate factors that precede or trigger the phenomenon of interest (here, intrusive thought)

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

Conceptualizing Intrusive Thought as an Emergent Property of Multiple Systems

Embedded in the argument that nonhuman experimental systems have utility in studying the etiology and neurobiology of intrusive thought are two fundamental notions:

1. Intrusive thought is, to some degree, conserved across rodent and primate species (and thus likely has some adaptive origins).
2. Corollaries and consequences of intrusive thought can be measured and quantified in the absence of speech.

To the first point, one can imagine instances in which multiple types of organisms would benefit from uninterrupted and intensive thought, such as when the goal is to escape from a predator. However, an individual must also be able to modify and shift focus when the situation changes (e.g., when the threat has been resolved) and engage in other behaviors that are more adaptive or otherwise suited to present and evolving contexts. A *failure* to inhibit intrusive thinking can distract from achieving adaptive goals.

As shown in Figure 5.1, we conceptualize intrusive thought to be an emergent property of multiple systems: emotional, cognitive, motor, and autonomic/somatic (for discussion of the origin of intrusive thought in these systems, see Roberts et al., this volume). We envision a world representation that contains these four coexisting elements, which homeostatically analyze and validate environmental and intrinsic (thoughts) stimuli to elicit adaptive behavior. Emotional and motivating content draw on circuitry in the central zone of Figure 5.1. Intrusive events trigger a deviation from the homeostatic condition; thoughts contain more excessive motivational and attentional relevance to the individual than is appropriate for the environment. In neuropsychiatric pathologies characterized in part by intrusive thoughts, this deviation is associated with a loss of proper regulation of the inner circuitry by the outer circuitry, as indicated in Figure 5.1 (for definitions of typical vs. intrusive thoughts envisioned by the model in Figure 5.1, see Table 5.2).

We envision that any given mental health disorder can coopt different domain hierarchies. Identifying these hierarchies could offer clues into the neurocircuits that one might explore in investigating etiologies and developing treatment strategies. For example, in disorders in which cognitive behavioral therapy can be effective, such as OCD, the cognitive domain plays a significant role in generating overall circuit feedback that restores homeostasis and control of the intrusions. We hypothesize that the distinct disorders or endophenotypes of disorders defined by DSM-5 have the order of domain dominance shown in Table 5.3.

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

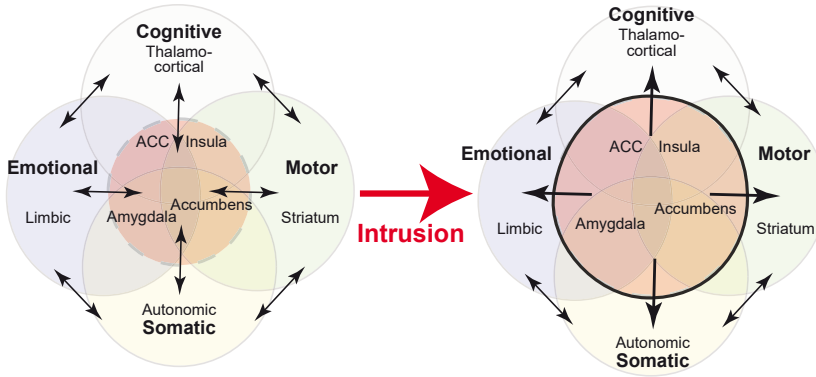

Figure 5.1 Intrusive thought is an emergent property of multiple systems. We propose that emotional, cognitive, motor, and somatic domains are recruited to interpret sensory or internal stimuli and generate adaptive responses. Navigating a complex world and adjusting our behaviors appropriately requires homeostatic involvement of these components. Depending on the arousal and motivation that emerges from this homeostatic interpretation, the central region is recruited (illustrated by the central red circle) to augment motivational value and attention. Each domain contains circuitry within the central region, as indicated by the brain nuclei listed (ACC, anterior cingulate cortex), that contributes salience to adaptive homeostatic interpretations and responses. Normally, we evaluate and modify our behavior, such as when in an aroused state that arises from a threat. In assessing the threat, we iterate between the external and internal circuits (bidirectional arrow), adjusting our appraisal of the stimuli through feedback between the circuits to generate the most appropriate responses. Though random thoughts occur, they are continuously appraised and only become problematic when the appraisal and motivation/arousal generated by the inner circuitry does not match the information received from the outer circuitry. Accordingly, the inner motivational circuitry becomes resistant to, or dominates, the outer cognitive circuitry (one-way black arrow from inner to outer circuitry). This leads to a loss of homeostatic response and manifests as excessive, maladaptive thoughts and possibly inappropriate behaviors. For instance, an intrusion producing a pronounced feeling of anxiety that cannot be regulated in a nonthreatening situation will produce autonomic, emotional hyperarousal and inaccurate conscious assessments of situations that could manifest as posttraumatic stress disorder.

Formulating a Road Map for Investigating Intrusive Thought

Keeping in mind the definitions, considerations, and concepts laid out above, we can begin to develop a meaningful list of behavioral and other factors that could be tractably measured in experimental systems (Table 5.4). For instance, we could capitalize on the ability of an external stimulus to distract an experimental subject from engaging in goal-directed actions:

- Will a rat in an operant-conditioning testing chamber respond for food reinforcers even in the presence of an opioid-related cue?

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

Table 5.2 Definitions of a typical versus pathological thought intrusion.

Typical	• Regulated by homeostatic cross talk between cognitive, emotional, motor, and somatic domains (see Figure 5.1).
	• Motivationally relevant information initially involves many classic limbic circuits and brain regions (e.g., ventral prefrontal and orbital cortices, insula, nucleus accumbens, and amygdala), see Figure 5.1, which act in concert with the outer circuitry to validate and regulate the state of motivation and arousal.
	• Level of arousal and motivation is appropriately managed by the outer circuitry, through feedback with the environment, to create and modulate the behavioral response.
Pathological	• Contains all of the elements of a normal event, except that the limbic circuitry (purple area, Figure 5.1) is not properly managed by the cognitive circuitry (outer circuit, Figure 5.1), thus creating an imbalance.
	• The resulting state of motivated hyperarousal leads to stress that is perpetuated without access to adaptive feedback and/or regulation by the outer circuitry and environment.
	• This imbalance can develop initially from any domain, and various combinations may be more typical in different neuropsychiatric disorders.
	• The different domain hierarchies which create the intrusion predominant in a given disorder offer a potential focus for experimental exploration into the underlying neuropathology of homeostatic loss of control.

Table 5.3 Utility of model organisms: behavioral measures relevant to intrusive thought. Here we broadly summarize behaviors and additional considerations relevant to investigations of intrusive thought in nonhuman experimental systems.

What can we measure?
• Disruption of goal-directed behavior (e.g., by drugs of abuse or aversive stimuli)
• Persistent avoidance of aversive stimuli (e.g., despite extinction conditions)
• Persistence of a given behavior, despite adverse consequences or punishment
• Persistence of a behavior in the absence of a conditioned stimulus or instrumental contingency (e.g., conditioned freezing that generalizes or fails to extinguish)
• Cognitive domains that are known to be affected in human conditions
Additional factors:
• Vulnerability factors (e.g., early-life stress)
• Individual differences
• Behavioral comorbidities that model known comorbidities in humans
• Recurrence and potential worsening with time (akin to sensitization or kindling)

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

- Alternatively, will the rat instead attend to the cue at the expense of goal-directed food seeking?
- What differs, at a neurobiological level, between the rat who becomes distracted and the rat who stays on task?

Similar to food-seeking behavior, drug-seeking behavior can have goal-directed properties, but the important measure that one might wish to collect in the context of intrusive thought is the degree to which it competes with a presumably more adaptive behavior, such as food seeking in a calorie-restricted organism.

Some recently developed procedures (a) force organisms to arbitrate between food-seeking behaviors and the avoidance of aversive stimuli and then (b) track the extinction of avoidance behavior in the absence of the stimulus (Bravo-Rivera et al. 2015; Rodriguez-Romaguera et al. 2016). Others measure the reaction of organisms to uncertainty (d'Angelo et al. 2014, 2017; Eagle et al. 2014; Morein-Zamir et al. 2018). Both strategies could be used to investigate mechanisms of intrusive thought, particularly when considered with factors such as individual differences or antecedents to intrusive thought, many of which can be recapitulated in the laboratory (Table 5.5).

Another type of intrusion that could be recapitulated in nonhuman experimental systems is the sense of incompletion of a task that looms until the process is completed. One example in humans draws from obsessive handwashing in OCD: the notion that one's hands must be washed is all consuming, generating hyperarousal and stress until one washes their hands, resolving the intrusion. Procedures in nonhuman experimental systems, such as persistent

Table 5.4 Examples of potential domain hierarchies involved in disorders containing maladaptive intrusions. The interacting domains defined in Figure 5.1 may be associated with particular neuropsychiatric conditions to a greater or lesser degree.

Condition	Proposed Hierarchy of Domains
OCD	Cognitive = motor > Emotional = somatic
PTSD	Emotional = somatic > Cognitive = motor
Craving in substance use disorders	Emotional = motor = somatic > cognitive
Rumination in depression	Emotional = somatic = cognitive > motor

Table 5.5 Antecedents to intrusive thought can be recapitulated in the laboratory. For arousal, long-term events are historical events that give rise to vulnerabilities and resiliencies, whereas short-term events refer to triggering factors.

Long-Term Events	Short-Term Events
Early-life experiences (early-life stressors)	Autonomic responses and stressors
Genetic correlates	Conflict
Environmental insults (drugs of abuse, trauma)	Emotional representations of environmental stimuli

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

response to a drug-paired cue or the "observing response task" (see d'Angelo et al. 2017), have utility in capturing both motor- and circuit-level aspects of incompleteness. In addition, most instrumental behavioral tasks include a discrete signal, such as a light, which designates the completion of a response requirement. Omission of that signal generates extended responding, a behavior that is potentially motivated by a sense of incompletion.

A third approach to studying the concept of incompleteness, which is not mutually exclusive, would be to measure neural signals that demarcate the completion of a response. For example, during cocaine self-administration, a phasic dopamine signal is observed in the nucleus accumbens (NAc) of rats upon completion of a response requirement (Phillips et al. 2003; Willuhn et al. 2012). In some animals, this feedback signal becomes diminished and, as a consequence, animals keep repeating the action, resulting in higher drug consumption (Willuhn et al. 2014).

Five Strategies for Investigating Intrusive Thought

To study intrusive thought in nonhuman experimental organisms, we propose five different approaches, summarized in Table 5.6 and discussed below.

Back-Translation, Susceptibility, and Resilience

It should go without saying that nonhuman experimental systems will be of greatest value if used in conjunction with appropriate types of behavioral analyses. There is, of course, inherent difficulty in translating clinical behaviors directly into an animal "model" of a mental health disorder, and one can debate whether it is even possible to model a mental health disorder in its entirety in animals (Bale et al. 2019). Back-translation offers a complementary approach. Broadly, this term refers to the identification of components of mental health disorders in humans that can be measured in nonhuman experimental systems. The concept of back-translation informs our first two categories: experimental approaches are driven by (a) factors implicated in human behavior, and/or by (b) vulnerability and resiliency factors that have been documented in humans. Ultimately, of course, the aim is to have "simultaneous translation" with different research approaches in humans and nonhuman experimental systems converging on the same findings (Milton and Holmes 2018).

Notably, one specific form of back-translation refers to the deconstruction of mental health disorders into specific psychological processes that can be studied in nonhuman experimental systems. In doing so, one should avoid measures that are subjective and self-reported; instead, the research scientist should look to data that are readily quantifiable and free of confounding influences. One example is performance on a battery of psychological tasks, such as the Cambridge Neuropsychological Test Automated Battery. By identifying

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

deficits exhibited by patients (e.g., in attentional set shifting or working memory), it should then be possible to identify populations of animals that manifest the same deficits. Population variation could be induced through genetic manipulations, early-life experiences, more proximal life experiences (e.g., exposure to drugs of abuse), or perturbation of neural circuits or neurochemical systems. In other words, one could expose the nonhuman experimental organism to putative vulnerability factors that would be expected to exacerbate intrusive thought. The readily quantifiable nature of the deficits in psychological processing should allow for identification of potential neural and neurochemical etiologies as well as hypothesis testing using the full range of tools available to animal researchers.

Treatment Mechanisms

Another strategy, which we call "treatment mechanisms," refers to our understanding of why, biologically, certain treatments are effective. Cingulotomies improve intrusive thought in chronic pain management by mitigating the distressing nature of pain, but not the pain itself. The mechanisms by which this phenomenon occurs remain elusive and could hypothetically be examined in nonhuman experimental systems, opening a window for identification and deep interrogation of cells that are excited, inhibited, or otherwise modulated by cingulotomy. Such cell populations could then be manipulated pharmacologically, genetically, or through other strategies, such as those detailed by Bruchas (this volume). This approach would allow one to isolate neurobiological correlates of successful interventions whose identification could ultimately point in the direction of new and better treatment strategies.

Decrypting the Ensemble

The recurrent nature of intrusive thoughts suggests that some type(s) of similarly recurrent oscillatory or reverberating neural processes could be identified in effective models. These processes could then be exploited to expand understanding into the etiology of recurrent thought, a strategy that we refer to as "decrypting the ensemble." This strategy is inspired, in part, by evidence that in subcortical areas, the plasticity of conditioned fear-related behaviors is accompanied by transient, measurable changes in the expression of calcium-permeable and calcium-impermeable α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors (Clem and Huganir 2010; Rao-Ruiz et al. 2011). Additionally, changes in the ratio of subtypes of N-methyl-D-aspartate (NMDA) receptors can determine whether a fear memory ensemble is stable or susceptible to strengthening following plasticity (Holehonnur et al. 2016).

The specific content of intrusions can be common across a large number of patients and interacts with the environment (e.g., an increase in the number of patients reporting obsessions regarding contamination with acquired immune

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

Table 5.6 Approaches to empirically investigate aspects of intrusive thought in nonhuman experimental systems. These approaches are not mutually exclusive. Each approach is named and briefly described. Each strategy could be combined with stimuli meant to trigger or exacerbate an intrusive thought and/or procedures that introduce conflict. The motivation to pursue a specific strategy is described (i.e., the strengths of a given approach) and examples of each strategy provided. These lists are not exhaustive and some examples cross categories.

Approach	Motivation	Examples
1. *Back-translation* investigates genetic/molecular factors, circuits, and behaviors that are highly (often, causally) implicated in, or comorbid with, human conditions	• Informed directly by the human condition (i.e., high construct validity)	• Genetics: study the role of the *Fmr1* gene in repetitive behavior in fragile X syndrome • Circuits: investigate interactions between the prefrontal cortex and nucleus accumbens in rodent addiction models, an approach originally inspired by clinical research • Behavioral: investigate cognitive functions disrupted in neuropsychiatric illness (e.g., attentional function)
2. *Susceptibility and resilience* investigate known vulnerability and resilience factors in humans	• Potential for informing precision medicine • Understanding resilience could shed light on mechanisms of coping	• Genetics: study known genetic risk factors in Tourette syndrome and addiction • Behavioral: identify neurobiological effects of early-life stress or enriched environments
3. *Treatment mechanisms* investigates the neurobiology of effective treatments	• Informed directly by "what works and what does not" in treatment • Could shed light on disease mechanism or reveal new targets for treatment	• Surgical: deep brain stimulation, used to treat multiple conditions, can be studied in nonhuman experimental systems • Behavioral: conditioned fear extinction procedures could recapitulate aspects of cognitive behavioral therapy • Pharmacological: use of buprenorphine treatment for craving in opioid misuse

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

Table 5.6 (continued)

Approach	Motivation	Examples
4. *Decrypting the ensemble* decodes and recreates neural representations associated with events or stimuli not currently present.	• Causality and biomarkers can be simultaneously identified	• Use electrophysiology, imaging, or molecular labeling to identify cellular changes associated with a given behavior • Thereafter, attempt to *simulate* the same behavior by recapitulating neural or molecular changes
5. *Quantify natural behaviors* measures species-appropriate, ethologically relevant physiological events that occur at atypical time points or frequencies.	• Unbiased, exploratory, and comprehensive • Could benefit from within-subjects measures to capture individual- and population-level differences	• Ultrasonic vocalization, heart rate variability, physiological events (e.g., sleep, pupil dilation), clustering of naturally occurring behaviors (e.g., feeding, grooming) that manifest spontaneously in conjunction with an external stimulus (e.g., foot shock, aggressive encounter)

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

deficiency syndrome in the 1980s and 1990s). It is likely, therefore, that there are particular concepts and associated neural ensembles that are recruited for such intrusions. This phenomenon may reflect intrinsic differences in excitability in specific neuronal populations, as is the case for amygdala neurons which show high levels of cAMP response element binding protein (CREB) phosphorylation (i.e., activation), which are subsequently more likely to be recruited to fear-related neuronal ensembles following an aversive experience than other neurons (Josselyn et al. 2001; Frankland and Josselyn 2014; Josselyn and Frankland 2018). In other words, this subset of neurons with higher levels of phosphorylated CREB are primed to respond to input. They therefore could be an ensemble that could predispose an individual to an emergent (intrusive) event. Similar phenomena have been described in the context of immediate early gene expression (Suto et al. 2016; Whitaker et al. 2017), and these kinds of strategies could be deployed formally to study intrusive thought.

Quantification of Naturally Occurring Behaviors

One way to minimize anthropomorphic bias in interpreting animals' behavior is to utilize a final approach, which we term "quantification of naturally occurring behaviors." Here, the investigator records species-appropriate, ethologically relevant physiological events (e.g., ultrasonic vocalization or naturally occurring grooming or feeding behaviors) and identifies neural correlates when these behaviors deviate in quality or quantity from what is typical. In this and all of our categories, the introduction of triggers (e.g., drugs of abuse or stressors) will likely have utility in creating a situation that would cause intrusive thought, which then manifests in a behavior that can be studied.

Considerations

It is important to acknowledge multiple limitations inherent in nonhuman experimental systems as well as in the strategies that we propose. These include, but are not limited to, the following:

1. There is a risk that phenomena unrelated to intrusive thoughts are measured.
2. False negatives may arise, due to an inability to collect sufficient information (e.g., in cases of multiple interacting brain regions and/or neuromodulators).
3. Manipulations of molecules/cells using many current tools may not be physiologically relevant (see Bruchas, this volume).
4. In experimental organisms, the emotional content of any given manipulation is difficult to measure conclusively.

Utilizing multiple strategies should help us overcome these challenges and identify convergences.

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

Another challenge that we anticipate is resolving the complexity of many modern cellular/physiological data sets. A comprehensive brain map whereby affective circuits are defined by key features at a complete reductionist level is needed. Such a map would include:

1. RNA profiling of millions of neurons within the mammalian brain in naive versus "intrusive event-like states" within discrete brain structures (e.g., Campbell et al. 2017; Saunders et al. 2018; Gouwens et al. 2019; Mickelsen et al. 2019).
2. Mapping a functional architecture of these cell types alongside discrete behavior epochs or states using imaging and/or physiological processes.
3. Environmental, genetic, behavioral, and pharmacological dissociation of critical manipulations which impinge upon and alter the transcriptional state and functional responsivity of these maps.

Finally, a fruitful conceptual strategy may be to switch from assigning special functions to genes or neurons to computational perspectives of how ensembles work together and coordinate complex behavior.

Naturally, behavioral data sets can also be quite complex. A fertile strategy may be to differentiate learned association structures, based on the complexity of the information that is stored to support that association, using the "model-free" and "model-based" terminology from reinforcement learning. A model-free computation assigns a single dimensional value to a stimulus based upon the reliability of its association with a motivationally relevant outcome. In contrast, a model-based computation establishes a model of the environment that can be used to explore potential and inferred connections between stimuli and states. These concepts are discussed at greater length by Phillips and Milton (this volume).

Where Did We Start, and Where Are We Going?

A primary goal of this chapter is to offer a road map for future investigations of intrusive thought. In this final section, we describe a selection of relevant published and ongoing investigations and consider how these investigations may be further developed to interrogate intrusive thought in nonhuman experimental systems.

Global Modulation of Networks and Brain Regions versus Individual Parts of the Circuit: A Place In Between

Intrusive thought emerges from an interaction between several functional domains—emotional, cognitive, motor (see Figure 5.1)—and is often triggered by internal or external sensory stimuli. Typically, in response to stimuli, the

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

interactions between these domains flow smoothly, with each taking a leadership role in the appropriate situation but then returning to the status quo. For example, an internal or external stimulus might invoke fear, overwhelming the individual until either an escape (movement) is executed or the cognitive control system takes over if the fear is unwarranted. Mediating between these functional domains requires the complex integration of information across them to resolve the situation. Intrusive thought can be considered to be a condition in which this mediation fails. Identifying the circuitry that underlies the integration of information processing across the different domains is a first step in understanding how and where these areas communicate, necessary to developing new therapeutic targets.

Karl Wernicke first recognized that connectivity of brain structures, rather than their locations, was the central feature of higher-order cognitive functions (Wernicke 1885/1994). Expanding on the idea, Geschwind suggested that this involves a combination of functional localization and connectivity, leading to the idea that the brain is comprised of complex, interrelated functional networks (Geschwind 1965a, b; Catani and Ffytche 2005). Functional imaging studies and graph theory techniques moved the field forward, demonstrating large-scale distributed networks and the existence of nodes and hubs (Sporns 2011). A node is an area that is connected locally or connected within a functional system. A hub is a node of a network that has unusually high connectivity to other nodes, or degree centrality, and high connectivity to other hubs, or eigenvalue centrality (van den Heuvel and Sporns 2013). Hubs are thought to represent regions for integrating and distributing information from multiple cortical regions. They likely play an important role in cross-functional computational tasks, such as integrating limbic, cognitive, and motor control calculations for decision making.

The rostral anterior cingulate cortex (rACC) is a good candidate for containing hubs, because it sits at the connectional intersection of the emotion, cognition, and executive control networks. Indeed, the entire rACC is considered a hub of the brain's global network (Buckner et al. 2009). However, the region is large, and inputs from different prefrontal cortical (PFC) functional domains vary across it. These connections could represent information processing sequentially across subregions (i.e., from valuation to cognition to action). Alternatively, a hub could be embedded within the rACC that integrates information across them. Consistent with the literature (Morecraft and Tanji 2009; Morecraft et al. 2012), mapping the distribution and relative strength of frontal cortical inputs across rACC in nonhuman primates reveals that PFC inputs to the rACC follow three general gradients:

1. Ventromedial PFC (vmPFC) and frontal pole inputs are strongest in the ventral and rostral parts of the rACC, with decreasing strengths in dorsal and caudal regions.

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

2. Frontal eye fields and premotor areas inputs are strongest in the dorsal and caudal regions, decreasing in rostral and ventral rACC regions.
3. Ventrolateral PFC (vlPFC) and dorsolateral PFC (dlPFC) inputs peak in a more central position.

One region embedded within these gradients, however, receives inputs from unexpected additional areas. In addition to inputs from expected connections from cognitive control areas, the dlPFC and vlPFC, this region is also connected with regions that are part of the emotional system, the orbitofrontal cortex (OFC) and vmPFC, as well as with the sensorimotor system, the frontal eye fields (Tang et al. 2019). Thus, this connectional hub within the rACC is in a position to *integrate information across emotional, cognitive, and sensorimotor systems*. It is perhaps unsurprising, then, that both PTSD and major depressive disorder show treatment response in an area in close proximity to the rACC hub (Mayberg et al. 1997; Pizzagalli 2011; Chakrabarty et al. 2016).

The striatum is also an important structure for integrating and distributing information. Although the striatum is classically divided into limbic, cognitive, and motor regions, embedded within this general topography, terminals from different frontal cortical areas interface in the rostral striatum, positioning it optimally to contain hubs (Haber et al. 2006; Averbeck et al. 2014). Indeed, in a specific location within the rostral caudate nucleus, terminal zones from the inferior parietal lobule, an area important for perception, converge not only with those from the dlPFC and vlPFC, as expected (Cavada and Goldman-Rakic 1991; Yeterian and Pandya 1993), but also with projections from the OFC and rACC. Thus, similar to the rACC, this hub combines inputs from several functional domains. The connections of the rACC and striatal hubs are examples of highly integrative, cross-functional regions with distinct combinational inputs that provide the anatomical substrate in which computations about motivation, internal states, cognition, perception, and motor control are linked to mediate adaptive behaviors based on the interaction of these functions. Disconnection of these hubs will likely result in an imbalance between goal-directed control, emotion, and higher cognition, and thus play a key role in maintaining intrusive thoughts.

Seeking the Source of Switching

One important aspect of countering or managing intrusive thought is the ability to *modify* thoughts and actions, a key ingredient in adaptive responding to external stimuli (Figure 5.1). Multiple structures discussed above are naturally involved in these processes and could be a focus for future investigations, particularly those that enjoy considerable homology between rodent and primate species. One such structure, the ventrolateral orbitofrontal cortex (vlOFC), has been intensively investigated using an instrumental contingency degradation procedure. In brief, the procedure requires nonhuman (or human) experimental

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

systems to generate operant responses for rewards such as food or juice. Then, the experimenter modifies the likelihood that a given behavior will be reinforced, and organisms must update learned action–outcome associations to modify their responding optimally. In a series of investigations, inactivation of the vlOFC blocked the ability of mice to update response strategies (Gourley et al. 2013a; Zimmermann et al. 2017, 2018; Whyte et al. 2019), consistent with evidence that certain vlOFC neurons represent outcome-related memories (Namboodiri et al. 2019).

The vlOFC interacts with aspects of the dorsal striatum, a key constituent of goal-directed action, to coordinate action–outcome response flexibility (Gourley et al. 2013a; Gremel and Costa 2013). Meanwhile, the use of viral-mediated gene silencing and behavioral pharmacological strategies has revealed several essential molecular factors within the vlOFC that optimize its function. These factors include, but are likely not limited to, brain-derived neurotrophic factor (Gourley et al. 2013a; Zimmermann et al. 2017; Pitts et al. 2020) and its high-affinity receptor trkB (Pitts et al. 2018, 2020), Abl2 kinase (DePoy et al. 2017), GABAAα1 receptor subunits (Swanson et al. 2015), fragile X mental retardation protein (Whyte et al. 2019), and developmental expression of integrin receptors (DePoy et al. 2019). These investigations provide overwhelming evidence that the vlOFC is necessary for behavioral switching, and they potentially shed light on molecular factors that are disrupted when intrusive thoughts interfere with behavioral flexibility essential to day-to-day function.

One common factor linking all of these proteins is that they regulate the stability or turnover of dendritic spines, the primary sites of excitatory plasticity in the brain. Whyte et al. (2019) revealed that updating expectations regarding whether an action was likely to be rewarded reduced thin-type dendritic spines, considered immature, on excitatory vlOFC neurons in mice. Meanwhile, the proportion of mushroom-shaped spines, considered mature and likely containing synapses, increased, potentially solidifying newly modified action–outcome associations to optimize future decision making and behavioral flexibility.

One function ascribed to the OFC as a whole is the updating of expectations, particularly under ambiguous circumstances; by extension, unbalancing these connections via spine loss or inappropriate excitatory plasticity could render expectations ambiguous and thereby vulnerable to intrusion by competing impulses. Consistent with these notions, exposure to cocaine (Gourley et al. 2012a; DePoy et al. 2017; Pitts et al. 2020) and stress hormones eliminates dendritic spines in the vlOFC, and identical procedures cause failures in the action–outcome updating of stress hormones (Gourley et al. 2012b, 2013b; Barfield et al. 2017; Barfield and Gourley 2019). Further, drugs that improve behavioral updating appear to recruit local cytoskeletal regulatory systems (DePoy et al. 2017). As a final note, artificially stimulating excitatory neurons in this region also causes failures in action–outcome updating (Hinton et al.

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

2019), potentially by activating circuits associated with OCD. This idea is discussed at length by Balleine (this volume). Understanding the conditions under which specific vlOFC connections are stimulated or quiescent could shed light on how thoughts and actions can fail to be updated and become intrusive, and how one recovers from their intrusion by switching cognitive strategies or behaviors, maintaining adaptive flexibility between emotional, cognitive, motor, or somatic domains (Figure 5.1).

Unfortunately, cellular heterogeneity within several brain regions pertinent to this discussion remains undefined. For instance, even within the striatum, where cells subtypes can readily be distinguished based on dopamine receptor constituents, dopamine D1 and D2 receptor-mediated neuronal ensemble responses to cues and rewards over temporal dimensions, and as a function of experience, remain opaque (for review, see Castro and Bruchas 2019). Determining the contribution of each cell to learning, memory, stressor, and drug reactivity, for instance, has previously been studied using *in vivo* physiological approaches, yet the process of defining circuit- and cell-type specifics in time and space is still in its infancy. As such, comprehensively understanding the neurobiological bases of neuronal ensembles is a critical step if we are to dissect and understand the aberrant patterns (signatures) by which intrusive events occur.

Arousal Systems

Intrusive thinking includes arousal, and deviations from typical arousal states can predispose one to, or acutely trigger, intrusive thoughts, a notion emphasized in Table 5.5. As such, understanding the mechanisms of arousal presents a point of entry into understanding intrusive thought itself. Within the framework that arousal can decrease the threshold for permeation of intrusive thoughts, we might consider the ability of acetylcholine (ACh) release elicited by salient stimuli to alter the strength of signaling in thalamo-cortico-thalamic loops, both acutely and persistently, via synaptic potentiation (Aramakis et al. 2000; Kawai et al. 2007). This ACh-mediated elevation in activity may alter the threshold for transmission of sensory information from subcortical to cortical structures. The directionality of this signaling can vary across development, with different ACh receptors mediating increases or decreases in the transmission of sensory information (Aramakis et al. 2000; Heath and Picciotto 2009).

Although we recognize that hallucinations may not reach a formal definition of intrusive thoughts, it could be useful to evaluate the particular circuits for which we have direct evidence of a causal relationship with this perceived, maladaptive mental event. Pharmacological studies (Warburton et al. 1985; Fisher 1991), as well as evaluations of patients with loss of cholinergic neurons (Dauwan et al. 2018), reveal that blocking muscarinic ACh receptors or decreasing ACh levels in patients with Lewy body dementia (Tsunoda et al.

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

2018; Dudley et al. 2019) results in hallucinations. One proposal is that loss of either nicotinic or muscarinic activity in corticothalamic circuits may underlie these hallucinations (Esmaeeli et al. 2019). Another suggestion is that cortical ACh increases the signal-to-noise ratio of perceived events, and that "muscarinic receptor activation in the cortex is involved in confining the contents of the discrete self-reported conscious 'stream' " (Perry and Perry 1995:240). When cholinergic input to the cortex is lost, irrelevant sensory information normally confined to subcortical circuits enters conscious awareness (Perry and Perry 1995), and hallucinations result from "a failure of the metacognitive skills involved in discriminating between self-generated and external sources of information" (Kumar et al. 2009:119).

An additional set of studies suggests that there is a pervasive increase in ACh levels throughout the brain in patients who are actively depressed: for unipolar depression, see Saricicek et al. (2012); for bipolar depression, see Hannestad et al. (2013). Elevated ACh may be a risk factor for depression since remitted patients have intermediate ACh levels between actively depressed individuals and healthy controls (Saricicek et al. 2012), as measured by competition with a cholinergic ligand and validated by within-subject challenge with a cholinesterase blocker (Esterlis et al. 2013). Relatedly, ACh is a critical mediator of arousal and rapid eye movement sleep (Ma et al. 2018). At baseline, ACh input to the basolateral amygdala is very high, and tonic activity of the cholinergic system can thus control both the level of arousal of stress-related systems and the likelihood that a stressful event will activate the basolateral amygdala (Picciotto et al. 2012). Currently, there is no information on whether this increase in ACh levels is associated with intrusive thoughts (e.g., rumination in depression), but this topic could guide future experiments.

Manipulations and measurements of both cholinergic signaling and circuits modulated by its receptors may represent a cross-species neurobiological approach ripe for translational evaluation. One consideration is that intrusive thoughts can be represented in experimental settings by regulation of arousal states along a multidimensional continuum. A possible dimension along this continuum includes asynchronous, decoupled activity of the filter or gain domain which prohibits "normal" function in a given circuit, steering an organism toward a hyperaroused state. One particular example of this phenomenon is found in the locus coeruleus, which projects broadly throughout the brain, and its activity (tonic vs. phasic) is dictated by salience, context, and stress responsivity. The ability of the locus coeruleus noradrenergic system to dissociate attention signals from stressful ones depends on which inhibitory filters are engaged. Along the temporal dimension, intrusive thought may have the effect of dysregulating the inhibitory gain signal (typically regulated by neuromodulators such as neuropeptides, monoamines, and steroids), thereby producing an unwanted hyperaroused state.

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

Lessons from the Study of Cocaine

Nonhuman experimental systems of many species will self-administer drugs of abuse, including cocaine, thereby providing a strong measure of face validity to drug self-administration in animal models of human addiction. As in humans, drug seeking can intensify with time and experience, take on compulsive properties, and persist despite adverse consequences, allowing for the intense investigation of neurobiological etiologies.

Over the last 15 years, several research groups have modeled compulsive drug-seeking behavior in the face of aversive consequences in preclinical models, with the ultimate goal of mimicking, as closely as possible, the symptoms, diagnostic criteria, and features of addiction manifested by human patients (Deroche-Gamonet et al. 2004; Belin et al. 2008; Economidou et al. 2009; Marchant et al. 2014; Belin-Rauscent et al. 2016). In well-validated rodent models, voluntary drug-taking and drug-seeking behaviors coincide with mild foot shock punishments, which are used to create negative consequences following drug use. Exposure to foot shocks has revealed a divide in rodent phenotypes into two separate groups: (a) those which are shock sensitive, whereby the rat ceases to press a lever after receiving the foot shock and (b) those which are shock resistant, whereby the rat keeps pressing the lever despite receiving the foot shock. Much like human populations who develop compulsive drug abuse or addiction, approximately 30% of rodents exhibit the shock-resistant phenotype.

The utility of the punished model of compulsive drug use in identifying and developing translational therapeutics was demonstrated by Chen et al. (2013a). In this study, the authors discovered that shock-resistant rats self-administering cocaine show a marked reduction of activity in the prelimbic PFC, a subregion of the PFC that is important in mediating behavioral flexibility and decision making. When Chen et al. reversed hypoactivity of this brain region via optogenetic activation of the prelimbic PFC, rats significantly, and almost instantaneously, reduced their cocaine-seeking behaviors. These findings led to clinical trials using a well-known, noninvasive form of brain stimulation, repetitive transcranial magnetic stimulation (rTMS), previously used as a treatment for depression. These clinical trials revealed that rTMS reduces cocaine craving and cocaine intake, thus paving the way for larger double-blind clinical trials that are the first to offer a promising treatment against cocaine use disorder (Terraneo et al. 2016; Pettorruso et al. 2018). The results of these translational studies highlight the importance of continuing efforts in developing increasingly sophisticated rodent models of substance use disorders, which are fundamental in leading to the next generation of treatments against substance use and other addictions, and more broadly, intrusions on adaptive functioning.

The term "incubation" refers to progressive, time-dependent elevations in drug craving and sensitivity to drug-related cues. Incubation is thought to

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

contribute to the maintenance and persistence of addiction and relapse (Lu et al. 2004; Venniro et al. 2016). By studying the incubation of cocaine-seeking behaviors in model organisms, we may also gain insight into certain forms of intrusive thought. Craving is considered a key factor in triggering relapse and can be triggered by drug-associated contexts or discrete drug cues. Although craving in humans is typically triggered through a combination of drug-associated contexts and cues, the two stimuli involve distinct and overlapping circuits. Thus, in animal models, it is useful to isolate them from each other during the incubation period; for instance, by extinguishing behavioral response to the drug-associated context with daily exposure to the context in the absence of drug availability. Accordingly, when the drug-paired cue is returned, it becomes possible to quantify behavior motivated only by the cue and to evaluate changes in circuitry and cell physiology produced by the cue-induced motivational state.

Within the context of dopamine-related pathologies (which presumably include substance use disorders), a change in the configuration of dopamine D3 receptors in the NAc has been observed, particularly when tonic dopamine levels are low. This change has not yet been fully characterized but may include a change in the ratio of the D3nf isoform; it also seems to enhance functional coupling with dopamine D1 receptors. This change has been associated with ticks in Tourette syndrome (Frau et al. 2016), L-DOPA induced dyskinesia (Fanni et al. 2019), and pathological gambling following dopamine agonist treatment in Parkinson disease (Pes et al. 2017). Treatment with the 5α-reductase inhibitor, finasteride, reverses associated molecular changes and ameliorates each of those pathological traits. Preliminary data indicate that finasteride also reverses incubation of cocaine craving and reduces escalated cocaine consumption (P. E. M. Phillips, pers. comm.), a finding that may be relevant to intrusive thought as well.

Distinct drugs of abuse (e.g., cocaine, heroin, alcohol) induce both similar and divergent neurobiological changes in brain regions like the NAc and frontal cortex. In light of the overlap in drug-seeking endophenotypes produced by, for example, the self-administration of opioids and psychostimulants, and the shared vulnerability to relapse in addiction, one argument has been that understanding the shared (rather than distinct) neurobiological factors might be particularly fruitful. One such example is elevated synaptic glutamate spillover from prelimbic PFC projections in the NAc during drug seeking for cocaine, heroin, alcohol, and nicotine. Under typical conditions, synaptic glutamate spillover is moderated by the glial glutamate transporter, GLT-1, located on astroglial end feet adjacent to the synaptic cleft. GLT-1 tightly controls basal extracellular glutamate, protecting against synaptic glutamate spillover. However, several drug classes downregulate GLT-1 and retract glial end feet from NAc synapses, modifications that have been directly linked to drug-seeking behavior (for a review, see Bobadilla et al. 2017).

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

Notably, acute stress has many overlapping effects, including decreased glial synapse coverage and reduced GLT-1 in the NAc (Garcia-Keller et al. 2016) and induction of glutamate spillover from unprotected glutamatergic synapses, which produces transient synaptic potentiation in NAc dopamine D1 receptor-expressing neurons (Scofield et al. 2016). Transient potentiation creates a situation that reinforces behavioral responses to stress- or drug-paired cues, making responding more persistent and more competitive with other stimuli (Kalivas and Kalivas 2016). Given that intrusive thought is similarly characterized by recurrence and resistance to outside stimuli that would typically redirect behavior, we should directly test whether these factors contribute to intrusive thought.

Sign Tracking, Goal Tracking, and the Interruption of Top-Down Control over Behavior

When rats are exposed to a Pavlovian-conditioned approach procedure, wherein an illuminated lever, a conditioned stimulus (CS), precedes the delivery of a food reward, an unconditioned stimulus (US), into an adjacent food cup, two distinct phenotypes may emerge (Flagel et al. 2009): *goal trackers* and *sign trackers*. Upon lever-CS presentation, goal trackers approach the location of reward delivery, whereas sign trackers approach and interact with the lever-CS itself.

For both sign trackers and goal trackers, the lever-CS is a predictor because it elicits a conditioned response. For sign trackers, however, the lever-CS also acquires incentive motivational value (incentive salience) and is transformed into a "motivational magnet" (Berridge and Robinson 2003). That is, for sign trackers, the lever-CS itself is attractive, elicits approach behavior, and acts as a conditioned reinforcer (Cardinal et al. 2002; Berridge and Robinson 2003; Flagel et al. 2009). Sign-tracking behavior can be considered compulsive because it will persist even if it results in omission of reward delivery, and is resistant to extinction (Tomie 1996; Flagel et al. 2009; Ahrens et al. 2016). Furthermore, relative to goal trackers, sign trackers are more impulsive (Lovic et al. 2011): they exhibit deficits in sustained attention (Paolone et al. 2013), show exaggerated responses to aversive stimuli (Morrow et al. 2011), and have an increased propensity for cue-induced reinstatement of drug-seeking behavior (Saunders and Robinson 2010). Recent evidence in rats (Eagle et al. 2014; Vousden et al., submitted) and humans (Albertella et al. 2019) indicates that sign trackers show greater levels of dysfunctional checking behavior of relevance to OCD.

The sign-tracker/goal-tracker animal model has been used to parse the neural mechanisms underlying two different learning strategies: predictive versus incentive learning. Sign tracking, or incentive learning, is dependent on dopamine in the NAc (Flagel et al. 2011b). In fact, using this model, it has been shown that the shift in dopamine in the NAc from the reward (US) to the cue

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

(CS) encodes the incentive value of the cue, not the predictive value (Flagel et al. 2011b). Relative to goal trackers, sign trackers show greater engagement of the cortico-thalamic-striatal "motive circuit" in response to a food cue (Flagel et al. 2011a). Within this circuit, the paraventricular nucleus of the thalamus (PVT) has emerged as a critical regulator of individual differences in cue-motivated behaviors (Haight and Flagel 2014; Haight et al. 2015; Kuhn et al. 2018).

The PVT is a midline thalamic nucleus ideally located to integrate cognitive, emotional, and arousal information from various areas of the brain and, in turn, to guide motivated behaviors (Kelley et al. 2005; Kirouac 2015). Specifically, the PVT receives dense input from the PFC, as well as subcortical areas, including brainstem nuclei such as the dorsal raphe and locus coeruleus and other areas such as the lateral hypothalamus and amygdala. The PVT then integrates this information and sends reciprocal output to some of the same regions, but also has dense glutamatergic projections to the shell of the NAc. In fact, the PVT can regulate dopamine release in the NAc, even in the absence of the ventral tegmental area (Parsons et al. 2007).

Within the context of the sign-tracker/goal-tracker animal model, neurons projecting from the prelimbic PFC to the PVT appear to encode the predictive value of reward cues, whereas subcortical systems surrounding the PVT encode the incentive value. Specifically, sign-tracking behavior is thought to result from hyperactivity of neurons projecting from the lateral hypothalamus to the PVT, and those projecting from the PVT to the NAc (Haight et al. 2017). The working hypothesis, therefore, is that cognitive representation, or the predictive value of the reward cue, is encoded in prelimbic PVC-PVT projecting neurons, and that this top-down process predominates in goal trackers. In sign trackers, however, where incentive learning prevails, subcortical processes are able to override this top-down mechanism. Thus, the PVT appears to act as a fulcrum between top-down cortical processes and bottom-up subcortical processes, and an imbalance between these processes may result in aberrant or psychopathological behavior. It is also intriguing, in light of our discussion of ACh systems above, that sign-tracking behavior in rats is associated with poor attentional control, mediated by an unresponsive basal forebrain cholinergic system (Kucinski et al. 2018). The neurobehavioral endophenotype of sign trackers may capture antecedents that predispose an individual to intrusive thoughts. For example, sign trackers appear to have an inherent imbalance between emotional and cognitive domains (with the PVT acting as a fulcrum between the two domains; see Figure 5.1), and this imbalance renders them more susceptible to behavioral control by intrusive experiences.

Avenues for Future Research

The concepts outlined above provide a number of avenues for future research. For example, the idea that we can capture and decode the neuronal ensembles

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

associated with disruptions of behavior and intrusive events suggests that we can use those neuronal ensembles to test whether activation of neurons, pathways, brain areas, or interregional networks recapitulate or disrupt intrusive events. A prototypical experiment might be to provoke an initial adaptive behavior, such as grooming in a rodent as a result of a sticky substance on its fur (e.g., peanut butter). During the adaptive behavior, molecular techniques could be used to capture active neurons that drive both the sensory representation of the sticky substance and the motor output. Driving the activity of these neuronal ensembles repeatedly should recapitulate the behavior and the sensory representation, without the feedback of the cleaning of the fur or a clear outcome of the motor behavior. Measuring the increasing connection between the sensory and motor systems could be achieved with electrophysiology, or behaviorally, by measuring the likelihood of grooming to other, less intrusive stimuli (i.e., stimuli that would not generally elicit a robust response). Generalization to other stimuli could also be measured. Finally, recruitment of circuitry related to anxiety or emotional behavior could be measured, which might result from the mismatch between the environment and the sensory perception or motor outcome. This hypothetical "peanut butter test" could be generalized to other behaviors in which a specific initial stimulus and its adaptive motor outcome are initially paired and then dysregulated, by driving the neuronal correlates of the event and subsequent outcome in the absence of appropriate feedback (e.g., cued fear memories, drug-associated stimuli). Capturing aberrant outcomes of neuronal recapitulation may identify common neural mechanisms, whether of plasticity or systems-level generalization, which could then be probed in patients with intrusive thought.

Conclusions

Intrusive thoughts are a hallmark of several psychiatric conditions (e.g., OCD, panic disorder, major depressive disorder, and addiction). For individuals suffering from these disorders, intrusive thoughts gain inordinate control over their emotions and actions, interfering with daily activities and disrupting lives. Given that intrusive thoughts are common to multiple mental illnesses, it is surprising how little we know about the underlying brain mechanisms. This gap in knowledge stems from the fact that we have not yet, as a field, tried to capture explicitly the commonalities of multiple disorders, such as intrusive thoughts. Here, we have highlighted standard animal models, behavioral tests, and outcome measures that could be exploited to shed light on the neurobiological components of intrusive thought. We defined intrusive thought within a biological framework that can be probed in the research laboratory, with resultant models optimized to yield novel therapeutic targets. We proposed a conceptual model that captures intrusive thoughts as an emergent property of multiple systems (emotional, cognitive, motor, and autonomic/somatic) that

From "Intrusive Thinking: From Molecules to Free Will," edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1

are represented in hubs throughout the brain. When the neural choreography between these hubs and their corresponding nodes becomes disrupted, there is a loss of homeostatic and/or cognitive control which leads to maladaptive thoughts and inappropriate behaviors. In the laboratory, with careful experimental design and multidimensional levels of analyses, we can model this loss of control on both behavioral and neural levels. Here, we provided a road map that illustrates multiple routes by which different approaches can be used in combination to expand our understanding of intrusive thought. This road map is not proscriptive; rather, we hope that it will serve as a foundation for a novel avenue of preclinical research to advance our knowledge and ultimately lead to more effective therapies for a number of psychiatric illnesses that are characterized by intrusive thoughts.

Acknowledgments

The authors wish to acknowledge support from the United States National Institutes of Health (grant numbers DA014242, DA044297, MH117103, MH007681, MH100023, OD011132, DA039687, AA024599, MH106428-5877, DA033396, MH111520, MH112355, MH045573, and MH106435).

From “Intrusive Thinking: From Molecules to Free Will,” edited by Peter W. Kalivas and Martin P. Paulus. Strüngmann Forum Reports, vol. 30, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978-0-262-54237-1