9

Neuropsychological Mechanisms of Intrusive Thinking

Renée M. Visser, Michael C. Anderson, Adam Aron,
Marie T. Banich, Kathleen T. Brady, Quentin J. M. Huys,
Marie-Hélène Monfils, Daniela Schiller, Florian Schlagenhauf,
Jonathan W. Schooler, and Trevor W. Robbins

Abstract

A classic definition of intrusive thinking is "any distinct, identifiable cognitive event that is unwanted, unintended, and recurrent. It interrupts the flow of thought, interferes in task performance, is associated with negative affect, and is difficult to control" (Clark 2005:4). While easy to understand and applicable to many cases, this definition does not seem to encompass the entire spectrum of intrusions. For example, intrusive thoughts may not always be experienced as unpleasant or unwanted, and may in some situations even be adaptive. This chapter revisits the definition of intrusive thinking, by systematically considering all the circumstances in which intrusions might occur, their manifestations across health and disorders, and develops an alternative, more inclusive definition of intrusions as being "interruptive, salient, experienced mental events." It proposes that clinical intrusive thinking differs from its nonclinical form with regard to frequency, intensity, and maladaptive reappraisal. Further, it discusses the neurocognitive processes underlying intrusive thinking and its control, including memory processes involved in action control, working memory and long-term memory encoding, retrieval, and suppression. As part of this, current methodologies used to study intrusive thinking are evaluated and areas are highlighted where more research and/or technical innovation is needed. It concludes with a discussion of the theoretical, therapeutic, and sociocultural implications of intrusive thinking and its control.

Group photos (top left to bottom right)   Renée Visser, Michael Anderson, Marie-Hélène Monfils, Trevor Robbins, Daniela Schiller, Marie Banich, Adam Aron, Kathleen Brady, Jonathan Schooler, Florian Schlagenhauf, Quentin Huys, Michael Anderson, Daniela Schiller, Jonathan Schooler, Kathleen Brady, Quentin Huys, Florian Schlagenhauf, Marie-Hélène Monfils, Trevor Robbins, Renée Visser, Marie Banich

Introduction

While reading this chapter, you might be sitting in a coffee shop somewhere or traveling on a train. The sunlight comes in through the window and you envision how warm the planet might get. You are having a meeting in a few hours and various scenarios of it are running through your head. Perhaps these few examples have triggered some of your own thoughts, as you are no longer reading this article: your eyes are just glazing over these words. If this just happened and you rejoin us a few sentences below, you have just experienced an intrusive thought.

We all seem to have an intuitive understanding of what an intrusion is. Yet, attempts to define it immediately sets off debate. According to the one textbook that has been written on this is, the classic definition of an intrusive thought is "any distinct, identifiable cognitive event that is unwanted, unintended, and recurrent. It interrupts the flow of thought, interferes in task performance, is associated with negative affect, and is difficult to control" (Clark 2005:4). Central to this definition is the assumption that an intrusive thought always constitutes an unpleasant experience which negatively impacts functioning. Here, we revisit the definition of intrusive thinking, by considering all the circumstances in which intrusions might occur, their manifestations across health and disorders, and their neurocognitive basis. We start with a rather narrow, presumably more commonly accepted definition of intrusions being *conscious, involuntary, unwanted thoughts,* and arrive at an alternative, more inclusive definition of intrusions as *interruptive, salient, experienced mental events.* We discuss current methodologies used to study intrusive thinking, highlighting existing strengths as well as areas where more research and/or technical innovation are needed. We conclude with a discussion of the theoretical, therapeutic, and societal implications of intrusive thinking and its control.

What Are the Everyday Manifestations of Intrusions and Their Control?

Although intrusive thoughts are often associated with mental health disorders, they also occur to healthy individuals in everyday life (Purdon and Clark 1993; Berntsen 1996). Identifying broad circumstances under which these thoughts occur to people in general, irrespective of mental health status, may highlight their adaptive functions in healthy individuals and illuminate the processes that generate such intrusions. Moreover, considering the motivations that people may have for controlling intrusions in daily life can shed light on important psychological and social functions that the capacity for control helps to support.

Types of Content That Drive Recurring Intrusive Thoughts in Healthy Individuals

Although unintended thoughts can occur for any content, not all such thoughts are intrusive and bothersome. There are, however, certain contents that consistently trigger intrusive thoughts in otherwise healthy people. Here we discuss several examples and consider why this content is intrusive.

Emotionally Salient Events.   Events that trigger intense emotions, such as psychological trauma, often lead to intrusive reminders, and the period of their intrusiveness extends for varying durations, depending on the event. Emotional events that trigger unwelcome intrusions are typically negative (e.g., anger, guilt, shame, sadness, fear, and embarrassment). However, positive events are also capable of triggering repetitive thoughts about pleasant memories, and this can be quite disruptive, especially when it interferes with focused attention that people need to perform certain tasks or activities (e.g., job responsibilities).

Incompletions.   When people initiate a process and are then unable to complete it, due to an interruption or some other impasse, thoughts related to that incompletion tend to recur until the process is completed (Horowitz 1975). This hypothesis is reflected in the classical Zeigarnik effect proposed in the early twentieth century by Gestalt psychologists, who posited that people had superior memory for interrupted processes (Zeigarnik 1938). Incompletions can include both physical and mental tasks, as long as there is an unresolved problem. An interesting possibility is that the tendency for incompletions to precipitate intrusive thoughts may be amplified by the salience or emotional intensity of the incomplete process (Horowitz 1975). Intrusive thoughts related to the interrupted process (i.e., seeking comprehension, solving a problem) may persist until the situation is understood. This suggests that emotionally intense events may not intrude merely due to salience or encoding strength, but because they are accompanied by a compelling desire to comprehend them.

Relatedly, when people reach a significant impasse in solving a difficult problem, they continue to "work on the problem" in the background. For example, research on creativity and insight problem solving suggests that when people allow a period of incubation, insights may emerge spontaneously (the "aha" phenomenon), as though a process has occurred in the background (Gable et al. 2019). The commonly experienced "tip of the tongue" phenomenon in memory retrieval (described initially by W. James and S. Freud) is another example of when a temporarily forgotten item finds its way back into consciousness unpredictably. The content of people's mind wandering (i.e., when thoughts distract us from a task at hand) often includes unresolved problems salient to the individual, whether emotionally significant or not (Smallwood and Schooler 2006; Klinger and Cox 2011).

Intentions.   A special case of "incompletions" worth distinguishing concerns incompletions arising from intended actions that must be deferred, an ability studied in research on prospective memory. This research indicates that although people often rely on the environment to remind them to perform intentions, or use intentional self-reminding strategies (e.g., lists), thoughts of the intended action can still pop into a person's mind, unbidden, particularly when intended actions have high importance or affect. Such cases suggest that intentions are maintained in an elevated state of accessibility, even without conscious rehearsal. Consistent with this possibility, deferred intentions often have a special active status in memory (outside of intentional rehearsal) that disrupts ongoing task performance, as illustrated by *intention interference* (Goschke and Kuhl 1993; Cohen et al. 2011; Bugg and Streeper 2019).

Anticipated Events.   Although intrusive thoughts often involve past events or general ideas, they also concern events that have yet to happen. The associated anticipation can carry emotionally positive, negative, or even both consequences. Under some circumstances, anticipated events elicit conflicting feelings of excitement and apprehension in parallel. For example, an upcoming biopsy may return a diagnosis of cancer or good health.

Uncertain Events.   When people are uncertain about a past or future event, this can promote intrusive thoughts for several reasons (Grupe and Nitschke 2013). For example, if someone is uncertain about whether they have already performed an action (e.g., locking the front door or taking one's pills), associated uncertainty may precipitate worry. Uncertainty about future outcomes can also result in persistently intruding thoughts, which in turn may prepare people for different outcomes, thus enabling them to be ready to respond appropriately.

Dissonant Facts, Events, or Beliefs.   When a new fact, experience, thought, or impulse conflicts with one's beliefs or self-image, the resulting dissonance creates a tension that must be resolved. For instance, if a person commits an action that conflicts with their self-image, the implications for how they should revise their self-perceptions can be distressing. Are they the good person they think they are or not? Dissonance creates conflict that can trigger recurring automatic thoughts until the conflict is resolved. The desire to resolve the mental discomfort and stress induced by such discrepancies suggests that people strive for psychological consistency, an idea first proposed by Leon Festinger (1962) in his theory of cognitive dissonance.

Frequent Events, Stimuli, or Ideas.   Not every intruding thought is emotional or related to incompletion or dissonance. Sometimes, the frequency of an event, image, or thought induces further repetitions. One striking example is the repeated hearing, in our imagination, of a song that we have recently heard on numerous occasions, known colloquially as "ear worms" (Hyman et al.

2013, 2015; Moeck et al. 2018). Research suggests that whether an ear worm develops is related to how often the song has recurred in recent experience.

Images.   As the foregoing examples illustrate, intrusive thoughts can occur in many representational formats. In some cases, intrusive thoughts can be verbal or propositional in nature, as might occur in persisting thoughts about incomplete processes, unsolved problems, deferred intentions, or incongruous, dissonant occurrences. In other cases, intrusive thoughts can be more sensory in character, including intrusions from any sensory modality. These mental images may play a special role in intrusive thinking, a possibility suggested by their prevalence in psychiatric disorders. Mental images are experiences of perception that occur in the absence of external sensory input (Kosslyn et al. 2006; Pearson et al. 2015). They are not limited to remembering the past, but can also include imagined future scenarios. While often benign, highly emotional images (e.g., of an upsetting event) can set off a cascade of other disruptive cognitive and emotional processes, including increased physiological responses and rumination (Lang 1977; Grey and Holmes 2008; Holmes and Mathews 2010; Ji et al. 2016; Holmes et al. 2017). Later, in discussions of clinical manifestations of intrusive thoughts, we revisit intrusive imagery in diverse psychiatric disorders.

External and Internal Factors Triggering Intrusive Thoughts in Healthy Individuals

The previous discussion reviews content that is especially prone to generate intrusive thoughts, but it is worth separately considering the conditions that generate intrusions. In general, the likelihood of intrusions occurring in a given moment is related to both the presence of retrieval cues as well as the availability of control resources to resist unwanted intrusions. Here we discuss the critical role of cue-driven retrieval, matching physiological and mood state, monitoring processes that may cue thoughts, and the availability of inhibitory control.

Cue-Driven Retrieval.   Intrusive thoughts are often triggered by associations to environmental cues (Berntsen 1996). The most obvious example arises when a stimulus elicits an unwelcome reminder of a past event. For instance, after an argument, a friend's face may remind someone of the altercation, or a song may bring back memories of a loved one who passed away. Although intrusive memories provide clear cases of cue-driven retrieval, other forms of intrusive thoughts, such as feared future events, or unpleasant ideas or images, also seem to be cue driven. When *the same content intrudes more than once*, the content is reemerging from memory of previous thoughts. One can frequently identify specific cues in the environment, or in a person's patterns of thinking, that drive retrieval of perseverative content (e.g., Gagnepain et al. 2014; Benoit et al. 2016).

Matching Mood and Physiological State.   Sometimes the cues that trigger retrieval are not specific stimuli in the world or concepts in memory, but rather are broad psychological or physiological states (Berntsen 1996). It is well established that information and experiences are often stored in memory in association with a representation of the context in which they took place (for a review, see Anderson and Hanslmayr 2014). This can include not only the spatiotemporal context (i.e., the environment), but also the mood, drug, or arousal state present at encoding. Later on, the chance of retrieving the content is higher when a person's state resembles the one at encoding. For example, experiences encoded in a sad or angry mood are more likely to be recalled at a later time, when a person is once again in a sad or angry mood than when they are in a happy mood.

Diminished Cognitive Control.   In addition to the particular cues at retrieval and their match to encoding, other elements of the state of the person at retrieval may influence the frequency and persistence of intrusive thoughts. An important category of state-related variables that can influence intrusion frequency is whether a person is suffering from diminished cognitive control that might otherwise help the person to limit involuntary retrievals. If mechanisms such as inhibitory control prevent or reduce intrusions, then anything that compromises these abilities is a risk factor for intrusiveness, even in healthy individuals.

Several factors common in healthy samples may give rise to such deficits in cognitive control, including sleep deprivation (Nilsson et al. 2005; Drummond et al. 2006), general fatigue, stress (Shields et al. 2016), lack of exercise (Hillman et al. 2008), and intoxication. Although such changes in state are temporary for most healthy individuals, chronic depletion of cognitive control, especially with several of the above factors contributing, could put healthy individuals on a trajectory to develop persistent intrusions and changes in mood that are clinically significant. In addition, consistent, trait-related deficits in cognitive control may give rise to significant risks in controlling intrusive memories and thoughts, and may be a risk factor for psychiatric disorders (Levy and Anderson 2008).

Desirability of Control

In healthy individuals, when intrusive thoughts occur, they are usually perceived as unwanted, at least at that moment. As a result, people often try to exclude them from awareness in an effort to regain control over thoughts and emotions. When successful, such efforts enable a person to put unwelcome thoughts out of mind, thus diminishing their accessibility in memory and reducing their tendency to return. Such attempts to facilitate the forgetting process often serve important behavioral, emotional, and social functions. Here, we consider several contexts in everyday life in which people are motivated to forget thoughts for a functional reason.

Concentration during Tasks. Sustained focus on a necessary task can be difficult when unrelated distracting thoughts intrude into awareness (Smallwood and Schooler 2006). Such thoughts can be of any valence. For example, receiving good news or bad news prior to leaving for work can undermine our focus. If distracting thoughts slow progress or increase the propensity for errors, it becomes essential and adaptive to put the distracting thoughts out of mind. Concentration is an extremely common selective attentional mechanism for trying to control intrusive thoughts, and successfully achieving concentration is clearly adaptive.

Executing High-Performance Cognitive and Motor Skills. Professional athletes are often under extreme pressure to perform to a very high standard, and when they do not, there can be significant consequences for themselves or their teammates. In sports psychology, a literature on the causes of "choking" under pressure (Beilock and Gonso 2008) has emerged to address the processes that lead an athlete's performance to deteriorate under pressure. This often includes the inability to overcome intrusive thoughts that undermine the focus needed for top performance.

Regulating Pain. Thoughts about pain, what it feels like, how uncomfortable it is, and worries about what it may mean often intrude persistently, even after the pain has resided. We consider people who can manage intrusive thoughts about their pain to be resilient and able to cope. In contrast, those who are overwhelmed by pain, who allow their discomforts to blossom into catastrophic thinking, not only pay a price for this distraction, their suffering becomes magnified and extended (Edwards et al. 2006). When confronted with a painful stimulus (e.g., the cold pressor task), people with a moderate history of adversity tolerate the pain longer and show less pain catastrophizing (the label given to excessive thoughts about the pain), relative to people who have no history of trauma (Seery et al. 2013). It appears, therefore, that in the face of physical discomfort, resilience emanates from the capacity to control intrusive thoughts about the painful stimulus, which not only reduces suffering but increases the ability to pursue other goals.

Regulating Affect. When unwelcome thoughts intrude, they often evoke unwelcome changes in emotional state. For example, being reminded of an argument may trigger anger; images of an upcoming doctor's visit may trigger fear or anxiety; or seeing the same car that your ex-partner used to drive may evoke sadness. As a result, intruding thoughts trigger mechanisms that regulate emotion to return a person to a neutral or positive state. Although these endogenously triggered emotions are common and contribute substantially to psychological disorders, they have been neglected in research on emotion regulation, which has focused on regulating affect triggered by external emotion-eliciting stimuli. Understanding how intrusive thoughts can be downregulated

in memory may provide critical insights into how people regulate the emotions that they elicit, contributing to successful *emotional homeostasis* (Engen and Anderson 2018).

Persisting in the Face of Failure.   When memories or thoughts of past failures dominate someone's thoughts, they can be disruptive. Intrusive thoughts about failure are unpleasant and undermine feelings of competence and control; this may lead a person to abandon their goals earlier than they should or to fail to improve their performance. When someone can set aside past failures, this enables them to improve their skills or knowledge or to find creative solutions to the problems that led them to fail. Persistence, especially dogged persistence in the face of challenges and setbacks, requires successful regulation of thoughts. If persistence is an adaptive trait enabling major personal achievements, the inability to control thoughts of failure can limit personal growth.

Protecting Self-Image.   People sometimes feel embarrassed or ashamed about things that they have done. Alternatively, others may say hurtful things that undermine self-confidence. Although people vary in how they respond to such events, they often try to put the unwelcome content out of mind. Correspondingly, they show greater forgetting for negative feedback, exhibiting a remarkable capacity to not remember their faults or misdeeds. Work on mnemic neglect (Sedikides and Green 2009; Sedikides et al. 2016) has established this pattern of forgetting and linked it to threats to a person's self-concept: recent threats to one's self-image are forgotten. Moreover, work on the positive self-illusion (Taylor and Brown 1988) has linked this tendency to improved mental and physical health: healthy individuals have a higher view of their capabilities than would be supported by observers. In contrast, depressed individuals, ironically, often have a more accurate view of their circumstances. Similarly, most individuals show a very powerful and replicable positivity bias in terms of the autobiographical memories that they ultimately retain (Walker et al. 2003), whereas depressed individuals very often show the reverse tendency.

Justifying Inappropriate Behavior.   Sometimes people commit unethical, immoral, or hurtful acts about which they feel shame. Despite this, most people like to think of themselves as decent, creating dissonance between one's beliefs and deeds. Intrusive thoughts about these discrepancies often develop and must be addressed. Research on ethical amnesia indicates that people often forget their ethical lapses (Kouchaki and Gino 2016; Stanley and De Brigard 2019), suggesting that people suppress these uncomfortable thoughts.

Maintaining Attitudes and Beliefs.   A person's political, religious, or personal beliefs are often resistant to contradictory evidence. When new facts contradict our beliefs, dissonance is created that must be resolved. This resolution often involves controlling intrusive thoughts to forget the inconvenient

information. For example, Republicans and Democrats show enhanced directed forgetting for attitude statements that are incongruent with their beliefs, compared with congruent statements (Waldum and Sahakyan 2012). Moreover, a person's memory can be shaped by selectively recounting an event (Cuc et al. 2007; Stone et al. 2012, 2013), a form of thought substitution (Benoit and Anderson 2012; Anderson 2001). Intriguingly, this can undermine memories of omitted facts, a phenomenon known as *socially shared retrieval-induced forgetting* (Cuc et al. 2007; Stone et al. 2012, 2013). The selective forgetting that people exhibit for facts that are incompatible with their beliefs suggests that healthy individuals control their thoughts to avoid the discomfort of inconsistency.

Forgiving Others and Maintaining Attachment.   Sometimes friends or relationship partners commit offenses that provoke anger, giving rise to persistent intrusive thoughts. Indeed, intrusive thoughts about anger can be intense for some people and can take considerable time to "get over," placing strain on personal relationships and requiring the ability to override thoughts and impulses originating from the anger. This ability is predicted by individual differences in inhibitory control, suggesting that overcoming intrusive thoughts involves, in part, control processes that suppress them in service of forgiveness. Consistent with this possibility, when people decide that an offense may be forgiven, it is easier to suppress intrusive thoughts about the offense, ultimately leading to worse memory (Noreen et al. 2014). Thus, suppressing intrusive thoughts of anger ultimately contributes to a healthy capacity to forgive and forget in social relationships.

Similar considerations apply when there is a need to maintain an attachment relationship with a parent, guardian, or powerful authority figure (e.g., a boss), which may be essential to survive or thrive in an environment. In such cases, the capacity to control recurring intrusive thoughts enables a person to maintain normal relationships despite this conflict and, in some cases, to forget about the discrepant content.

Mind Wandering: An Everyday Manifestation of Intrusive Thinking?

One interesting example of everyday intrusive thinking is the widely studied experience of mind wandering. Returning to our example at the beginning of this chapter, we are all familiar with suddenly realizing that while our eyes have been moving across the page, our minds have been temporarily sidetracked by thoughts unrelated to the text. Such a situation is known as *mind wandering* (Smallwood and Schooler 2006). Using thought sampling techniques in which individuals are intermittently queried as to whether their thoughts are engaged in the task at hand, research suggests that people mind wander up to fifty percent of their waking hours (Killingsworth and Gilbert 2010). The surprising frequency with which people engage in mind

wandering raises an intriguing question: Should mind wandering be viewed as everyday intrusive thinking? Let us consider the following aspects of this question.

Unwanted.   Mind wandering is often unwanted because it disrupts task performance across many cognitively demanding domains (for a review, see Smallwood and Schooler 2015). It is, for instance, a major source of car accidents (Gil-Jardiné et al. 2017). Yet while mind wandering interferes with current tasks, it can be useful for completing more distal goals, such as planning (Baird et al. 2011) and creative problem solving. For example, when a vigilance task was interposed between two trials of a creativity task, mind wandering interfered with the vigilance task but enhanced performance on the second round of the creativity task (Baird et al. 2012). In a diary study, both creative writers and physicists indicated that twenty percent of their ideas occurred outside of work (Gable et al. 2019). Thus, although mind wandering can interfere with the task at hand, it often contributes to progress on other problems.

Unintended.   Individuals frequently report mind wandering and are often not even aware that they are engaged in it, until they are caught, as in the experience sampling probe (Schooler et al. 2011). Nevertheless, individuals may sometimes deliberately abandon a task in favor of other thoughts (Seli et al. 2015). The lack of awareness that one is mind wandering (i.e., lack of meta-awareness; Schooler 2002) contributes to more disruptive mind wandering episodes (Schooler et al. 2011), which are neurologically distinct from on-task thinking (Christoff et al. 2009). Notably, lack of meta-awareness has been associated with mind wandering about unwanted thoughts (Baird et al. 2013a). This suggests that intrusive thoughts may similarly slip below the radar of meta-awareness (Takarangi et al. 2014).

Recurrent.   Although up to fifty percent of people's waking hours is spent mind wandering (making mind wandering a recurring form of thought), the content of mind wandering routinely varies from one episode to the next. Hence, whether mind wandering is recurrent depends on the level of analysis.

Associated with Negative Affect.   In their influential paper, Killingsworth and Gilbert (2010) hold that "a wandering mind is an unhappy mind." This characterization reflects the observation that when individuals mind wander, they are routinely less happy than when they are on task. Nevertheless, the degree to which episodes of mind wandering drive negative affect may be partly determined by content. For example, Franklin et al. (2013) found that participants were generally less happy when mind wandering than when on task, unless they were mind wandering about something interesting, in which case they were actually happier.

Difficulty of Control. Individuals' perceptions regarding the controllability of mind wandering varies considerably (Zedelius and Schooler 2017), and these perceptions influence its frequency and impact. Individuals who view mind wandering as outside of their control mind wander more often, and their performance is more disrupted than those who view it as a mental state they can manage. Collectively, evaluation of the relationship between mind wandering and the definition of intrusive thoughts indicates that mind wandering is often, but not always, a form of everyday intrusive thought.

Summary and Commentary

As the above discussion illustrates, intrusive thinking occurs in a range of circumstances in healthy individuals. Such thoughts do not, by themselves, indicate a mental health disorder but are a normal feature of life. Moreover, attempts to suppress intrusive thoughts are often not only highly successful in healthy individuals, they are in fact essential to achieving numerous social, emotional, and cognitive goals, including the successful regulation of emotional state (Engen and Anderson 2018). Not all healthy people are equally effective at controlling intrusive thoughts, however, and vulnerability to persistent intrusions is critical to understand. These observations highlight the importance of understanding when and how intrusive thoughts become pathological and contribute to psychological disorders.

What Are the Main Manifestations of Intrusive Thinking in Mental Health Disorders?

As indicated in the preceding sections, occasional intrusive thoughts are normal occurrences and even commonplace in certain circumstances. However, there are a number of mental health disorders for which intrusive thinking is a core symptom, becomes problematic, and can interfere with function. Some common disorders in which intrusive thinking is a critical element are briefly described below; for further detail, see Schlagenhauf et al., Banich, as well as Brewer et al. (this volume).

Posttraumatic Stress Disorder

Symptoms of intrusions in posttraumatic stress disorder (PTSD) include recurrent, unwanted, involuntary, and distressing memories of a traumatic event (intrusive memories); reliving the traumatic event viscerally, as if it were happening again (flashbacks); upsetting dreams or nightmares about a traumatic event; and severe emotional distress or physical reactions to cues that remind an individual of the event (e.g., the smell of burned rubber reminding someone of a car accident or a ski mask that triggers memory of an armed robbery).

Intrusive memories or flashbacks are often precipitated by environmental cues that remind an individual of the traumatic event. Flashbacks are immersive, causing individuals to "dissociate" from their bodies and feel that they are back in the situation in which the trauma occurred. It is not clear whether intrusions and flashbacks are qualitatively different phenomena or fall on a spectrum of severity in which flashbacks represent a severe form of intrusions. The intensity and content may be different, yet both are multimodal sensory images that are interruptive, salient, experienced, unwanted, involuntary, and often recurrent.

Obsessive-Compulsive Disorder and Related Disorders

In obsessive-compulsive disorder (OCD), intrusive means the feeling of "being out of control." This feeling is experienced in different phenomenological domains since OCD develops over time. OCD is a process with different clinical stages rather than one single stage. These stages follow a dialectic interaction in which an intrusive event elicits a response and the response amplifies the intrusive event. Through different neurobiological adaptations, the course of OCD eventually worsens. Thus, OCD should be regarded as a disease process that develops through the amplifying interaction between (the reflection and resistance of) the person (mind) and the disorder (brain). The following example describes how intrusions may develop in OCD.

A young mother recently gave birth for the first time and now carries the responsibility of brand-new motherhood. Her partner leaves daily for work, and thus she is home alone with the child. Viewing her young baby in the crib, a thought appears in her mind: she imagines that she could strangle her baby in the crib and that no one is there to prevent her. The mere presence of the thought is intrusive because it occurs beyond her free will. She feels out of control and is unable to control her thinking. She is worried because the idea of strangling her baby does not fit her ideal of motherhood. The content of the thought is intrusive and ego-dystonic because it is not in line with her self-image. She wonders whether she could really strangle her baby and begins to feel out of control. How could she be certain that she would not strangle her baby, given the fact that humans are notoriously unpredictable? As the implication of the thought is intrusive, she feels anxious because the thought confronts her with being out of control.

The emotional value (anxiety) of the thought is intrusive. The presence, content, implication, and emotional value of the thought all have an intrusive quality. She actively resists the thought because it annoys her and feels intrusive. The process of reflecting or resisting, however, serves to reinforce the frequency and intrusive strength of the thought. The thought becomes obsessional. Her attention is completely drawn to that one single thought. Obsessionality is a dysfunction of intentionality: the incapacity to shift focus or attention to another topic due to a stronger and longer intentional relation

with the mental act. The thought is intrusive because of its obsessive nature. She cannot suppress the thought. Moreover, she is compelled to think about her obsession. Compulsivity is a dysfunction of sense of agency: she is forced to think about the intrusion, contrary to her willpower. The thought is intrusive because of its compulsive nature. Gradually the thought becomes more present and repetitive; it loses its original meaning, but becomes intrusive because of its duration and repetition. The thought has become a full-blown obsession. Obsessions are answered with compulsions (e.g., obsessing about germs could lead to compulsive handwashing). Though initially successful in reducing anxiety, gradually they become intrusive since the acts have to be performed compulsively. Note that both obsessions and compulsions are intrusive, and both have an obsessional and compulsive quality. Eventually, the anticipatory power of the intrusion is so overwhelming that reality testing gets disturbed. She does not know anymore whether she has or has not strangled her baby. Thoughts may take on a delusion-like character, with psychotic features.

Substance Use Disorder

Intrusive thinking in individuals with substance use disorders (SUDs) may include planning to procure the drug, recall of the experience related to its use, as well as the anxiety related to lack of access to or possession of the drug. However, there is large variability between patients, dependent on disease stage. In the initial stages, the individual may perceive thoughts of the drug as non-intrusive and innocuous or even pleasurable, at least as long as they believe to be in control of drug intake, and that such thoughts do not significantly interfere with daily activities. Once the disease progresses toward more severe stages, the individual begins to perceive thoughts of the drug as intrusions. This usually happens when the individual realizes that they have lost the ability to make decisions regarding whether or not to obtain and/or take the drug.

In individuals with SUD, the onset of intrusive thinking varies widely. Intrusive thoughts may appear during acute withdrawal from the drug, or years after the last exposure to the drug. Furthermore, they can be elicited by an internal cue (e.g., a memory of an event related to the drug, or a particularly emotionally negative moment while the patient is alone at home) or external cues, such as other people taking the drug, movie scenes, or physical cues around the patient (e.g., restaurants, bars, or images related to the drug or addiction).

In conclusion, intrusive thinking is a major component of SUD as well as in other addictions, such as pathological gambling. It occurs throughout every step of the life cycle of SUDs and, in some cases, can be the *primus movens* of a series of actions and events leading to relapse and drug taking.

Mood Disorders

Clinically, intrusive thoughts are neither a core nor a necessary feature of mood disorders. However, they appear regularly in a manner that has important bearing on its treatment, severity, and outcome, specifically in the form of rumination, suicidal thoughts, negative automatic thoughts and "flight of ideas."

Rumination is a frequently observed cognitive feature of depression, involving the repetitive and persistent focusing on the causes of the current state of distress and its likely consequences. Depressive rumination is typically centered around personal shortcomings, faults, failings, and mistakes (Treynor et al. 2003). Rumination has intrusive features: it is highly interruptive and distracts individuals from engaging in other tasks. While individuals often state they do not want to ruminate, they also hold metacognitive beliefs that it is important to do so (Borkovec and Roemer 1995). Rumination is associated with the onset of depression and, when combined with negative cognitive styles, predicts the duration of depressive symptoms (Nolen-Hoeksema et al. 2008). The extent to which it is hence desired and indeed under volitional control is unclear.

Individuals suffering from suicidal thoughts report at times imagery related to potential ways of committing suicide (Holmes et al. 2007). This imagery can be experienced as unwanted, intrusive, and interruptive.

Negative automatic thoughts are regarded as a central feature of depression in cognitive models of depression (Beck 1976). In this view, events that are related in some form to core negative beliefs or schemata can trigger fast interpretations or "automatic thoughts," which, due to their negative nature, promote a depressed, negative mood. These automatic thoughts appear very rapidly and can profoundly influence behavior. However, as these automatic thoughts closely relate to core beliefs, they tend not to be experienced as intrusive or unwanted.

Individuals with manic episodes of bipolar disorder can display a "flight of ideas," whereby they exhibit a rapid sequence of unrelated thoughts. This state is often experienced in the context of mania and tends to be positively experienced as a phase of heightened creativity. In mixed states, however, it can also be perceived negatively and exhibit loss of control of one's thinking. Hence, although intrusive thoughts are not central to mood disorders, they do feature prominently and in areas that are thought to be closely related to the core mechanisms of the illnesses.

Anxiety

Two subtypes of anxiety have been distinguished: anxious arousal (e.g., panic) and anxious apprehension (e.g., worry). In cases of anxious arousal, the individual has fearful reactions and thoughts that lead to somatic symptoms and/or somatic symptoms are interpreted in a fearful manner. For example,

in simple phobias, an individual will see a phobic object (e.g., a spider) and have intrusive thoughts regarding that object (e.g., "The spider is about to crawl on me and bite me") that are often accompanied by somatic symptoms (e.g., sweating, increased heart rate). In cases of panic disorder, the individual has somatic symptoms, such as increased heart rate and dizziness, associated with intrusive thoughts (e.g., "I am having a heart attack and am going to die"). This situation of a bodily state leading to intrusive thoughts is somewhat akin to SUDs, in which a somatic state (i.e., withdrawal) can lead to intrusive thoughts of craving.

In contrast, in cases of anxious apprehension, intrusive thoughts tend to be related to a future event: about the feared event itself, ways to avoid the fear, or the discomfort/harm that might be associated with the event. In some cases, intrusive thoughts can be more distressful when a person anticipates the event, than during the event itself. This appears to be associated with the fact that individuals with anxious apprehension have an intolerance of uncertainty (e.g., not being 100% sure that the flight will not involve severe turbulence).

Individuals that experience anxious arousal and/or anxious apprehension may know that their fears are likely unrealistic or overblown. This knowledge, occurring simultaneously with the experienced fear and intrusive thought, can cause distress and lead to additional intrusive thoughts related to poor self-evaluation similar to depressive rumination: "What is so wrong with me that I cannot get on an airplane like everyone else? Why am I such a baby?" However, intrusive thoughts associated with anxiety tend to be more future oriented whereas intrusive thoughts associated with depression tend to be more concerned with past events.

Psychosis

Psychosis is a state characterized by hallucinations and delusions, which is common in schizophrenia spectrum disorders but also observed transdiagnostically. Hallucinations are false perceptions, mostly experienced with similar sensory quality, as if originating from an outside stimulus but without an external stimulus being present. For example, a patient with schizophrenia might repeatedly experience hearing voices that are negatively commenting on what they are doing and perceive the voices to come from outside their head. Auditory verbal hallucinations of this type can repeatedly occur and severely interfere with daily life, although frequency and distress vary.

Delusions are false beliefs held with high subjective certainty and confidence despite contrary evidence. Delusional ideation is very salient for the patient and center on the person. For example, common topics are delusions of reference or prosecution. Deluded persons do not experience the delusional ideation as interruptive to one's thoughts. Delusional ideations are part of the person's belief system and thus are ego-syntonic; still, they can be disruptive

for the person's relation to their interpersonal and social surrounding. A special case is delusion of control such as thought insertions, when (mostly verbal) thoughts are experienced as being "inserted" by another agent into one's head. Inserted thoughts can be salient, interruptive to the train of thought, and unwanted, but are attributed to an outside agent rather than to oneself, and therefore cannot be experienced as ego-syntonic nor ego-dystonic.

Attention Deficit Hyperactivity Disorder

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with everyday functioning (American Psychiatric Association 2013). People with ADHD are highly distracted by things happening in the outside world as well as by internal thoughts. The issue is whether these distracting internal thoughts should be classified as intrusions. To an external observer, they are certainly disruptive and maladaptive. Compared to other disorders, however, the intrusions may, on average, be much less salient since an individual with ADHD can be distracted by any event, salient or trivial. The distraction can be relatively diffuse, and thoughts may not have specific, recurrent content. While certainly maladaptive and poorly controlled, these thoughts are also not necessarily unwanted; they may function as a welcome distraction (e.g., when feeling bored). An experience sampling study found that individuals with ADHD symptoms experienced excessive disruptive mind wandering, together with little meta-awareness on how to regulate this (Franklin et al. 2017). In general, one could ask whether having an exceptionally low threshold for internal and external distraction might render irrelevant the concept of intrusions; in such cases, the inability to focus and follow goals (i.e., stay on task) changes the goal hierarchy. Thought meandering becomes the default; there is simply no focused thought process upon which this meandering intrudes. This lower threshold could make people with ADHD more vulnerable to other disorders and the experience of intrusions, in general (Abramovitch and Schweiger 2009).

Revisiting the Definition of Intrusive Thinking: A Synthesis

From the description of all the possible manifestations of intrusive thinking in everyday life, as well as across clinical syndromes, a number of features can be distilled that best capture what intrusions are. Each of these features will be discussed below. From this, we select what we think are the key features of intrusive thinking to arrive at two definitions: one narrow, probably more typical definition of intrusions, and a broader, more inclusive one.

Consciousness.   To what degree are intrusions conscious? Is it possible to have unconscious intrusions? It is hard to conceive of an example where an intrusion is unconscious. However, the term consciousness is often confused with meta-consciousness (Schooler 2002); that is, being aware that something is an intrusion rather than merely experiencing it. Because of this ambiguity, experience may be a less controversial term, though arguably more abstract. In addition, while intrusions can encompass a range of unexpected events or actions, intrusive thinking explicitly refers to cognition. This may be incompatible with nonhuman animal models and may also overestimate how voluntary our usual thoughts are (see also Liu and Lau, this volume). Instead of conscious thought, we suggest the broader term experienced mental events.

Unwantedness/desirability.   Although intrusions are not necessarily negative in content, they may be unwanted in the moment, as they distract from the task at hand. The question then is: Does it matter what the task at hand is? Are intrusions during daydreaming unwanted, as long as they have no clear negative content? Can something even be identified as an intrusive thought if it occurs during daydreaming? One can even think of examples of clear pathological intrusions that are not necessarily unwanted. In depression, for instance, rumination is experienced as a strategy to solve a problem. In some cases, it reflects long-term aspirations or a society that makes something unwanted. It boils down to the question of what wanting is; different individuals (or even different voices within an individual) interpret what is wanted differently in the moment. Because unwantedness may not be a defining feature in every instance of intrusive thinking, it seems necessary to develop a narrower clinical definition that is more closely related to the patient's experience of being out of control.

Involuntary/Controllability.   Generally we think of intrusions as being involuntary, though this can mean different things. Involuntary can refer to the fact that a thought was unintended or uncontrollable once it appears. Are thoughts that are unbidden, but easily dismissed, intrusive? We suggest including involuntary in a narrow definition of intrusive thoughts, but not in a broader one (see below).

Disruptiveness.   While unwantedness and controllability may not be a defining feature of every instance of intrusive thinking, a necessary characteristic is that intrusive mental events interrupt and disrupt current cognition. By definition, they intrude upon ongoing processes (e.g., a task or a gentle, natural cadence of unconstrained thought). The disruptiveness may, however, not be experienced as such by the person with intrusive thoughts, though clearly it has maladaptive consequences to an external critic.

Salience.   Something can only interrupt if it has gravitational pull; that is, if it captures attention (see also Fedota and Stein, this volume). Such gravitational pull may lie in different, not mutually exclusive, aspects (e.g., its valence,

highly negative or positive), its vividness, its novelty, or its incongruence with a current state.

Valence.   Inherent to Clark's definition (Clark 2005:4) is the assumption that intrusive thoughts are unpleasant, but this is not always the case. Examples of potentially positive intrusive thoughts may include being in love, or experiencing feelings of grandeur (as occurs during manic or psychotic episodes), or when one has an insight ("aha" moments). Therefore, we suggest characterizing intrusive events as salient to the person, but not necessarily as negative in valence.

Content and Shape.   Intrusions can take different shapes: they can present as verbal thoughts, slips of action, or mental images. Content varies in valence as well as in time frame, ranging from the past, present, or future. We suggest not including content or shape as a defining feature of intrusive mental events as it might vary widely between different instantiations of intrusive thinking in healthy and clinical states.

Punctate versus Extended.   Intrusions usually appear unexpectedly, with a sudden onset and a limited duration. They can, however, last for an extended period of time, as in the case of an uncontrolled flashback episodes in PTSD, or if we include delusions, mind wandering, worry, craving, and rumination as forms of intrusive thinking. In these latter cases, one could argue that only the brief, initial episode is the intrusion that sets in motion a cascade of secondary cognitive processes. Intrusive suggests that something intrudes upon something; if these secondary cognitive processes continue for an extended period of time, they become the primary cognitive process, perhaps even changing the goal hierarchy. Should such a primary cognitive process be labeled intrusive thinking?

Recurrence.   While an ongoing process should perhaps not be labeled intrusive thinking, intrusions can repeat themselves and still qualify as intrusions. In fact, a key feature of intrusive memories in PTSD and other clinical syndromes is their recurrent nature. It may be the frequency as well as the recurrent content that distinguishes clinical intrusions from nonclinical intrusions.

Trigger.   As described at the outset, intrusions can be triggered by external cues, such as certain sights or smells or through internal cues such as mood states. However, intrusions are more than reflexive orientation to salient stimuli: the stimulus-evoked mental event has to interrupt the ongoing process significantly.

Agency.   As described above, to be defined as intrusive, thoughts must be attributed to oneself rather than an external agency.

In conclusion, a commonly accepted definition of intrusions may be that they are *conscious, involuntary, unwanted thoughts*. While easy to understand and applicable to many cases, this definition may mean different things to different people and not encompass the entire spectrum of intrusions. An alternative, more inclusive definition may be that of intrusions being *interruptive, salient, experienced mental events* that generally recur, particularly in clinical syndromes. While the latter definition encompasses the clinical definition of intrusions, it does not differentiate nonclinical from clinical phenomena. Clinical intrusive thinking differs from its nonclinical form with regard to frequency, intensity, and maladaptive reappraisal. Intrusive thinking with clinical significance is often part of a clinical syndrome and results in distress and disability as a general identifier of mental illness. In Table 9.1, we evaluate to what degree each of these key features for both definitions is manifested across symptom clusters; for a slightly different conceptualization of intrusions in pathology, see Monfils and Buss (this volume).

From Table 9.1, it is clear that symptoms occurring in PTSD and OCD best exemplify the construct of intrusive thinking. Less prototypical examples of intrusive thinking are flash-forward events, drug craving, suicidal ideation, rumination, and worry. In these instances, intrusions may not always be involuntary, unwanted, or clearly interrupt a task at hand. Rather, patients may voluntarily engage in these thoughts, for example, as a way to dampen present agony (suicidal ideation) or to mentally expose oneself to a feared situation in the future (worry), thereby reducing the risk of being overtaken by panic unexpectedly. In drug craving, these thoughts are only unwanted and interruptive to the degree that an individual intends to abstain from using. It is important to note that while *worry* is part of the DSM-5 criteria for generalized anxiety disorder and as such is a real symptom, rumination is considered a risk factor for depression, but not a diagnostic criterion. However, both constitute repetitive negative thoughts, with common underlying features as well as unique features (e.g., Hur et al. 2017), and seem to qualify as intrusive thinking.

It is less clear to what degree symptoms in ADHD qualify as intrusive thinking and, relatedly, mind wandering in healthy individuals. Typical daily thought meandering certainly does not elicit the derailing effects that clinical intrusions do and may not even be interruptive. Impulsive actions and tics do not seem to qualify as they are not clear experienced mental events (i.e., they do not necessarily involve awareness). Finally, it is unclear to what degree hallucinations, thought insertions, or delusions should be considered examples of intrusive thinking. These thoughts may not generate the same feeling of uncontrollability and intrusiveness, and an individual may not even realize that these are intrusive thoughts, such as when an individual truly believes that the police or secret service is monitoring their thoughts, or when someone is getting commands via imagined voices. Hallucinations and thought insertions

Table 9.1 Features of intrusive thinking across symptom clusters.

Symptom	Related disorder	Thought			Mental event			Further specifier
		Conscious	Involuntary	Unwanted	Experienced	Salient	Interruptive	
Intrusive memories	PTSD	x	x	x	x	x	x	Recurrent, past oriented
Flashback	PTSD	x	x	x	x	x	x	Recurrent, past oriented
Flash-forward	Bipolar disorder	x	?	?	x	x	?	Future oriented
Obsessions	OCD	x	x	x	x	x	x	Recurrent
Compulsions	OCD	x	x	x	x	x	x	Recurrent
Drug craving	SUD	x	x	?	x	x	?	Depends on motivation for abstinence; recurrent, future oriented
Pathological worry	Anxiety	x	?	?	x	x	?	Future oriented
Depressive rumination	Depression	x	?	?	x	x	?	Past oriented
Suicidal thought	Depression	x	?	?	x	x	?	Future oriented
Thought insertions	Psychosis	x	?	?	x	x	?	No agency, attributed to outside agent
Auditory hallucinations	Psychosis	x	?	?	x	x	?	No agency, attributed to outside agent
Delusions	Psychosis, depression, mania	x	–	–	x	x	–	No agency
Inattention	ADHD	x	x	?	x	–	x	Relatively diffuse
Impulsive action	ADHD	x	x	?	–	?	x	
Mind wandering	ADHD	x	?	?	x	?	?	
Tic	Tourette syndrome	–	x	x	–	?	x	Recurrent
Urge	Tourette syndrome	x	x	?	x	x	x	Recurrent

are usually attributed to an outside agent. In our definition of an "experienced mental event," a sense of agency is implied.

How to Best Study the Processes Underlying Intrusive Thinking and Its Control

In this section, we consider a number of general mechanisms that underlie the likelihood for a person to experience intrusive thoughts, and how they are

assessed. As suggested above, situations that trigger intrusions all tap into mechanisms related to the ability to control thoughts, actions, and mnemonic processes. Most of these processes come into play at the time of (involuntary) memory retrieval or when habits or compulsions interfere with goal-directed behavior. Here, we elaborate on what is known about these mechanisms and the methods to investigate them.

Cognitive Control Processes

Action Control

Goal-directed actions are those that we acquire to bring about a change in the world that accords with our basic desires. In this view, I perform an action, *A*, because I want a certain outcome or goal, *O*, and believe that *A* is the way to achieve *O*. Action *A* may start life under goal-directed control; however, with extended training performance, it can become more automatic and invariant, elicited by environmental stimuli, instead of to achieve specific consequences, thus becoming habitual. Both actions and habits are acquired and represented in parallel rather than serially. Therefore, it is possible for certain errors to emerge as a consequence of two controllers attempting to control action simultaneously: suddenly switching from goal-directed to habitual control can, for instance, result in the intrusive performance of an action that is adaptive in another situation, but not in the present situation where it is unwanted (e.g., continually turning on the windscreen wipers rather than the blinkers in a rental car). Conversely, intrusive goal-directed control can disrupt the smooth operation of a well-learned habit. It is well known, for instance, that suddenly exerting goal-directed control while playing a complex piece of music on the piano can result in reduced capacity to perform the piece accurately. The same holds for well-trained athletes (see Balleine, this volume).

Both actions and habits require top-down control to prevent excessive and hence maladaptive impulsive behavior. Clues about neural systems supporting the regulation of intrusive thoughts come from the field of action control, capitalizing on putative similarities between inhibition of motor responses and thought control. In the simple Stop-Signal paradigm, for each trial, subjects get ready to respond when a Go signal occurs; then, on a minority of trials, they try to stop in response to a subsequent Stop signal (Lappin and Eriksen 1966; Verbruggen and Logan 2008). Anderson and Green (2001) have long suggested that stopping a memory or thought from intruding is like stopping a motor response. When the stopping process fails, an intrusion occurs. Several studies have provided evidence consistent with this, although it is not yet certain whether there is a common process, common principle, or common circuitry (e.g., Morein-Zamir et al. 2010; Depue et al. 2016; Guo et al. 2018; Castiglione et al. 2019). Assuming there *is* a commonality between the stopping of a motor response and the termination of a memory or thought, it may be

useful to adopt a distinction from the motor stopping literature between global suppression and selective suppression (Aron 2011). Stopping a memory or thought can be achieved by a global systemic suppression mechanism, which suppresses all memories (Hulbert and Anderson 2018), or through a selective stopping process in which the target is suppressed by retrieving an alternative memory (Hertel and Calcaterra 2005; Bergström et al. 2009; Benoit and Anderson 2012).

Pavlovian-instrumental transfer (PIT) provides another experimental paradigm relevant to intrusive thinking and may provide evidence of intrusive thinking in nonhuman animals as well as humans. In the case of positive PIT, experimental animals (or humans; Freeman et al. 2014) are trained on two actions to earn distinct reward outcomes: A1 → O1 and A2 → O2. They are then exposed to two stimuli paired with the same outcomes but without the opportunity to perform the actions: S1 – O1, S2 – O2. They are then allowed to perform the two actions in extinction (without reward) but in the presence of the two stimuli. In this situation, the animals will perform the two actions at a low baseline rate. When one of the other stimuli is presented, however, this leads to an interruption in the animals' performance and causes them to immediately shift performance to the action which, in training, earned the goal or outcome predicted by the stimulus (specific PIT). General PIT occurs when the stimulus reminds the subject instead of the valence of the goal and similarly may enhance responding nonspecifically. Probably even more relevant to intrusive thinking is negative or aversive PIT in which the occurrence of a negative Pavlovian conditioned stimulus (associated, e.g., with painful electric shock or loud white noise) enhances avoidance or withdrawal behavior or disrupts appetitive behavior, as in conditioned suppression.

These findings suggest that the presentation of the stimulus during the task might *intrude* on the animals' ability to perform their continued action. In this sense, the stimulus might be regarded as an intrusion, bringing to mind an outcome. The occurrence of the interruptive stimulus reminds the animal of its associated action, the performance of which provides a readout of that intrusion. PIT is not influenced by extinction of the predictive learning nor by the devaluation of the instrumental outcome. In these cases, the stimuli continue to drive the performance of a specific action, demonstrating continued stimulus-mediated retrieval of the specific, now unwanted, outcome.

Working Memory

While much research relevant to intrusive thoughts, such as those that occur in PTSD, has focused on long-term memory formation and retrieval (the topic of the next sections), significantly less work has been done on the role that working memory mechanisms may play in intrusive thoughts. Yet, by their very nature, intrusive thoughts can only be *intrusive* to the degree that they can gain access to working memory and hijack an individual's attention. Working

memory is both a capacity-limited and time-limited buffer that allows information to be held online in service of current goals or objectives. As with long-term memory, there are distinct general processes that may be relevant to intrusive thoughts. Each of these processes is discussed in more detail below as well as by Banich (this volume).

The first process allows an intrusive thought to gain access to working memory or to be gated into working memory. Thoughts may be intrusive when they have priority for access to working memory. Information that enters working memory may be drawn from the external or the internal environment (e.g., current thoughts, long-term semantic or episodic memory, habitual actions), and may do so via its salience. If the information is drawn from the external world, it may be salient due to perceptual characteristics; if drawn from either the external world or internal milieu, it may be salient due to conceptual or abstract characteristics. For example, negative information may be quite salient for individuals who are depressed, and drug-related information may be quite salient for those with SUDs. Information may also gain access to working memory through a more controlled selection process (Feldmann-Wüstefeld and Vogel 2019). Individual differences in the efficiency of such selection processes have been observed (Vogel et al. 2005), raising the possibility that processes which allow for selection of information into working memory are altered in individuals who experience intrusive thoughts.

In the second process, an intrusive thought needs to be given priority within working memory so that it dominates current thought and actions. Thoughts may be intrusive when they are stronger and "out compete" other items within working memory, so being at the focus of attention even when they are not relevant for the goals or processes at hand. Generally, it is assumed that this selection process requires cognitive control, and one that can be examined, for example, using the emotional Stroop paradigm in which individuals are shown a bivalent stimulus: one feature is task relevant and should guide responding, the other is not. For instance, an emotional word (e.g., "joy") is overlaid on a face with an emotional expression, and individuals must respond based on the emotional valence of the word (positive, negative). When the word and the face are incongruent (e.g., the word "joy" on an angry face), reaction time and error rates are higher than when the items are congruent (e.g., the word "joy" on a smiling face). This effect is thought to arise because the task-irrelevant information, the face, is salient—so much so that it interferes with processing of the task-relevant information (in this case, the word). At least some part of the interference observed for incongruent as compared to congruent trials is thought to arise at the level of working memory representations (e.g., Banich 2009). In this paradigm, the salient information is provided by its stimulus characteristics (i.e., being a word not a face). In other cases, such as the "recent negatives," task information can be selected based on its temporal tag. Thus, individuals are shown a series of four items on a screen and, after a brief delay, a probe prompts them to decide if that item appeared in the immediately

preceding set of four. On certain trials, the probe is one of the four items from the prior trial, but not the current one. In such cases, individuals show reductions in performance compared to a less recent probe. This paradigm demonstrates that time can be a marker for controlling what information is within the current focus on attention. Future research might explore the degree to which selection (either on the basis of perceptual, topical, or temporal characteristics) is affected in individuals who experience intrusive thoughts. To the degree that these control mechanisms are defective, they may not allow for intrusive thoughts to be moved from the focus of attention, or manipulated or reintegrated in such a way as to reduce their saliency (e.g., via reappraisal).

In the third process, information must be removed at some point from working memory to allow more relevant information to be placed in this limited buffer. Theoretically it has been argued that removal may occur via three potential mechanisms: (a) passive decay, (b) being replaced by something else, or (c) being actively removed (Lewis-Peacock et al. 2018). Thoughts may be intrusive to the degree that they are particularly resistant to such processes. These processes may be especially important for disorders of intrusive thoughts which are more abstractly cognitive (i.e., not sensory based) and more repetitive in nature. Moreover, in certain cases it may just be specific categories of information that are specifically resistant to removal. For example, the degree of rumination observed in depressed individuals is associated with difficulty in removing negatively biased information from working memory (Joormann and Gotlib 2008). This difficulty may occur because of an entrenched overriding schema for thought that focuses attention to negative material.

Long-Term Memory

Many intrusive thoughts concern contents that have been encoded into long-term memory. This content includes not only personal autobiographical memories of past events, but also other contents that clearly rely on declarative memory but are not autobiographical memories per se (e.g., images of feared future scenarios or memories of prior thoughts). As noted earlier in this chapter, these intrusive thoughts are often elicited by retrieval cues in the environment or, in some cases, cues that are internally generated by the person. While sometimes positive or neutral, intrusions often have a negative tone. In their extreme form, intrusive memories involve vivid, multisensory images from highly aversive events, constituting the hallmark symptom of disorders such as PTSD. When looking at the role of mnemonic processes underlying intrusive memories and opportunities to intervene, there are different stages of memory that need to be considered. A distinction can be made between the encoding of a memory and its retrieval: within the context of encoding, processes of encoding suppression can limit the impact of an event; within the context of retrieval, one can focus on opportunities for memory modification as well as mechanisms for suppression. Critically, intrusive thoughts involving content encoded into long-term memory

(whether autobiographical, future scenarios, or memories of prior thoughts) can be conceptualized as involuntary retrieval episodes.

Memory Encoding

Some evidence suggests that the later intrusiveness of an emotional memory is influenced by processes at the time of encoding and shortly thereafter. Although emotional intensity contributes to recurring intrusive thoughts for a variety of reasons, the impact of emotion on both the strength and nature of the memory that is encoded may be especially important. As reviewed later in the chapter (see section, What Are the Neural Systems Relevant for Intrusions and Their Control?), abundant evidence has documented interactions between the amygdala and the hippocampus during emotional events, putatively contributing to a strong encoding that makes a memory more prone to later retrieval. Indeed, one view is that intrusive emotional memories are simply stronger memories, making them more accessible and easily triggered (Berntsen and Rubin 2008, 2013; Rubin et al. 2016b). In line with this, emotional memories are typically vivid and are often particularly resistant to the passive processes that usually lead memories to be forgotten over time (Hamann 2001); active forgetting processes, however, may behave differently, as discussed next. Another view is that memories for the specific episodes of which elements intrude may actually be worse; that is, they are less contextualized and therefore more fragmented (Brewin 2016; for a challenge to this view, see Rubin et al. 2016a; Bisby and Burgess 2017). This has led to the proposal of a "dual representation" of traumatic memories, with a hippocampus-based system underlying the neutral, declarative aspects of a memory, and a system involving the amygdala and sensory areas representing sensory, emotional aspects of a memory (Brewin 2014; Bisby and Burgess 2017). Although this may be an oversimplification, there is evidence that the intrusiveness of a memory can be selectively reduced while leaving the voluntary recall of a memory intact (Holmes et al. 2009, 2010; James et al. 2015; Gagnepain et al. 2017; Lau-Zhu et al. 2019) and vice versa (Bourne et al. 2010). This finding is compatible with the possibility that these types of memories may rely on distinct systems (Visser et al. 2018), or at least distinct representations (Gagnepain et al. 2017).

Whereas emotional intensity at encoding can amplify a memory's intrusiveness, this can often be mitigated by control processes that limit encoding and consolidation. For example, studies using experimental models for trauma show that engaging in a competing visuospatial task, during or shortly after experiencing an event, can reduce the frequency of intrusions of that event over the subsequent week, as evidenced through self-reports in a subject's daily diary (Holmes et al. 2004; 2010; Lau-Zhu et al. 2019). In addition, a sizeable body of work on item-method directed forgetting shows that when people are instructed to forget the immediately preceding memory item

in a list, the instruction has a large impact on later retention of the memory, rendering to-be-forgotten items significantly less accessible, even for emotionally unpleasant scenes (Anderson and Hanslmayr 2014). Such directed forgetting effects arise not only on recall tests, but also on recognition and implicit memory tasks, indicating that participants can successfully block the encoding of recent events into long-term memory when motivated to do so. This encoding suppression mechanism is likely one coping process that healthy individuals use to limit the impact of upsetting events, reducing their intrusiveness (Anderson and Hanslmayr 2014). It may also be the key process underlying the previously discussed phenomenon of mnemic neglect (Sedikides et al. 2016).

Memory Retrieval: Modification

Sometimes the encoding of unwanted content occurs despite a person's efforts to prevent it. When this happens, people are vulnerable to later involuntary retrievals when reminders occur in the environment. During these retrievals, several processes become relevant that could modify the memory and play a critical role in whether intrusions develop or are mitigated. Understanding these memory modification processes presents opportunities for intervention, even for memories that have been successfully consolidated (see Figure 9.1). Much of the research on memory modification has been conducted within the tradition of classical conditioning in experimental animals, although memory modification work also extends to declarative memory.

In research on memory modification, retrieval or recall of a previously consolidated memory has been proposed to engage two possible mechanisms: extinction and reconsolidation (Suzuki et al. 2004; Lee et al. 2006; Clem and Schiller 2016). In extinction, the repeated presentation of a conditioned stimulus leads to the progressive decrease in, and a resultant reduction in, the behavioral expression of a memory. In reconsolidation, the presentation of an isolated retrieval is thought to initiate a molecular cascade that can be bidirectionally modulated to either strengthen or weaken a memory trace (Dudai 2004; Suzuki et al. 2004; Lee 2009). For example, blocking *de novo* protein synthesis or other cellular and molecular processes critical for memory destabilization and reconsolidation prevents subsequent memory expression (Nader and Hardt 2009; Flavell et al. 2011; Elsey et al. 2018; Orederu and Schiller 2018).

Studies utilizing a noninvasive behavioral approach combined principles of reconsolidation and extinction to update a conditioned fear memory. Specifically, one day after rats acquired associative memories through fear conditioning, they were presented with an isolated retrieval trial; then, within the reconsolidation window, they received an extinction session. This paradigm led to an enduring decrease in conditioned responding, which unlike standard extinction (i.e., extinction not preceded by an isolated retrieval trial) did not result in the return of fear as assessed via renewal, reinstatement, and
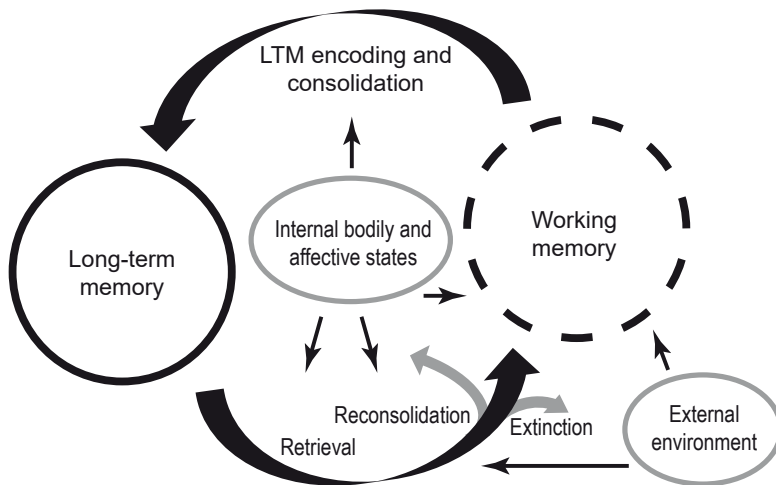
Figure 9.1   Interactive interplay in memory formation, maintenance, and modification. When initially acquired, memories are fragile and short lived. Some items or events may only be required for a brief moment and thus may be only temporarily maintained in working memory. Others may enter the process of consolidation and be stored in long-term memory (LTM). Upon retrieval, a consolidated LTM may reenter working memory and be temporarily put to use or trigger mechanisms of extinction or reconsolidation.

spontaneous recovery, and led to a retardation of fear reacquisition (Monfils et al. 2009). The findings provide evidence that retrieval followed by extinction promotes an updating of the original emotional memory rather than the formation of a competing memory trace (Cahill and Milton 2019). This idea was extended to fear conditioning in healthy humans (Schiller and Delgado 2010), individuals with phobias that received *in vivo* exposure (Telch et al. 2017), as well as a number of other forms of learning in rodents and humans (Xue et al. 2012; Sartor and Aston-Jones 2014; Luo et al. 2015; Björkstrand et al. 2016; Germeroth et al. 2017; Lee et al. 2017). In a clever twist, researchers have started to employ a procedure involving a distracting task (the computer game *Tetris*) following retrieval to update a traumatic memory (James et al. 2015; Iyadurai et al. 2018; Kessler et al. 2020; see also Holmes et al., this volume).

Memory retrievals serve as an opportunity to modify memories and indeed, other aforementioned protocols could actually be understood through the lens of reconsolidation updating (i.e., a change in the memory during a period of destabilization/restabilization). If intrusions are viewed as involuntary retrievals, they may provide a window of opportunity to modify an unwanted memory (see Figure 14.3 in Holmes et al., this volume), using either extinction mechanisms with extended exposures (Cassini et al. 2017; Hu et

al. 2018), reconsolidation updating (with extinction, cognitive restructuring, a distracting task or pharmacological disruption), or suppression mechanisms (described below). It may be important to consider that failure to act in the face of pathological intrusions could lead to their reinforcing a memory trace. Alternatively, the unique nature of intrusions may represent a form of memory that is resistant to modification. For instance, it may not destabilize upon retrieval or fail to promote extinction or reconsolidation processes. In this case, more targeted behavioral or pharmacological interventions (or a combination of the two) may help.

Memory Retrieval: Suppression

As discussed, intrusions can be conceptualized as involuntary retrievals, whether they are autobiographical memories, images of feared future events, or persistent thoughts. Efforts to control intrusions therefore involve controlling retrieval. Here, we discuss two processes that contribute to controlling such thoughts: retrieval suppression and retrieval-induced forgetting.

Retrieval suppression refers to the act of trying to stop an ongoing retrieval process, usually triggered by a reminder (Anderson and Green 2001). For instance, a loved one's face may elicit an intrusion of a recent argument that caused upset or a dog barking may trigger memories of the night a person got news of a friend's death. Other retrievals may be of a feared future: a medical bill may remind someone of an upcoming medical test. Because people do not enjoy the aversive emotional states that such reminders can create, they often exclude the offending content from awareness to regain their focus and composure. The aim is not merely to stop the intrusive thought at the moment, but to diminish its recurrence. Before describing retrieval suppression, we first discuss differing views on the value of suppression and whether it is desirable to encourage.

The Utility of Suppression. The clinical psychology literature contains conflicting observations on the value of thought suppression. Many clinicians, particularly those concerned with trauma or anxiety, maintain that suppression is an intrinsically unhealthy response to intrusive thoughts or memories. Indeed, it is common to view suppression as not only unhelpful, but as a significant risk factor for developing psychiatric disorders. Some studies have found that questionnaires putatively measuring thought suppression (Wegner and Zanakos 1994) are related to worse symptom severity in PTSD and other disorders characterized by intrusions. Given these facts, one can readily understand why clinicians would conclude that suppression is maladaptive. As a result, many therapists and therapies (e.g., acceptance and commitment therapy) discourage suppression and encourage people to accept and interact with their intrusive thoughts.

Several observations, however, are incompatible with the view that suppression is maladaptive. First, suppressing unwelcome thoughts is a widespread behavior that serves many different functional purposes, as outlined above. Simply concentrating on an important task, by its nature, requires that an individual excludes distracting content. Second, achieving emotional balance after an upsetting event involves "getting over" unwelcome thoughts and feelings, a process that requires active regulation of thoughts. People who have difficulty with upsetting occurrences (e.g., getting over arguments or perceived slights) are viewed as coping less well than people who quickly recover. Third, after a trauma, most people initially experience intrusive thoughts and memories that usually diminish over time. People who cannot reduce their intrusions are classified as having mental health concerns; those who are capable of reducing intrusive thoughts effectively are considered resilient, a clearly positive attribute. Fourth, people with anxious thoughts about a feared event that they cannot discard are considered as less healthy, compared to those who are able to set aside fears and cope well. Finally, the proposal that trying to stop unwanted thoughts is intrinsically ineffective is oddly discordant with the massive literature on inhibitory control and, more broadly, attentional control, in which controlling cognition and behavior is not only desirable but an essential capacity of intelligence. Indeed, the same clinicians that classify suppression as maladaptive often maintain beliefs that are contradictory to that stance upon closer inspection. For example, the belief that attention bias modification (i.e., training people to ignore unpleasant interpretations) is a desirable therapy amounts to a belief that it is desirable to become skilled at ignoring unpleasant contents, exactly what people do through suppression. The belief that people with intrusive thoughts suffer from cognitive control deficits that make them vulnerable to such thoughts assumes that resilient people can set aside upsetting thoughts, which must therefore be beneficial.

What do we make of these contradictions? What accounts for the view that suppression is maladaptive, given its ubiquity and utility? One reason derives from work on thought suppression using the white bear paradigm (Wegner et al. 1987). In this paradigm, participants are instructed not to think about a particular thought (e.g., a white bear) over a five-minute period. During that time, if they happen to think about the thought, they were asked to report this. Afterward, they are given a five-minute free expression period. Wegner found that participants asked to suppress thoughts of a white bear ironically experienced more white bear thoughts than participants who were not asked to suppress. This enhancement of the unwanted thought after suppression was termed the *rebound effect* (Wenzlaff and Wegner 2000). Across nonclinical samples, meta-analyses show that people exhibit a small to medium rebound effect compared to control instructions (Abramowitz et al. 2001). Wegner and Zanakos (1994) also introduced the White Bear Thought Suppression Inventory, a scale intended to identify people prone to thought suppression. This instrument became widely used in clinical research and was found to correlate with both

rebound effects and clinical symptomatology. Coupled with this question-naire, the white bear paradigm provided a compelling narrative that explained why psychopathologies were associated with intrusive thinking: unpleasant thoughts led to suppression, which led, in turn, to a paradoxical rebound effect, yielding a vicious cycle that worsened psychological symptoms (Wegner and Zanakos 1994).

There are significant difficulties with treating the white bear task as a model for controlling intrusive thoughts. First, the task requires the integration of the very thought to be suppressed with the task goal; that is, to know what task one is doing and to monitor whether one is achieving it, one reactivates the thought by asking: "Am I thinking about white bears?" The task is self-defeating be-cause it requires one to violate the task's goals to check whether the goal is being accomplished (Anderson and Huddleston 2012; Engen and Anderson 2018). Although the situation modeled by this task may model a slice of clini-cal reality, work using this paradigm has been overgeneralized to all instances of suppression (as will be seen shortly). Second, the White Bear Suppression Inventory—thought to quantify thought suppression frequency—has now been shown to have two factors opposed to one another: one measures thought sup-pression frequency while the other, the experience of intrusiveness (Blumberg 2000; Höping and de Jong-Meyer 2003; Rassin 2003). Research indicates that only the latter correlates with clinical symptoms, which is not surprising. Third, meta-analytic treatments demonstrate that suppression, as measured with the white bear task (and its rebound effects), shows few reliable differences across control and psychiatric populations, raising questions about the validity of the relationship of the process being measured and disordered control over intru-sions in everyday life (Magee et al. 2012). Finally, the clinical observations that led clinicians to be attracted to the claim that thought suppression is counterpro-ductive—reduced suppression in patients—can be explained more simply by positing that patients have difficulties applying inhibitory control rather than by suppression being intrinsically maladaptive. In line with the former view, sup-pression is an otherwise healthy coping response which, when disordered, poses a risk factor in developing intrusive symptoms—a very different interpretation with different clinical implications. For example, if suppression is not intrinsi-cally maladaptive, but simply not functioning properly in some populations, it is reasonable to expect that interventions can be developed to improve it. This possibility would be ruled out by the view that suppression is maladaptive.

Another reason why some clinicians have been opposed to suppression in-volves confusions in terminology. Researchers have not distinguished retrieval suppression from other constructs that may well be maladaptive and that might appear equivalent to retrieval suppression. For example, researchers have falsely equated retrieval suppression with the emotion regulation construct of expressive suppression and other seemingly related terms, such as avoid-ance, cognitive avoidance, and distraction. Careful analysis of the situations

captured by these terms reveals that retrieval suppression is cognitively distinct from these other concepts:

- White bear suppression (Wegner et al. 1987) refers to the explicit intention to not think about a specific unwanted thought for a specific period of time (e.g., 5 min). The unwanted content is incorporated as part of the goal and integrated with it, making it, by definition, not possible to avoid periodic attention to the thought. In contrast, retrieval suppression involves the attempt to stop or cancel the retrieval of content associated with a reminder while sustaining attention to the reminder. A person's goal does not specify the content to be inhibited but rather concerns the desire to shut down "whatever is associated with the reminder."

- Expressive suppression (Gross and John 2003) refers to the strategy of inhibiting behaviors associated with emotional states, such as adopting a poker face to hide emotions. Expressive suppression involves motor control, not control over thoughts or feelings. Retrieval suppression, on the other hand, involves suppressing mnemonic processes and content.

- (Cognitive) avoidance (Ehlers and Steil 1995; Williams and Moulds 2007) refers to entirely avoiding reminders that could otherwise trigger unwanted thoughts or emotions. It fundamentally involves not confronting reminders and not adapting one's internal response to them so that unwanted contents are preserved without alteration and may continue to intrude at the slightest provocation. It is widely considered a maladaptive coping process, a "not dealing with the problem." In contrast, retrieval suppression requires a person to confront reminders directly, attend to them, and adjust their internal retrieval and affective, conditioned responses to them, features shared with cognitive behavioral therapy and exposure therapy in particular.

- Distraction refers to the removal of attention from emotions or thoughts and refocusing it onto other innocuous stimuli in the world or to unrelated thoughts. It is worth distinguishing general distraction (where attention shifts to stimuli, topics, or activities unrelated to the intruding thought) from specific distraction, which involves interacting with reminders. If general distraction leads one to remove attention from reminders to intrusive thoughts or emotions, it functions in a manner similar to avoidance and fails to engage suppressive processes that might adjust one's internal mnemonic and emotional responses. If distraction takes the form of generating alternative associations, thoughts, or memories in response to a reminder, this form of specific distraction would constitute thought substitution, which retrains the response to the cue to elicit different content, potentially leading to altered patterns of thought in the future. Sometimes, because retrieval suppression involves attending to reminders and not the memory, or instead, attention

to thought substitutes, researchers falsely equate retrieval suppression with general distraction, from which it clearly differs.

Because clinical research has often not attended to these distinctions, it is easy to see how, with the paradoxical rebound effects reported in the white bear paradigm, clinicians would have taken a dim view of suppression. In recent years, retrieval suppression has been studied intensively. As we look at next, this work integrates models of the control over intrusive thinking within the bedrock of cognitive neuroscience research on inhibitory control.

Retrieval Suppression: Methods of Study and Basic Findings. Work on retrieval stopping focuses on a situation that is similar to motor response inhibition. Participants are confronted with a cue that reminds them of a memory and are asked to focus on that cue for trials of several seconds, while preventing the memory from entering awareness. This situation parallels motor response inhibition tasks, such as the Go/No-Go and Stop-Signal tasks, in which participants are presented with stimuli designed to elicit a motor response, yet are required to stop the action. To study retrieval suppression, Anderson and Green (2001) introduced the Think/No-Think paradigm. Participants are trained on cue-target pairs until they can recall the target items when given the cues. These pairs can be unrelated words (e.g., ordeal roach), picture pairs, or even cues related to autobiographical memories. Participants then enter the Think/No-Think task. On each trial, participants receive a reminder from one of the pairs, colored either in green or red. Green reminders signal the participant to retrieve the associated item and keep it in awareness for the trial's duration; red reminders signal that they should attend to the stimulus but prevent the associated item from entering awareness at all. Usually, a given item is only suppressed or retrieved (not both), and will be repeated many times (usually, 8–16 times). A final memory test follows, giving participants all studied cues to determine the impact of retrieval suppression on the suppressed content. Performance on the final memory test is computed not only for the Think items, which were repeatedly retrieved, and No-Think items, which were suppressed, but also for baseline pairs studied in the training phase but which did not appear in the Think/No-Think phase.

A key finding that emerged from this work is that if people consistently suppress the retrieval process, memory for the suppressed No-Think items declines compared to Think items as well as to baseline items receiving the same initial training as No-Think items. Because suppressing retrieval impairs retention of the suppressed content, this phenomenon is known as suppression-induced forgetting (for a review, see Anderson and Hanslmayr 2014). Suppression-induced forgetting increases with the number of times that people suppress the unwanted content and resists monetary incentives on the final test for successful recall (Anderson and Green 2001; Hulbert and Anderson 2018). Forgetting generalizes to novel test cues for the suppressed items, exhibiting a property

known as cue independence: suppression-induced forgetting occurs not only for simple word pairs, but also for objects or complex scenes, irrespective of whether they are neutral or negatively valenced (Depue et al. 2007; Anderson and Hanslmayr 2014). Suppression-induced forgetting has even been observed with autobiographical memories, although the effect emerges more in memory for details (Noreen and MacLeod 2013). Importantly, a growing body of work shows that suppression-induced forgetting does not merely influence explicit memory, it also affects implicit memory regardless of whether perceptually driven (Kim and Yi 2013; Gagnepain et al. 2014) or conceptually driven tasks are used (Hertel et al. 2012, 2018; Taubenfeld et al. 2019). Of clinical relevance, retrieval suppression also reduces affective responses to the suppressed content when participants view it later, whether measured with ratings (Gagnepain et al. 2017) or psychophysiology (Legrand et al. 2019; Harrington et al. 2020). These findings suggest that retrieval suppression, for healthy participants at least, regulates the negative affect associated with intrusive memories (Gagnepain et al. 2017; Engen and Anderson 2018).

Although most research on retrieval suppression focuses on final test performance, recent studies have increasingly focused on online measurements of intrusive memories during the Think/No-Think phase (Levy and Anderson 2012; Benoit et al. 2015; Hellerstedt et al. 2016; van Schie and Anderson 2017; Legrand et al. 2019; Harrington et al. 2020). In these studies, the Think/No-Think task is unchanged except that after every trial, participants provide a phenomenological report of their experience. Specifically, they judge whether the associated memory entered awareness. On Think trials, awareness of the memory is the intended outcome, and people usually report awareness on nearly all trials (averages of 90–100% of trials). On No-Think trials, however, because participants are instructed to prevent awareness of the memory, any awareness that does occur constitutes *clear evidence for an intrusive thought*. Importantly, when intrusions occur in this task context (e.g., on 30% of No-Think trials), the inference of involuntariness is particularly clear. Indeed, one of the hallmark features of an involuntary, automatic behavior is its tendency to occur, despite efforts to stop it. Intrusions during No-Think trials therefore provide a robust and theoretically clean measure of a vulnerability to involuntary retrieval. Other paradigms hoping to measure intrusions (e.g., diary studies) cannot make a clear attribution of involuntariness (for discussion, see van Schie and Anderson 2017).

Contrary to what might be inferred from the white bear paradigm, people can suppress the retrieval process and reduce intrusive thoughts elicited by reminders. For example, across repeated suppressions one observes a dramatic downregulation of intrusions between the initial trials (e.g., 60% intrusions) and final trials (e.g., 25% by the tenth trial). Moreover, the slope of the downregulation in intrusions is often correlated to suppression-induced forgetting (Levy and Anderson 2012; Hellerstedt et al. 2016), and suppression effects persist to produce memory deficits on both direct and indirect tests, contrary to the

notion of rebound. Work on retrieval suppression aligns the control of intrusive thoughts with work on inhibitory control in general, especially work on action stopping. Indeed, given that nobody questions that actions can be stopped and that a giant literature has developed which documents action-stopping mechanisms in humans and other animals, it would be more surprising to find that internal actions, such as retrieval, were not also subject to control (Anderson and Green 2001). This conclusion fits the observation that psychiatric populations with intrusive thinking also have deficits in cognitive control (Catarino et al. 2015; Waldhauser et al. 2018). Suppression-induced forgetting has also been related to self-reports of rumination (Fawcett et al. 2015), perceived success at thought control in real-life settings (Küpper et al. 2014), and self-reports of intrusive memories collected in the week following exposure to a trauma film (Streb et al. 2016). These findings support the view that suppression is a helpful coping process that is compromised in clinical populations, rather than an intrinsically maladaptive approach to intrusive thoughts.

Retrieval-Induced Forgetting: Methods and Basic Findings.   Thus far we have focused on the role of inhibitory control in intentional retrieval stopping. Inhibitory control, however, can also be engaged in another context to forget distracting memories: selective memory retrieval. During selective memory retrieval, a person seeks to retrieve a particular event, idea, or fact. Reminders, however, are often associated to many traces, creating interference known as retrieval competition. Retrieval interference is ubiquitous, whether one is retrieving an episodic memory, a fact, a general idea, or simply an object's location (Anderson and Neely 1996). Isolating the desired trace despite this interference engages prefrontally mediated inhibitory control mechanisms to suppress competing memories. This suppression induces aftereffects on the inhibited traces causing forgetting. Thus, ironically, the very act of remembering something can cause forgetting of competing memories; this is known as *retrieval-induced forgetting* (Anderson et al. 1994; Anderson and Spellman 1995; Levy and Anderson 2002; Anderson 2003; Bäuml et al. 2010; Storm and Levy 2012). Although retrieval-induced forgetting is not intentional, this process can be exploited in a motivated manner to deliberately forget unwelcome memories (Benoit and Anderson 2012). The procedure by which this phenomenon is studied will be described below.

Retrieval-induced forgetting becomes particularly relevant to controlling intrusive thoughts when it is deployed in a motivated way; that is, as a key strategy for deflecting unwelcome thoughts, refocusing on benign knowledge. For example, after an argument with a friend or partner, one might (upon seeing them again) try to recall more pleasant thoughts about them (Anderson 2001). Optimists routinely seek the positive in any event, even if unpleasant things happened, and in so doing, may reshape their memories to fit a more pleasant existence. Indeed, the widely demonstrated positivity bias in autobiographical memory retrieval (the disproportionate accessibility of positive

over negative autobiographical memories) significantly correlates with the amount of retrieval-induced forgetting that people show on laboratory tests with neutral materials (Storm and Jobe 2012). Thus, the capacity to exhibit inhibitory control over memory predicts a sunnier view of one's past than might be expected. Critically, retrieval-induced forgetting may be a mechanism that explains why reappraisal is effective as an emotion regulation strategy, given that reappraising an unpleasant memory involves selectively retrieving the contents of an event or thought (Engen and Anderson 2018). If conventional therapies for addressing intrusive thinking rely on reframing, rescripting, or reappraisal, retrieval-induced forgetting may be central to reducing the accessibility of negative interpretations and intrusions.

Selective retrieval's role in motivated forgetting has been studied with the retrieval-practice paradigm as well as with a variant of the Think/No-Think paradigm, known as thought substitution (Hertel and Calcaterra 2005; Bergstrom et al. 2009; Joormann et al. 2009; Benoit and Anderson 2012). In thought substitution, participants are instructed to stop retrieval of the unwanted memory by retrieving diversionary content, either self-generated or supplied, related to the cue (Hertel and Calcaterra 2005; Bergstrom et al. 2009; Benoit and Anderson 2012). Retrieval of such thought substitutes significantly impairs retention of the intrusive thoughts, although the neural mechanism underlying this process is dissociable from that involved in retrieval suppression (Benoit and Anderson 2012). Thought substitution as a strategy for controlling intrusive memories has been studied extensively using the Think/No-Think task, especially in dysphoria and depression (Joormann et al. 2009; Stramaccia et al., unpublished). Given people's natural pre-disposition to replace intrusive contents with distracting thoughts, how thought substitution affects memory is important to study.

Summary and Commentary

Most research on intrusive thoughts has focused on long-term memory formation, retrieval, and their control. The substrate for intrusive thoughts is laid down when memories are encoded. Many intrusive thoughts are related to past experiences, such as those associated with substance use or specific events (e.g., feeling rejected, being traumatized). Other intrusive thoughts may not refer to specific episodic memories (e.g., images of feared future events or ruminations) but are nonetheless stored in memory and governed by its mechanisms. If a memory is encoded in a particular way, it may be more accessible to reactivation at a later time. Highly salient memories may, in particular, be easily triggered by external and internal cues, many of which do not reach conscious awareness. Intrusions, therefore, appear to pop spontaneously into a person's mind. Such potent memories may be related to specific types of emotional information (e.g., negative information as in depression and anxiety) or

to a particular topic (e.g., danger and threat in PTSD and anxiety; substance-related information in substance use disorders).

Substantial research has focused on cognitive control processes, including the role of inhibitory control mechanisms in intrusive thinking. Memory control mechanisms related to action control are known to suppress retrieval of specific, episodic long-term memories and can work more globally to block all episodic memory retrieval in response to a cue. Alternatively, reminder cues can be used to selectively retrieve alternative distracting content, suppressing the intruding thought via the mechanisms of retrieval-induced forgetting. Moreover, retrieving information may trigger destabilization and/or restabilization of a thought. As such, retrieval provides an opportunity not only to increase the strength of a memory, but also to reduce it. In addition, working memory processes are critically linked to the encoding and retrieval of long-term memory and are important to guiding the selection of information for consolidation into, and retrieval from, long-term memory (see Figure 9.1). It seems likely that for an intrusive thought to be experienced, it would enter working memory, making this system central to having an experienced (conscious) mental event.

Summary of Commonly Used Methods and Desiderata for Investigating Intrusive Thinking

Methods and Paradigms

The previous section described examples of research into the mechanisms underlying intrusive thinking and their control using a variety of methods. Here we present an overview of the methods and paradigms that are most commonly used to study intrusive thinking in humans, nonhuman animals, or both, each with its own strengths and weaknesses.

- *Experience/thought sampling* is a form of self-report that is often used to assess mind wandering in the present moment (Larson and Csikszentmihalyi 1983; Bolger et al. 2003; McVay et al. 2009; Schooler et al. 2011; Baird et al. 2014; Fraley and Hudson 2014). One form has people pressing buttons when they catch themselves mind wandering or experiencing other intrusions. This can be done in the lab or in real life (e.g., using a diary or smartphone). Alternatively, one can also probe people at irregular intervals to report whether they were on task or not. A comparison of probed versus unprobed self-report shows that having an experience and knowing that you are having an experience are different things (Schooler et al. 2011; Baird et al. 2014). The first requires metacognition and this is imperfect, and possibly worse in clinical populations.

- In the *emotional Stroop task*, individuals view items with two overlapping dimensions, such as a word superimposed on a face (Williams et al. 1996). The participant bases their response on a task-relevant dimension (e.g., valence of a word) while ignoring the task-irrelevant dimension (e.g., valence of a face). Proneness to intrusions is defined as the degree to which reaction times or error rates are increased on incongruent trials (e.g., the word "suicide" on a happy face) as compared to congruent trials (e.g., the word "joy" on a happy face).

- The *trauma film paradigm* is used to model intrusive memories of distressing events (Horowitz 1969; James et al. 2016a). It uses film stimuli in the laboratory, which contain traumatic content that can bring about intrusive memories subsequently in daily life. These memories are typically recorded in a diary, allowing for a frequency count of intrusive memories, or via button presses during a provocation task (both are forms of experience sampling).

- *Pavlovian conditioning* is the classic model by which simple associative learning and memory are studied, and it is well suited for research across species (Pavlov 1927; Rescorla and Holland 1982; LeDoux 2003; Nader and Hardt 2009). In this paradigm, an initially neutral stimulus (conditioned stimulus, CS+), such as a triangle, is repeatedly paired with an intrinsically aversive stimulus (unconditioned stimulus, UCS), such as an electric shock, while another conditioned stimulus (CS–), such as a circle, is never paired with the UCS. With sufficient CS+/UCS pairings, the CS+ acquires the same aversive qualities as the UCS and will elicit a conditioned defensive response that can be measured, for example, by the amount of freezing or avoidance behavior in nonhuman animals, or skin conductance, acoustic startle response, heart rate, pupil dilation, action tendencies, UCS expectancies, and subjective distress in humans. After repeated presentations of the CS+ without the UCS, the defensive response usually diminishes, a process referred to as *extinction learning*. Intrusions may often be conceptualized as conditioned responses to external or internal reminders.

- *Stop-Signal and Go/No-Go* are tasks that require inhibition of already initiated responses. Error rates in response to Stop-Signal or No-Go trials may reflect the outcome of diminished control of actions in a similar way as intrusions reflect a diminished control of thoughts.

- In the *Think/No-Think paradigm*, participants learn a set of cue-target pairs, either word pairs, picture pairs, or autobiographical memory cues (Anderson and Green 2001). Next, trials appear, each presenting a reminder cue and a task cue: a green box around the reminder cue signals that the associated memory item needs to be brought to mind whereas a red box signals that retrieval of the associated memory item needs to

be suppressed. This method allows one to measure, during the suppression task itself, whether participants experienced intrusions (i.e., intrusion reports during the No-Think trials) in response to the reminders, despite their best efforts to stop retrieval.

- Retrieval-induced forgetting is studied using the *retrieval-practice paradigm* in which participants learn to associate a cue with multiple associates; participants are then set the task of retrieving a subset of these items, but not the rest (Anderson et al. 2000; Bekinschtein et al. 2018). Rather than providing explicit instructions to forget, this paradigm capitalizes on a natural process to select relevant information. It leads to forgetting of nonselected, competing memory items through a process of inhibition. Retrieval-induced forgetting is related to a variant of the Think/No-Think paradigm that uses thought substitution rather than retrieval suppression, and when deployed in a motivated way, may be a model for how individuals control and diminish intrusive thoughts. Retrieval-induced forgetting is now also being studied in rodents, capitalizing on rats' intrinsic curiosity (spontaneous object recognition paradigm). After encoding multiple objects in an environment, selective retrieval of one object leads to forgetting of other objects in a manner directly analogous to that seen in humans and that depends on the prefrontal cortex. An advantage that this paradigm enjoys over retrieval suppression is that it allows us to understand memory inhibition mechanisms, not only at the level of neural systems (through imaging in humans) but also through foundational neurobiological work in nonhumans using electrophysiological, lesion, optogenetic, and molecular biological methods.

- With *Pavlovian-instrumental transfer*, a subject undergoes Pavlovian threat conditioning and is also trained on an instrumental reinforcement task (reward or avoidance) in a different context (Campese et al. 2013; Cartoni et al. 2016; Watson et al. 2018). In the test phase, during extinction (or baseline performance) of the instrumental task, the CS+ is presented and its possible rate-altering effects measured as evidence of negative Pavlovian-instrumental transfer. There are several variations of this basic procedure, which can also be appetitive in nature and have general and specific aspects of transfer in relation to the overall goal. The degree to which the presentation of the CS+ interrupts an animal's (human or nonhuman) goal-directed behavior is taken as a proxy for the experience of an intrusion.

These paradigms should continue to be developed. In addition, we wish to stress that promising new approaches are emerging from computational neuroscience, as we illustrate next.

Computational Approaches to Intrusive Thinking

We turn now to a few considerations from a computational perspective. A common feature of intrusive thoughts is that they occur at the "wrong time." They are remembered or imagined contents that are not appropriate for the situation or the particular problem our minds are focused on at that time. But what constitutes the "wrong" time? How do we, our minds, or our brains choose the "right" thoughts or the "right" time for particular thoughts? Are there any general principles that guide the selection of thoughts, and what might they be? Computational modeling of decision making provides a framework for this problem and suggests a number of potential answers that might apply to different mental illnesses (Huys et al. 2016; Redish and Gordon 2016). At their core, these accounts rely on the notion that thoughts are expensive and that organisms have much to gain from deploying them well.

Our starting point is the acknowledgment that most decision problems faced by humans on a daily basis are so complex that they radically outstrip our computational abilities. The mismatch between computational demands and resources requires us to invest our computational resources with wisdom. Unfortunately, exerting any such wisdom further complicates the problem: the optimal deployment of limited computational resources is itself a decision problem that has the same form as the original decision problem, but which is vastly more demanding (Russell and Wefald 1991). This problem is displayed in Figure 9.2a. Hence, we are faced with the double challenge consisting of hard problems and the even harder problem of apportioning our cognitive resources.

There is, however, a silver lining for the study of thoughts. As the problem of optimally apportioning computational resources is essentially of the same nature as the original decision-making problem, the same formalisms used to study choice can potentially be brought to bear on the problem of thought selection (Lieder et al. 2018a, b). Errors in these processes might, in turn, link the emergence of intrusive thoughts to well-defined optimal approaches to solving decision-making problems under resource constraints (Huys and Renz 2017). Below, we review several computational accounts of thought choice and discuss how intrusive thoughts could arise.

Affective Biases

A bias toward thinking about negative events can arise through utility-weighted sampling, a process whereby the samples (in this case, what we choose to think about) are not just proportional to the probability of the event but also to its importance. Humans routinely overestimate the probability of extreme negative events (Lichtenstein et al. 1978), and such biases have long been viewed as a signature of irrationality. However, they may alternatively reflect a rational use of limited resources. When many different outcomes are
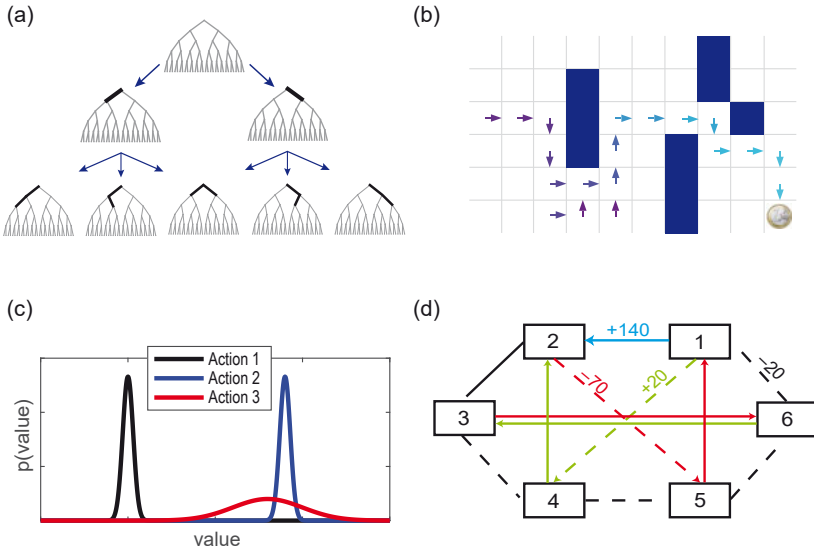
(a)

(b)

(c)

(d)

Figure 9.2 Computational approaches to thought guidance: (a) The meta-reasoning problem of optimally apportioning limited resources can be formalized as a decision-making problem over decision-making problems. While the standard decision-making problem is the tree at the top without thick lines, the meta-reasoning problem is a tree of such trees, where each branch corresponds to choosing to evaluate one branch (thick lines) in the original decision problem. Intrusions could relate to a tendency to only sample one set of options (arrows). (b) Prioritized replay: In this state-space, if the reward (Euro coin) is received in the bottom right corner, then standard model-free learning updates only the state immediately adjacent to it. Prioritized replay allows memories to be reused multiple times to update more distant states; here, propagating the information about the reward all the way back to the starting state. If the need or gain functions determining what memories to replay are altered, then this could result in repetitive replay of the same memory (arrows) as well as a failure to update distant states (i.e., to integrate the memory with other memories). Adapted after Mattar and Daw (2018). (c) Guidance of thoughts by habits: Consider the situation where a valuation system provides distributions for likely values of actions. Here, action 1 is clearly inferior to the other two, and it appears that action 2 is the best. If the agent were to invest computational effort into refining the estimates of the values of the actions, it would be best to examine action 3, as it may be even better than action 2. In this manner, a goal-directed or model-based system could elaborate on approximations provided by a simpler valuation system. This would also mean that the goal-directed thought choice could be (mis)guided by habits. (d) Stopping aversive thoughts: In this task, participants are extensively trained to learn to navigate a maze where each transition incurs some gains or losses. When given the opportunity to plan freely, they stop internal simulations when they encounter one of the salient losses in red. Adapted after Huys et al. (2015).

possible, it becomes difficult to consider them all, and humans are thought to simulate instead a few outcomes in their mind and average the results of these few simulations. If these simulations are too few in number, they are likely to miss very important outcomes. Simulating in proportion not only to the

probability, but also in proportion to the utility, allows the resulting estimates to be more robust. That is, the apparent overestimation of extreme events may help mitigate cognitive limitations and explains a number of apparent irrationalities (Lieder et al. 2018b). Applied to the setting of intrusive thoughts, it provides one argument for why salient negative events should be simulated even when they are very unlikely. In fact, the more aversive the event, the higher the likelihood of simulation, suggesting one path by which the perceived negative valence of a simulated event might increase the frequency with which it is thought about.

Trauma Replay

A computational process that may be related to intrusive thoughts in PTSD is that of prioritized replay. Briefly, learning from experience is often slow. A prominent way of learning from experience through iterative updates of expectations with prediction errors is particularly slow. Indeed, this is one reason why it has been considered to be a computational account of habitual learning (Daw et al. 2005). Figure 9.2b shows that this is because experience at any one state or stimulus, s only leads to learning at that particular state or stimulus and is only propagated to adjacent states upon the next transition from those adjacent states back into state s. Such experience influences "knowledge" only very locally and does not generalize. One solution to this is to store episode-like chunks of autobiographical memory and replay them multiple times so as to spread the effect of any one experience to other states. This can substantially accelerate learning (Schaul et al. 2016). However, replaying memories is also costly, and hence the key here is to again deploy the resources (in this case, which memory to replay) efficiently. Computational models of this process identify two terms that determine which memory to replay: a need term and a gain term (Mattar and Daw 2018). The need term captures how likely state s is to be encountered in the future (Russek et al. 2017). Clearly, using computational resources to learn about states that will never be visited is not useful. The gain term captures how much the memory is likely to change behavior.

As this gives a normative account of when to optimally replay particular memories, it should reflect the tendency to replay both intrusive and non-intrusive memories experienced after laboratory induction procedures. It also provides an interesting window on features of intrusive memories in PTSD. The gain term depends on the implied change in behavior and, as such, a memory should only be replayed if it implies a change in behavior. This suggests an important modulatory role of generalization processes seen across anxiety disorders (Laufer et al. 2016). For instance, replay tendency after traumatic abuse by a trusted person should be influenced by the perceived importance of this event for other relationships. The more relevant it is judged for other relationships, the higher the replay tendency should be. More generally, it captures

the notion that excessive negative appraisals of the sequelae of a trauma might relate to the emergence of PTSD symptoms (Ehlers and Clark 2000).

Obsessions

In the context of OCD, intrusive thoughts exhibit a different quality to those experienced in PTSD (see Monfils and Buss, this volume). Extensive evidence suggests that OCD involves a shift from goal-directed toward habitual behavior, with most evidence pointing to an impairment in goal-directed or model-free control rather than an explicit change to habitual or model-based processes (Robbins et al. 2012; Gillan et al. 2013, 2015; Voon et al. 2015). How could an impairment in model-based control give rise to the emergence of intrusive repetitive thoughts as seen in OCD? One avenue arises from the notion of value of information (Keramati et al. 2011). Consider the situation in Figure 9.2c, where one option is clearly good and a second one clearly bad. The third option appears slightly worse than the best one, but could be better. The optimal investment of limited cognitive resources in this case would be to examine this one option, again because this investment of cognitive resources has the potential of altering behavior. This could underlie intrusive thoughts in OCD if there was a drive provided by uncertain habitual or model-free evaluations, such as through distributional reinforcement learning (Dabney et al. 2017), coupled with impairments in goal-directed evaluations which fail to result in improved predictions.

Rumination

Frequently seen in depressive disorders, ruminations are usually described as a tendency to think repetitively about the causes of distress without engaging in active problem solving (Nolen-Hoeksema et al. 1993, 2008; Treynor et al. 2003). Viewed in the context of intrusive thinking, they appear to involve a prominent failure to inhibit or discontinue aversive sequences of thoughts. This raises a theoretical point not yet addressed above—that of thought inhibition. Thus far we have focused on which thoughts, evaluations, or memories should be chosen and have not yet addressed the monitoring of a thought that appears not to be fruitful nor the question of when it should be terminated.

This question has been examined in some detail using computational accounts of the task in Figure 9.2d. Here, individuals are trained to navigate a maze where each transition yields rewards or losses. They are then dropped into one state randomly and asked to search a route of a given length through the maze that maximizes their total earnings. The problem they face is a binary decision tree, as in Figure 9.2a. Individuals are much more likely to identify optimal routes that do not transition through salient losses than those which do, independently of the size of the large loss (Huys

et al. 2012, 2015). Computational models of the choices in this task suggest that participants internally simulate potential routes through the maze and terminate simulation when they encounter a salient loss. Functional magnetic resonance imaging suggests that this inhibition recruits the subgenual anterior cingulate cortex (Lally et al. 2017), a region known to be important in depression (Drevets et al. 1997), its treatment (Mayberg 2009), and rumination (Hamilton et al. 2015). Indeed, individuals who score high on self-reported rumination show reduced pruning; that is, a reduced tendency to terminate thoughts when encountering large losses during their internal simulations (Q. Huys, pers. comm.), suggesting that rumination might directly relate to an inability to inhibit aversive thoughts, possibly via impairments involving the subgenual anterior cingulate.

Desiderata for More Sophisticated Behavioral Paradigms to Measure Intrusive Thinking and Its Control

Studying intrusive thinking is challenging for several reasons. Intrusive thoughts are typically spontaneous, making it hard to predict when they occur, and therefore, when to measure them. Although paradigms exist that can reliably induce intrusive thinking under controlled circumstances in the laboratory (as detailed above), additionally allowing for the investigation of its neural correlates, these paradigms may not capture all circumstances under which intrusions occur in real environments. In addition, while it is possible to infer the occurrence of intrusion-like events from nonverbal behavior, measuring an "experienced mental event" ultimately relies on self-report. Indeed, historically intrusive thinking has been studied with self-report questionnaires (for an overview, see Banich, this volume).

A general issue is how well the more objective, laboratory-based measures of intrusive thinking correlate with subjective measures. A common finding for other constructs (e.g., impulsivity) is that they do not intercorrelate particularly well (Nombela et al. 2014). Why might this be the case? One notion is that the subjective measures are in some sense "noisier" and more prone to error. It is well known, for instance, that people sometimes have great difficulty in expressing their conscious evaluations or descriptions of their thinking, and there may be considerable interindividual ability in this capacity. However, this notion can perhaps be dismissed, as there is also considerable evidence of superior test-retest reliability for questionnaires than objective measures, such as those based on reinforcement learning parameters (see Table 5 in Bland et al. 2016). It does appear likely that objective tests may capture much narrower aspects of the construct under study, thus resulting in a looser association with the more generic aspects captured by a composite measure of a psychometrically well-designed questionnaire. Another important conclusion, however, is that the objective measures and the subjective responses obtained from questionnaires are simply

tapping into quite different processes, the latter most obviously addressing the contents of monitoring operations in meta-consciousness. This reliance on meta-awareness can be problematic and may be especially compromised in clinical populations. More importantly, in the field of intrusive thinking, there is a clear need for the development of better self-report instruments: at present, there does not appear to be a measure that captures all aspects of intrusive thinking in one instrument. In particular, there are four gaps in this field (for discussion, see Banich, this volume).

1. Most questionnaires were designed with a particular clinical syndrome in mind (e.g., PTSD, OCD, craving). As such, a general intrusive thinking scale does not exist.
2. For the most part, the questionnaires do not assess both the content of the thought as well as the ability to control those thoughts, broken down by content.
3. There are few, if any, questionnaires that specify the quality of intrusions across numerous dimensions, such as their form (verbal, images, urges), vividness, salience, and frequency.
4. There are no questionnaires that systematically assess how intrusions are triggered or the context in which people experience intrusions.

Looking forward, future research should focus on designing a new, theoretically motivated questionnaire, refining existing objective measures and capitalizing on the potential of computational approaches.

What Are the Neural Systems Relevant for Intrusions and Their Control?

The exact neural underpinnings of intrusions are currently unknown, but we could speculate about processes involved in intrusive thinking by combining what we know from the study of cognitive control, including action control, and the study of emotional memory processes.

Cognitive Control

Action Control

A critical prefrontal region for stopping movements is the right inferior frontal cortex (rIFC) as shown by lesion and transcranial magnetic stimulation studies (Aron et al. 2014). It is thought that the rIFC, in concert with the presupplementary motor area, implements stopping through a (fast) hyperdirect pathway to the subthalamic nucleus (STN) of the basal ganglia, and this suppresses thalamic drive back to cortex, leading to movement cancellation (reviewed by Bari and Robbins 2013; Jahanshahi et al. 2015; Wessel and Aron 2017). Importantly,

in the standard case, subjects apparently stop using a global mechanism. For example, stopping the voice suppresses the hand (Badry et al. 2009; Cai et al. 2012). This broad skeletomotor suppression begins around 120 ms after the Stop signal and lasts for 100 ms or more. It is thought that the broad suppression relates to a wider putative impact of the STN on basal ganglia output (Aron 2011), just as the degree of global motor suppression in the Stop-Signal paradigm relates to the level of oscillatory power in the STN (Wessel et al. 2016).

Yet subjects can also stop selectively when they are *forced* to do so (Aron 2011). For example, they are able to stop one response (or hand) while continuing with another with minimal interference, and this does not result in broad motor suppression (Majid et al. 2012). It is possible that this form of selective response suppression relates not to a hyperdirect cortico-STN connection but rather to a frontal-striatal-pallidal, so-called, indirect pathway. For example, people with degeneration of striatum and pallidum, who are thought to have an indirect pathway disorder, could not stop as selectively in the paradigm and did not show typical physiological signatures of selective stopping (Majid et al. 2013).

As will be discussed below, this global versus selective picture in motor stopping may have relevance for global versus selective control over memory. Future work could develop behavioral paradigms to look at this selective control on the analogy of selective response suppression, taking into account the observation that selective stopping is best done when the subject proactively sets it up ahead of time (Cai et al. 2011; Majid et al. 2012); that is, it could be possible to go into a situation preparing to suppress a particular memory intrusion.

Working Memory

The three working memory operations that we discussed above may be particularly relevant to intrusions:

1. Gain access or gate information into working memory.
2. Select information within working memory to be given priority.
3. Remove information from working memory.

In general, the neural correlates of working memory mechanisms have been well described (D'Esposito and Postle 2015) and likely involve prefrontal areas that are involved in executive control processes (e.g., selecting among information in working memory), basal ganglia mechanisms that work to gate information into working memory, and posterior brain regions that help to provide sensory or abstract (e.g., semantic) representations of information to be activated and/or placed in working memory. Most of the paradigms utilized in cognitive neuroscience require the confluence of processes whereby information enters, is selected, and updated in working memory (e.g., the N-back task). Nonetheless, the neural bases of some of these three main processes have been distinguished.

With regard to gaining access to working memory, research suggests that the basal ganglia may play an important role in determining when the gate to working memory should be opened so as to let in new information (see Badre, this volume). It is proposed that dopaminergic reinforcement learning mechanisms help to provide information on when the gate to working memory should be opened. Hence, this process is not conscious and controlled. The basal ganglia form multiple loops with distinct regions of the cortex, which are semi-segregated based on the nature of information they carry (e.g., sensory, motoric). As such, these basal ganglia mechanisms might selectively or concurrently act to allow access to working memory. In the case of intrusive thoughts, especially when they are repetitive in nature, it may be that through habitual learning, the intrusive content can more easily open the gate and gain access to working memory. While alterations in the accessibility of information to working memory may affect many different syndromes in which intrusive thoughts are observed, the specific basal ganglia loop involved may differ according to the disorders (e.g., more motoric in substance use, more visual in anxiety), as discussed by Balleine (this volume).

With regard to selecting information within working memory to be prioritized (or buffered from interference), processes often considered to be executive aspects of working memory, meta-analyses suggest that selection relies on frontal and parietal regions across a variety of tasks (Nee et al. 2013). For example, using Stroop and emotional Stroop tasks, research suggests that these tasks tend to involve regions associated with the frontoparietal, cingulo-opercular as well as dorsal and ventral attention networks (e.g., a meta-analysis by Chen et al. 2018b), regions associated with cognitive control. Activation in these areas has been found to be altered by characteristics related to intrusive thought. For example, an individual's degree of worry influences the degree of activation (Engels et al. 2007) as well as the time-course of activation (Levin Stilton et al. 2011) across lateral and medial prefrontal regions. Portions of the superior parietal lobe may be particularly important in selecting or shifting what information is currently within the attention focus of working memory (Tamber-Rosenau et al. 2011). These regions are engaged in executive processes more generally, not just specifically with regard to working memory.

With regard to removing information out of working memory, the picture is less clear. As discussed by Banich (this volume), most methods examine removal from the perspective of replacement; that is, how new information is placed into working memory and/or how the current information is buffered from removal. These operations involve cognitive control regions, including both the frontoparietal network and the cingulo-opercular network. Furthermore, variance in activation in these areas has been observed in individuals, like those who are depressed, who behaviorally show difficulty in removing certain types of information (e.g., negative emotional information), as indexed by subsequent interference from those items (Foland-Ross et al. 2013).

Less research has examined how one can take information currently held within working memory and dispose of it completely. This has been a difficult question to address because it is hard to verify that an item has indeed been removed. Other than with behavioral methods such as self-report, which may not be reliable, few (if any) methods are able to confirm that a thought has been removed or inhibited. Yet understanding such processes could have important implications for how interventions for intrusive thoughts might be created.

While there is little research in this area, some work has demonstrated that brain imaging techniques can be used to help verify the removal of information and simultaneously to examine the control mechanisms by which such removal occurs. In one of the few studies of this nature, Banich et al. (2015) utilized activity in the sensory cortex as a rough proxy for whether information was currently active in working memory. With this approach, they verified the presence of such activity when individuals were maintaining information about a picture just viewed or a short tune just heard, or when they replaced the item with something else. They also observed a lack of such activity when the item was removed from current thought, both when participants specifically suppressed that item as well as when they did so by clearing their mind of all thought. With regard to control mechanisms, a hierarchy of function was observed. Common across the replace, suppress, and clear conditions—all of which require a shift in attention away from the original item to something else, compared to the maintain condition—was activation in superior parietal regions implicated in shifting attention among items in working memory (Koenigs et al. 2009). When information had to be removed from working memory (the suppress and clear conditions) as compared to when an item was present in working memory (the maintain and replace conditions), activation was observed over regions of lateral prefrontal cortex involved in executive processes that act on working memory. Finally, for the clear condition as compared to all the other conditions, there was increased activation in the insula, suggesting a shift of attention to bodily states (Craig 2011) as well as activation in inferior parietal cortex, which may represent a mechanism for altering the bottom-up salience of information (Cabeza et al. 2008), either by reducing the salience of visual information or by increasing salience of information derived from bodily states (see also Fedota and Stein, this volume). In relation to long-term memory, the global "clear" condition on the contents of working memory may be analogous to a global stopping process that inhibits retrieval of all information from the hippocampus, whereas the "suppress" condition may be analogous to suppression of a specific memory, as discussed in more detail below.

In summary, the brain regions relevant to working memory processes involved in intrusive thought likely involve (a) content-specific cortical regions needed to access the representation of information underlying the intrusive thought: visual areas for visual images, language- or semantically related regions for thoughts, limbic regions for emotional information, and hippocampus

for episodic memory, (b), basal ganglia mechanisms that influence what gains access to working memory (see Badre, this volume), and (c) prefrontal mechanisms involved in control processes that select and manipulate information within working memory or act to engender its removal.

An unresolved issue concerns the extent to which the processes deployed in suppressing contents actively held in working memory are distinct from or overlapping with those involved in suppressing the long-term memory retrieval of unwanted thoughts, prompted by reminders (see section on retrieval suppression), and whether these different types of mental control tap into unique processes that may be differently affected in psychiatric disorders.

Long-Term Memory

Memory Encoding

For over a century, it has been recognized that memories are initially unstable, subject to interference, until they are stabilized through a process of consolidation (Ribot 1882; Muller and Pilzecker 1900; Burnham 1903). Once consolidated, long-term memories are largely protected from interference, save windows of destabilization that may occur upon retrieval (Sara 2000; Nader 2003). Reconsolidation involves a relatively brief (few hours) cascade of molecular and cellular processes enhancing synaptic efficacy via structural changes; a longer process (days and weeks) involves system-level connectivity changes between the hippocampus and cortical areas (McGaugh 2000; Dudai 2012). Emotionally significant experiences trigger the release of adrenal stress hormones, stimulating norepinephrine in the amygdala, which in turn modulates plasticity in the hippocampus, cortex, and other brain regions (Cahill et al. 1994; Southwick et al. 2002; Cahill and Alkire 2003; Strange and Dolan 2004; Hurlemann et al. 2005). Abundant evidence links dysfunction in these circuits to psychological disorders in which intrusions are a symptom. Figure 9.3 presents a schematic overview of the neural circuits underlying these major domains and their possible (dys)function in intrusive thinking.

To the degree that encoding of information is tied to a specific event, such as that often associated with PTSD, aberrant functioning in the emotion circuits as well as the hippocampal and medial temporal structures that support episodic memory are likely involved (Liberzon and Martis 2006; Pitman et al. 2012). Evidence points to hyperactivity in the amygdala and dorsal anterior cingulate cortex (the putative human homologue of the prelimbic medial prefrontal cortex), whereas the ventromedial prefrontal cortex and hippocampus evince hypofunction, accompanied by impaired extinction learning and recall (Milad et al. 2009; Milad and Quirk 2012; Logue et al. 2018). Most recently, specific computations of threat, including learning parameters such as value, prediction error, and learning rate, have been characterized in PTSD. It was found that PTSD severity was related to overweighing of prediction errors
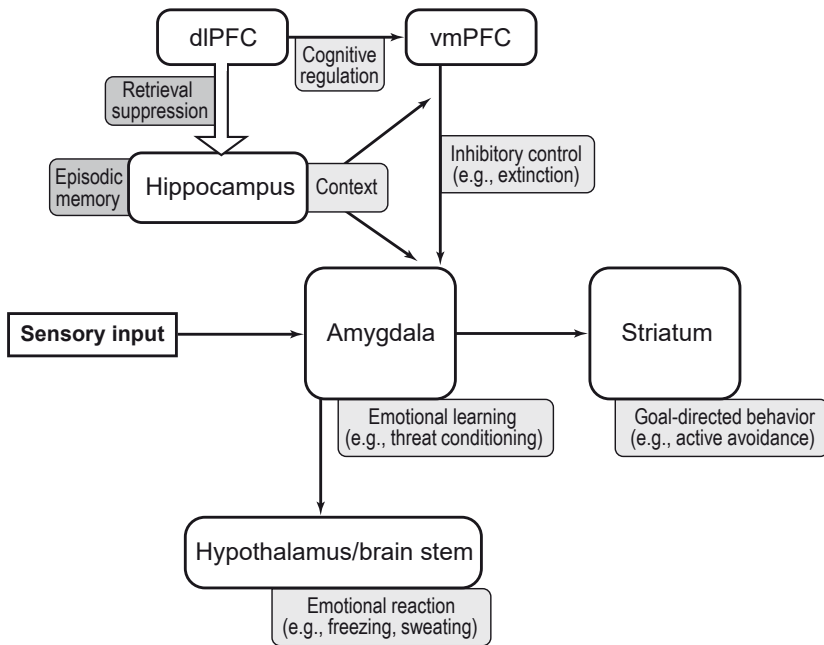
Figure 9.3   Most of what we know about the neural mechanisms of emotional learning and memory comes from animal studies that utilize threat conditioning and extinction as a model. Based on these studies, sensory information from neutral stimuli in the environment that reliably coincide with emotionally significant outcomes converge in the lateral nucleus of the amygdala where associative learning occurs, conferring emotional value on the conditioned cues. Projections from the lateral to the central nucleus engage descending projections to the hypothalamus and brain stem, which mediate the expression of the conditioned response (e.g., freezing). Projections from the lateral to the basal nucleus onto the striatum form a path that promotes active coping and goal-directed behavior. The prelimbic region of the medial prefrontal cortex (mPFC) connects with the amygdala to sustain the expression of emotional responses, while the infralimbic region acts to counteract amygdala output and diminish emotional responses. As such, the infralimbic mPFC (ventral mPFC in humans) is a critical region in the acquisition and recall of extinction learning. The hippocampus exerts contextual control over the expression of threat learning. The dorsolateral prefrontal cortex (dlPFC) is the region involved in top-down cognitive regulation of emotional responses by influencing amygdala via the vmPFC (Hartley and Phelps 2010; Schiller et al. 2010; Milad and Quirk 2012). The dlPFC also exerts top-down modulation of hippocampal activity to induce retrieval suppression (Anderson et al. 2016).

(akin to enhanced sensitivity to negative surprise) and to impaired amygdala and striatal tracking of the negative value of conditioned cues (Homan et al. 2019). Neuroimaging studies that examine the relation between viewing analog trauma and subsequent intrusions are sparse but do tentatively suggest that activation of regions in the salience network distinguishes distressing scenes that later intrude compared to equally distressing scenes that do not (Bourne et

al. 2013; Battaglini et al. 2016; Clark et al. 2016), highlighting that consolidation processes are important in the formation of intrusive memories. Together, these studies invoke neural mechanisms for the learning and flexible modulation of emotional memories. Aberrant functioning of the salience and memory circuitries might induce emotional inflexibility, lack of proper contextual control, and impaired ability to diminish inappropriate emotional reactions, allowing strong negative memories to persist and intrude.

Similar mechanisms may come into play if a memory is associated with particular actions or procedures, such as may occur in substance use disorders. Here, the encoding likely involves basal ganglia systems and alterations in specific cortical processes (Balleine, this volume), such as visual regions that may become tuned to particular stimuli (e.g., paraphernalia associated with substance use). Across disorders, when information is associated with emotional salience or significance (e.g., salience of negative emotional information in depression, association of loud sounds with a traumatic event in PTSD), such encoding likely involves the amygdala, as well as the striatum, insula, and dorsal anterior cingulate cortex (Fedota and Stein, this volume). However, whether intrusive memories are stored in a qualitatively different way than other emotional memories, or are merely strong memories and therefore more accessible, is still a topic of debate (Berntsen and Rubin 2008, 2013; Brewin 2014; Bisby and Burgess 2017).

Retrieval Suppression

Although we know little about the neural mechanisms of intrusions, we know a great deal about the inhibitory control processes deployed to reduce their occurrence. For an in-depth review of the neural basis of memory control, see Banich et al. (2009) and Anderson and Hanslmayr (2014). For a review of memory control in relation to emotion regulation, see Engen and Anderson (2018). For a detailed consideration of memory control in relation to primate anatomical pathways and neural circuits, see Anderson et al. (2016).

Prefrontal-Hippocampal Interactions as a Basis for Retrieval Suppression. Research on retrieval suppression was initially premised on the parallel between stopping prepotent actions in response to triggering stimuli and stopping internal processes, such as memory retrieval (in response to reminders), to control intrusive memories and thoughts (Anderson and Green 2001). In both cases, a stimulus (or, sometimes in the case of retrieval stopping, an internally generated cue) initiates an automatic process (action preparation or retrieval) that the person wishes to stop, and both trigger a race between a Go and a Stop process. Given this analogy, one might expect both similarities and differences between retrieval and action stopping: similarities in the prefrontal control regions engaged in service of stopping, but differences in the target regions with which control processes interact to implement stopping. Because retrieval

stopping involves stopping retrieval and not physical actions, a plausible suppression target would be the hippocampus, a brain structure involved not only in encoding new memories but also in their retrieval, at least for recently acquired events and thoughts (Anderson and Green 2001; Anderson et al. 2004; Anderson and Hanslmayr 2014).

Abundant evidence indicates that retrieval suppression, unlike motor response inhibition, relies on prefrontally mediated downregulation of activity in the hippocampus and other medial temporal lobe regions to stop retrieval, presumably by preventing pattern completion (Anderson et al. 2004, 2016; Depue et al. 2007; Banich et al. 2009; Levy and Anderson 2012; Gagnepain et al. 2014, 2017; Benoit et al. 2015; Schmitz et al. 2017). Suppression reduces hippocampal activation not merely relative to active retrieval in the Think condition, but also relative to passive baseline conditions, and this negative BOLD response arises from negative coupling between the right lateral prefrontal cortex and the hippocampus, established using effective connectivity analysis (Benoit and Anderson 2012; Gagnepain et al. 2014, 2017; Benoit et al. 2015; Schmitz et al. 2017). Indeed, within-subject comparisons of action and memory stopping establish a clear double dissociation, with retrieval suppression downregulating the hippocampus more than action stopping, but action stopping downregulating motor cortical areas (M1) more than retrieval suppression (Schmitz et al. 2017). Stopping of unwanted memories and thoughts thus involves a distinct frontohippocampal inhibitory control pathway that suppresses hippocampal activity. Hippocampal suppression appears to be a blunt instrument that acts globally on the hippocampal state. For example, when a person tries to suppress retrieval to depress awareness of a particular unwanted thought, the forgetting arising from that suppression is not limited to the item people intend to suppress; rather, any other recently encoded memories that occur either before or after the act of suppressing something are also forgotten, even if they are entirely unrelated to the content being suppressed (Hulbert et al. 2016). Thus, suppression of unwanted thoughts induces an *amnesic shadow* in the temporal surround of the suppression attempt, creating both anterograde and retrograde amnesia effects in healthy people. This finding has been linked to the global suppression of hippocampal processes that not only stop retrieval but also disrupt encoding and stabilization processes necessary to retain recent experiences (Hulbert et al. 2016). This global, systemic disruption of hippocampal activity is analogous to the global stopping identified in motor response inhibition.

What do we know about how the prefrontal cortex achieves this form of inhibitory control over hippocampal activity? Primate anatomical studies tell us that top-down suppression of hippocampal activity is unlikely to be direct, not only because there are no direct connections between the lateral prefrontal cortex and the hippocampus (Anderson et al. 2016), but also because long-range projections from the hippocampus are largely excitatory. To achieve an inhibitory effect in the hippocampus, if the negative BOLD response is truly inhibitory, the prefrontal cortex must drive local populations of inhibitory

interneurons within the hippocampus to disrupt its function. Because all inter-neurons in the hippocampus are GABAergic, this observation suggests that individuals with higher concentrations of hippocampal GABA may show a superior ability to suppress hippocampal activity by prefrontal influence. Recently, Schmitz et al. (2017) found evidence of this, using a multimodal imaging study that combined fMRI with magnetic resonance spectroscopy. These findings indicate that people with higher concentrations of hippocampal GABA showed greater downregulation during retrieval suppression, more successful forgetting of intruding thoughts, and greater negative coupling between the right prefrontal cortex and the hippocampus. As such, local concentrations of hippocampal GABA may provide a pivotal function that enables the prefrontal cortex to implement long-range inhibitory influence necessary for control over intrusive thoughts. This discovery sheds new light on evidence of diminished hippocampal GABA in many disorders characterized by intrusive thoughts (Schmitz et al. 2017), which may be a heretofore unrecognized risk factor in the pathogenesis of disordered control over intrusive thoughts.

Despite the unique hippocampal targets involved in implementing retrieval suppression, it is equally clear that both retrieval and action stopping processes engage overlapping regions in the dorsolateral (BA 9/46/10) and ventrolateral prefrontal (BA 44/45) cortex, suggesting the existence of domain general supramodal inhibitory control regions that may dynamically recouple with task-specific target regions (Depue et al. 2016; Schmitz et al. 2017; Guo et al. 2018). These domain general regions are strikingly right lateralized, strongly consistent with the long-standing claim by Aron et al. (2004, 2014) that inhibitory control is right lateralized, although the regions clearly include both dorsolateral and ventrolateral regions, and not simply ventrolateral prefrontal cortex. These supramodal regions have been identified both through within-subjects conjunction analyses of action and retrieval stopping (Depue et al. 2016; Schmitz et al. 2017) as well as conjunctions performed on quantitative meta-analyses of independent studies from many laboratories (Guo et al. 2018). Interestingly, action stopping and retrieval stopping also appear to engage highly colocalized regions within the basal ganglia (Anderson et al. 2016); for a detailed discussion of anatomical hypotheses, see Depue (2012), Guo et al. (2018), Paz-Alonso et al. (2013), and Balleine (this volume).

Cortical Modulation and the Reinstatement Principle.   Although the discussion of direct retrieval suppression emphasized hippocampal suppression as the principal mechanism through which intrusive memories and thoughts are controlled, the hippocampus is not the only target of inhibitory control during suppression (see Banich as well as Balleine, this volume). For example, when people suppress visual objects or scenes, downregulation takes place in the hippocampus as well as in visual cortical regions, such as the fusiform cortex and the parahippocampal place area, respectively. This leads to the

generalization that areas outside the hippocampus involved in reinstating the unwanted memory in awareness are suppressed in parallel with the hippocampus by the right lateral prefrontal cortex, an idea known as the reinstatement principle (Gagnepain et al. 2014; Gagnepain et al. 2017).

Intrusive Thoughts and the Triggering of Control.   Intrusions of unwanted contents into awareness during retrieval suppression are particularly important for triggering top-down inhibitory control processes that suppress the hippocampus and cortical regions. Using the intrusion judgment procedure outlined earlier, Levy and Anderson (2012) found that downregulation of activity in the hippocampus and other medial-temporal lobe regions was largely confined to trials in which participants reported that a memory had intruded into consciousness. Strikingly, the extent of hippocampal downregulation during intrusions predicted subsequent forgetting with a correlation of 0.7, whereas hippocampal activity during non-intrusions was unrelated to later forgetting. These findings suggest that intrusions play an important role in triggering top-down control by the prefrontal cortex to cancel the retrieval process, and that the critical inhibitory action that disrupts later retention of the intrusive thought arises in the purging of the intrusion from awareness. Consistent with this possibility, Gagnepain et al. (2017) replicated Levy and Anderson's (2012) evidence for intrusion-specific downregulation in the hippocampus with aversive scenes and showed that negative coupling between the right dorsolateral prefrontal cortex and the hippocampus was significantly greater during intrusions than during non-intrusions.

The findings by Gagnepain et al. (2017) are especially relevant to the current discussion, given that they involve the suppression of unpleasant and intrusive images, prevalent in psychiatric disorders. Gagnepain et al. (2017) also found intrusion-specific downregulation in both the amygdala and the parahippocampus. These downregulations robustly predicted both intrusion frequency and later reductions in negative affect for the suppressed content: the stronger the downregulation in the amygdala, the lower the number of intrusions and the greater the subsequent reduction in negative affect for the suppressed scene. These findings indicate that top-down inhibition during intrusive thoughts plays a critical role in modifying the representations that support both memory and emotion about the offending content. Consistent with this, Legrand et al. (2019) found that suppressing unpleasant images from awareness also significantly reduced later psychophysiological measures of emotion elicited by the scenes, such as heart rate deceleration. Similar findings have now been observed with skin conductance responses (Harrington et al. 2020).

The importance of purging intruding thoughts from awareness highlighted here points back to the discussion above of what relationship, if any, the processes involved in stopping retrieval have to those involved in regulating working memory. One assumes that when an intrusion is retrieved and enters awareness, there is a good chance that the intruding content has entered

working memory, however briefly. Some evidence supports this. For example, Hellerstedt et al. (2016), using event-related potentials, found evidence that the frontal negative slow wave (NSW), observed over the right prefrontal cortex, is modulated during intrusions. This component has been linked in prior work to the storage of information in working memory. Consistent with this, participants during Think trials show a prolonged NSW that lasts throughout the full several seconds of the trial; in contrast, non-intrusions show little evidence of this, consistent with the exclusion of items from working memory. Intrusions, however, showed a brief increase in the NSW, which was rapidly eliminated within the first seconds of the trial, suggesting a brief penetration of the intruding item into working memory. Perhaps relatedly, Castiglione et al. (2019), using time frequency analysis, found evidence that during No-Think trials, there is a robust increase in frontal beta component during non-intrusions. Given the prior linkage of this component to motor response inhibition, these findings suggest that intrusions may reflect an initial failure of inhibitory control that allows the intruding content to penetrate working memory.

Related Phenomena Observed with Item-Method Directed Forgetting. Although we have focused primarily on retrieval suppression, related work on item-method directed forgetting supports the hypothesis that, in parallel with retrieval suppression, people can also suppress encoding. By encoding suppression, we mean the possibility that the same inhibitory control processes which modulate hippocampal activity during memory retrieval to disrupt retention may also be deployed shortly after encoding to terminate stabilization processes in the hippocampus that might promote the formation of an enduring memory. For example, several studies indicate that when participants are instructed to forget an item that they just encoded into memory on the preceding trial, activation increases in the right dorsolateral prefrontal cortex and decreases in the hippocampus. As with retrieval suppression, connectivity analyses indicate that the prefrontal cortex couples with the hippocampus during Forget trials, especially on trials when the item is successfully forgotten (Rizio and Dennis 2013; Wierzba et al. 2018). In another compelling study, intracranial recordings indicated that lateral prefrontal cortex interacts with the hippocampus during instructions to forget to promote forgetting (Oehrn et al. 2018). These findings converge to suggest that suppressive processes are not limited to controlling intrusive retrievals, but may be also be used prophylactically shortly after an unpleasant experience to limit the footprint of that experience in memory. Moreover, this application of inhibitory control to suppress unwanted contents immediately after they are encountered seems related to inhibitory processes involved in purging the contents of working memory (e.g., Holmes et al. as well as Banich, this volume), although these two strands of research are not usually considered together.

Summary and Commentary

At present, little is known about the neural underpinnings of intrusive thinking, and the underpinnings presumably vary across its different manifestations. In contrast, research on cognitive control and emotional memory has identified key systems involved in generating and suppressing intrusions. Concerning intrusion generation, areas involved in signaling salience, such as the amygdala and striatum, not only play a role during the encoding of salient events or thoughts, but also respond to reminders of those events (external cues). Their salience signal may help bring back to mind memories of these events or thoughts, possibly via reinstatement of multimodal cortical representations of these events. In addition, these areas may themselves generate internal cues (mood states) that trigger intrusions. With regard to the suppression of intrusions, frontal areas implicated in inhibitory control over actions and thoughts play a clear role. In particular, the dorsolateral prefrontal cortex can suppress the reinstatement of a memory or of an imagined future event by downregulating hippocampal and neocortical activity, while also downregulating emotional responses via amygdala suppression. Suppression reduces the frequency of intrusions and impairs memory, also for new information presented around the time of retrieval suppression (i.e., creating an *amnesic shadow*). Some evidence indicates that it also reduces negative affect associated with suppressed content.

What Are the Implications of Intrusive Thinking?

In this chapter we have revisited the definition of intrusive thinking by systematically considering all the circumstances in which intrusions might occur and their manifestations across health and disorders. We define intrusions as being *interruptive, salient, experienced mental events* and propose that clinical intrusive thinking differs from its nonclinical form with regard to frequency, intensity, and maladaptive reappraisal. We have reviewed the neurocognitive processes underlying intrusive thinking and their control, including action control, working memory processes, long-term memory encoding, retrieval and suppression, and methods for studying them.

Functional Perspective: The Adaptive Nature of Intrusive Thinking and the Desirability of Suppression

Despite being commonly associated with mental health disorders, intrusive thinking commonly occurs in the absence of psychological problems and is thus by itself not indicative of pathology. In fact, there are many instances in which having a thought pop into mind to disrupt ongoing cognitive processes is actually beneficial, such as recalling an action that is required (e.g., paying

a bill) or solving a problem (an "aha" moment) (see also Monfils and Buss, this volume). If mind wandering is a form of intrusive thinking, research shows that the capacity to "jump out of the present set" can be adaptive and relates to creativity. Its adaptiveness is likely reduced when mind wandering becomes excessive or too diffuse, such as in ADHD, or when the content is unpleasant.

The potential adaptive nature of intrusive thinking may be independent of the desirability to control it. In healthy individuals, frequent and exciting thoughts (e.g., love infatuations) might disrupt concentration on the task at hand and therefore not be adaptive, yet still wanted in the moment. Alternatively, intrusive thinking about mistakes that we have made may allow us to adjust our behavior and become a better person, yet unwanted in the moment. The discussion below on sociocultural implications provides compelling examples of this. In addition, as a clinical symptom, intrusive thinking is not always unwanted (e.g., exciting flashforward thoughts that occur during a manic episode), yet often maladaptive. Whether the opposite is also possible (i.e., unwanted intrusive thinking in mental health disorders may in some cases serve an adaptive purpose, such as in acute stress disorder as part of processing trauma), is an open question that requires more research.

Clinical and Therapeutic Implications

Could excessive intrusive thinking be an endophenotype (or neuroendophenotype) for mental disorders? According to the novel, more dimensional approach to understanding psychiatric nosology (Cuthbert and Insel 2013), the propensity toward intrusive thinking could underlie several otherwise distinct, categorically defined disorders, such as OCD and depression, the implications being comorbid, shared dysfunctions of common neural systems or networks, with obvious implications for treatment. Considering some of the dimensions already postulated, there are clear relationships with such constructs as inhibitory control and working memory. It remains unclear, however, whether intrusive thinking comprises a unitary construct itself or is a collection of phenomena that meet our definitions. For example, intrusive thinking could arise as an emergent feature of different neural networks processing perceptual inputs, on one hand, or neural networks underlying internal factors such as intentions and mood states, on the other. However, such malfunctions in different networks could depend, for example, on a common molecular or neurotransmitter deficit. This issue can perhaps best be resolved when we have firmer information about the neural substrates of intrusive thinking. There are some promising indications of this (see chapters by Philips, Fedota and Stein, Balleine, Badre, Gourley et al., and Roberts et al., this volume) as well as possible genetic relationships, which will depend on having precise definitions of the phenotype obtained, for example, through definitive objective tests and a standardized general instrument for assessing intrusive thinking.

The mechanisms and paradigms discussed have implications for therapeutic applications. An idea often discussed in the clinical literature is that suppressing intrusive thoughts in clinical disorders is counterproductive, an idea substantially influenced by Wegner's white bear thought suppression procedure (Wegner et al. 1987; Wegner 1997). However, substantial recent evidence with the Think/No-Think procedure suggests that findings from the white bear paradigm have been overgeneralized and that other processes, such as retrieval suppression, are effective in modulating thoughts, at least in healthy individuals. If so, there may be situations in which suppressing thoughts could be useful to improve an individual's ability to function, even in clinical disorders. For example, it may be that both trauma-focused interventions, such as eye movement desensitization reprocessing, and paradigms using cognitive interference, such as memory retrieval procedure and visuospatial task (Holmes et al., this volume), which appear to be effective in reducing the frequency of intrusive memories of trauma, work through a process similar to the retrieval-suppression mechanism described above. While it is clear that in some disorders (e.g., OCD), attempts to suppress intrusive thoughts can lead to rebound, clinical disorders involving intrusive thoughts are heterogeneous, as are patient populations who have these disorders. Thus, it is critical to explore the therapeutic implications of paradigms which have demonstrated efficacy in controlling, replacing, and suppressing intrusive thoughts in human laboratory settings (see also Brewer et al., this volume).

A particularly promising therapeutic approach for fostering the effective control of intrusive thoughts is mindfulness-based cognitive therapy, which integrates meditation techniques with cognitive behavioral strategies (Külz et al. 2014). One of the primary skills taught in such approaches is the ability to control one's thoughts to focus on breathing. The act of releasing thoughts in mindfulness practices is highly reminiscent of the control strategies invoked in the Think/No-Think paradigm, further illustrating their potential pertinence to the treatment of intrusive thoughts. Mindfulness practices may also be helpful in furthering individuals' meta-awareness of having intrusive thoughts in the first place (Baird et al. 2014). Specifically, such practices may enable individuals to identify episodes of unwanted thoughts that might have otherwise been experienced but evaded explicit acknowledgment (Baird et al. 2013a; Takarangi et al. 2014). Identifying intrusive thoughts in the light of meta-awareness may enable individuals to invoke the necessary mental control strategies required to release them.

Sociocultural Context

Intrusive thinking and its control (or lack thereof) take place within a sociocultural context. Examples of this include the AIDS-HIV epidemic of the 1980s. In his book, *The Man Who Couldn't Stop: OCD and the True Story of a Life Lost in Thought*, David Adam (2014) provides a vivid example of how frequent

concerns about infection drove the induction of his own obsessive and compulsive symptoms. Certainly, there are many other examples of widespread general concern (e.g., nuclear war, Brexit, global pandemics) that intrude into our everyday consciousness and may lead to pathological consequences.

Although many of us successfully resist such concerns, this too can lead to societal consequences, as can be currently observed through the striking example of the climate crisis. Since the Industrial Revolution, global temperatures have increased by 1°C and humankind is well on track to experiencing an additional increase of 0.5°C by 2030 (Xu et al. 2018). In line with the worst-case scenario put forth by the Intergovernmental Panel on Climate Change (IPCC 2018), this half a degree will likely correspond to a 50% increase in extreme weather events worldwide (e.g., droughts, floods, snowstorms, hurricanes, cyclones) and exact devastating consequences, including mass migration, agricultural failure, and deadly heat events (Xu et al. 2018). The near total consensus of climate scientists, exemplified by the Paris Agreement, which almost all governments signed (and almost none are honoring), is that we have to reduce emissions soon. Failure to do so will result in continued temperature increases that will soon be beyond human control and ultimately lead to a "hothouse planet" by 2100, or perhaps even sooner (Wallace-Wells 2019).

The climate crisis is surely creating daily thought intrusions in hundreds of millions of people. Such intrusions are likely characterized by interrupting, salient, experienced events (imagery, emotion, moral feeling) that recur. This example illustrates the complexity of judging whether intrusive thinking is maladaptive or not, and the range of responses that people can have. For instance, against the unimpeachable backdrop of scientific knowledge, such intrusions appear highly adaptive: *they compel action* and yet only a very small minority of people *are* currently engaged in action—the great majority of global citizens are not taking action. Of these, some may deny the science or the predicted impacts, or they may accept the science and impacts but deny that any serious action is warranted (e.g., because they are ideologically committed to the current economic system). In Western societies, polling shows that most people fall in the latter category (accepting the science and the impacts), yet they are not acting beyond some minor adjustments in their personal lives, possibly because action would be inconvenient to one's career or lifestyle or because it would require confronting grief and fear in a way that one is not yet prepared to do. Since such people do understand and accept the terrifying imperative to act, but are not doing so, they could be characterized as exerting control over the intrusive thinking caused by the climate crisis. The form of control being used is probably based on reappraisal: people express degrees of fatalism ("the problem is too big or too hopeless," "it is too late"), nihilism ("humans deserve what is coming to them"), a deferral of responsibility to policy makers ("it is a government problem"), or presently unwarranted faith in technology alone to solve the problem (Hansen 2018). Nonetheless, as more extreme weather

events, for instance, increase over time, so too will levels of anxiety, until the intrusion breaks through the threshold of control and people clamor for action.

In summary and in accordance with the definition of intrusive thinking as interrupting, salient, experienced events (imagery, emotion, moral feeling) that are also recurrent, the climate crisis is clearly generating intrusive thinking in a wide range of people around the world. This, in turn, creates conflicts which can inculcate attempts at control (e.g., at the reappraisal level). Such tensions exacerbate poor mental health as well as the feelings of instability and fear, which are already driving populist political regimes (Latour 2018). Meanwhile fake news disseminated within the broader system functions to degrade the salience signal that is currently essential to generate the type of population-level intrusive thinking that would compel quick action. This sociocultural example is interesting because it represents the flip side of many of the psychiatric-related examples discussed in this chapter. Whereas intrusions were often characterized as pathological, the climate crisis example demonstrates how intrusive thinking is a good thing for our biosphere and civilization, and attempts to control these intrusions are maladaptive. The survival of human civilization depends on more intrusive thinking right now.

Concluding Summary

Our discussion focused on the psychological bases of intrusive thinking and its control that may occur every day in healthy individuals, as well as in psychiatric disorders. We surveyed the range of phenomena that can be construed as examples of intrusive thinking and endeavored to reach satisfactory definitions and classifications of the different forms of intrusive thinking that will aid further research. The least constrained definition was one emphasizing the interruptive, salient, and experienced nature of intrusive mental events (as compared with a common definition which specifies unwanted and conscious, as well as interruptive criteria). Recurrence is a further property which may be more important for psychiatric manifestations. Agency, meta-consciousness, and (mal)adaptiveness or desirability are considerations that further define the boundaries of what can be considered as intrusive thinking. Based on this analysis, PTSD and OCD appear to be prototypical disorders of intrusive thinking, although its elements appear in a range of other psychiatric diagnoses, including addiction and depression, though probably not psychosis.

The main part of our discussion focused on neurocognitive mechanisms of intrusive thinking and its control, which included impairments particularly in the regulation of memory retrieval, as well as of affect and action control. We brainstormed future possible approaches to investigating intrusive thinking, with priorities for designing a new, theoretically motivated questionnaire, refining existing objective measures and capitalizing on the potential of

computational approaches. Finally, we considered the importance of this in the context of broader social-cultural and clinical-therapeutic issues.