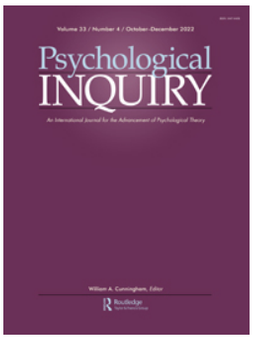

Routledge
Taylor & Francis Group

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hpli20

Clarifying Internally-Directed Cognition: A Commentary on the Attention to Thoughts Model

David R. Vago, Norman Farb & R. Nathan Spreng

To cite this article: David R. Vago, Norman Farb & R. Nathan Spreng (2022) Clarifying Internally-Directed Cognition: A Commentary on the Attention to Thoughts Model, Psychological Inquiry, 33:4, 261-272, DOI: 10.1080/1047840X.2022.2141005

To link to this article: https://doi.org/10.1080/1047840X.2022.2141005

PSYCHOLOGICAL INQUIRY
2022, VOL. 33, NO. 4, 261–272
https://doi.org/10.1080/1047840X.2022.2141005

Routledge
Taylor & Francis Group

COMMENTARIES

Clarifying Internally-Directed Cognition: A Commentary on the Attention to Thoughts Model

David R. Vago[a,b], Norman Farb[c], and R. Nathan Spreng[d,e,f,g]

[a]Department of Psychology, Vanderbilt University, Nashville, Tennessee; [b]Contemplative Sciences Center, University of Virginia, Charlottesville, Virginia; [c]Department of Psychology, University of Toronto Mississauga, Mississauga, Canada; [d]Laboratory of Brain and Cognition, Montreal Neurological Institute, Department of Neurology and Neurosurgery, Faculty of Medicine, McGill University, Montreal, Canada; [e]Departments of Psychiatry and Psychology, McGill University, Montreal, Canada; [f]Douglas Mental Health University Institute, Verdun, Canada; [g]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

Introduction

Self-generated thought is one of the most ubiquitous and familiar human experiences, a 'stream of consciousness' uniquely relevant to the individual and their experience. Internally-directed cognition (IDC) represents the internally-oriented stream of conscious thought and associated cognitive processes that occupy inner mental life during wake and sleep.

Given the limitations of a global workspace for working memory and attentional capacity, the degree to which we prioritize the immediate internal or external environment depends on current goals, motivation, level of alertness, the complexity of the ongoing task demands and the context in which it is being performed. At any given moment in time, the stream of human thought is rich and self-relevant, reflecting deliberately or automatically constrained content. It most often reflects an individual's greatest personal concerns, interpersonal feelings, future goals and plans, creative incubation of ideas, hopes and fears, unresolved challenges and decisions, simulations of the past or future, self-monitoring and evaluation, daydreams and fantasy, intrusive memories, and other constructed experiences.

Given how fundamental IDC is to the construction of self-identity, human behavior and particularly social interactions, a comprehensive theoretical model that clarifies the form, content, frequency, timing, neurobiology, and adaptive nature of IDC would be beneficial for the study and understanding of human cognition. Over the last decade, there has been considerable progress from the psychological and neurobiological literature to elucidate some of the functions, mechanisms, and potential theoretical understanding of processes subserving IDC, including selective attention and associated content that occupy working memory (Andrews-Hanna, Smallwood, & Spreng, 2014; Axelrod, Rees, & Bar, 2017; Baird et al., 2012; Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Chun, Golomb, & Turk-Browne, 2011; Fox, Andrews-Hanna, & Christoff, 2016; Fox & Christoff, 2018; Hommel et al., 2019; Spreng, Mar, & Kim, 2009; van Vugt, van der Velde, & ESM-MERGE Investigators, 2018). Yet, conceptual gaps remain and there is a paucity of empirical work on IDC.

The Attention-to-Thoughts (A2T) model proposed in this issue by Amir and Bernstein is a formal computational and theoretical model that provides a novel algorithmic approach to simulate and predict how internal attention selection is expressed from moment-to-moment, including the psychosocial relevance of contextual demands, and modulation by neurobiological systems supporting working memory and emotion. The A2T model describes IDC as encompassing self-generated or endogenously driven attention to thoughts and mental imagery as well as working memory processes and associated interactions with long-term memory stores. Amir and Bernstein focus on internal attention selection, including the content and the act of processing selected information. Although the A2T model provides some helpful computational and descriptive aspects of the intended (mal)adaptive phenomenology, the model also contains limitations relative to the extant theoretical and empirical literature. Here, we reflect on IDC and assess whether it is: (i) accurately operationalized by the A2T model, (ii) sufficiently reflective of the extant literature, and (iii) adequate for widespread utilization. We describe some of the most salient issues to inform the further development of the model so it can be considered a general model of attention to internal thought and expand the model to include other aspects of IDC.

The most general initial response to the target article is that IDC represents a rich variety of internal experiences that are not captured by the relatively narrow formulation expressed in the A2T model. We describe the formulation as narrow because the A2T appears to fixate on the very specific scenario of engaging in an attentionally-demanding internal cognitive task, while resisting attentional selection of other intrusive and disruptive internal representations. To this end, the model provides valuable formalization of dynamic processes that emphasize maladaptive forms of "higher order IDC" processes implicated in mental health, including, (1) repetitive negative thinking and mind-wandering, (2) the influence of cognitive dyscontrol

CONTACT David R. Vago david.vago@vanderbilt.edu Department of Psychology, Vanderbilt University, 2301 Vanderbilt Place, Wilson Hall, Nashville, TN 37240-7817, USA.

on attentional selection, (3) the competition between external vs. internal attentional processes, and (4) modulation of IDC by working memory, emotional reactivity, and cognitive fusion. Yet, while there are undoubtedly many situations where resisting distraction by negatively valenced, internal thought processes, is the chief "problem" for the mind to solve, IDC must also balance a variety of other constraints that are not formally included in the model. In the following sections, we evaluate the specificity and sufficiency of the A2T model and focus on several considerations for expansion and growth which address some of the limitations operationalizing and modeling IDC. If these limitations could be addressed in an expanded version of the A2T model, it may help elucidate IDC and basic principles for the deployment of attention in IDC contexts as well as contribute to a more comprehensive framework for IDC.

Specificity and Sufficiency of the A2T: Suggestions for Expansion and Growth

There are at least two levels at which the adequacy of the A2T model can be evaluated: specificity and sufficiency. By *specificity*, we mean whether the model provides enough granularity to explain its principal exemplars. By *sufficiency*, we mean whether the model captures a diverse enough set of scenarios to adequately predict the internal selection of attention and reasonably approximate the breadth of IDC. When evaluated for specificity, the A2T model provides strong explanatory power for a particular subset of IDC, namely attentional selection, cognitive demand, and the disruptive influence of Task-Unrelated Thought (TUT). The model emphasizes the modulatory influence of affective tone and current working memory processes on internal attentional selection.

The A2T model provides good specificity as a thought dysregulation model for generating and sustaining dysphoric rumination, anxious thoughts, and associated biased attention; yet its current formulation seems insufficient as a comprehensive characterization of IDC. Specifically, there is inadequate consideration of "internally-directed" beyond forms of dysregulated thinking. Within this area of inquiry, there are a number of other existing models that highlight the tendency for cognitive dyscontrol and negative orientation of self-reflective processing, including a process-occurrence framework (Smallwood, 2013), a focus on cognitive control failures (McVay & Kane, 2010), and a product of spontaneous thought (Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016). Most of these existing models and others (Farb et al., 2007; Fox et al., 2016; Klinger, Marchetti, & Koster, 2018; Spreng et al., 2009; van Vugt et al., 2018) clearly describe how mind-wandering and rumination are active with TUT during low cognitive demand and arise independently of perceptual input. Each model provides some further characterization of IDC; yet, there remain conceptual gaps and room for further development reflect the current state of the literature in this field.

Internally-directed cognition (IDC) is an umbrella term for a variety of distinct cognitive process and ubiquitous categories of internal representation that are not fully characterized by the A2T model. A few conceptual areas and processes that are alluded to, but not sufficiently described include, the role of interoceptive sensory processing, prospective forms of memory, meta-cognition and planning, mindfulness, and constructive forms of IDC such as creative idea incubation, analytical problem solving, goal setting, and most other mental computations that are not limited to processing of external sensation or enacting external behaviors. Put simply, internal life is so rich and varied that a comprehensive model requires an extraordinary number of potential influencing factors if it hopes to capture even a fragment of this richness. Herein, we suggest that the A2T model may be too narrow precisely because the authors claim to provide a general account of the qualities of IDC from which mental life emerges. Concrete points of concern, critique, and suggestions for improved specificity and sufficiency are indicated in italics throughout the text.

Below we suggest specific areas for clarification, expansion, and growth to improve specificity and sufficiency of the A2T model, grounded in a broader conceptualization of IDC. Five specific critiques are offered: (1) Narrow *criteria for attentional selection* (i.e., degree of negative affect) (2) Insufficient *characterization of contextual demands*, (3) Limited *characterization of task-(un)related thinking*, (4) Confounding *contributions of working memory, internal, and external attention*, and (5) Inadequate *neurobiological specificity*.

Narrow Criteria for Attentional Selection

The A2T model emphasizes the degree of negative affect as the key dimension by which attentional selection of content is based. This characterization of attentional selection is primarily based on the high frequency of negatively-valenced, self-reflective processing often reported in association with mind-wandering, rumination, and negative correlation with wellbeing (Killingsworth & Gilbert, 2010; Smallwood & Schooler, 2006). Although it is not explicitly stated in the model, "degree" is assumed to be a dimensional variable of intensity associated with the current emotional experience, and by which the stimuli competing for attentional selection are modulated. Amir and Bernstein emphasize high intensity negative affect and cognitive reactivity increase the likelihood working memory will be populated by mood-congruent memories and thought content. They further suggest a reciprocal relationship between high intensity negative affect and thought content with phasic influence of affect over attentional selection. High intensity emotional experience involves a broad range of factors that influence attentional selection, including immediate personal relevance, sympathetic arousal, feelings of threat or awe that are variable with valence of the experience and current mood, and depth of memory encoding. The more intense the emotional experience, the more likely depth of memory encoding and selection history will play a strong role in biasing future attentional selection. Attentional selection is more likely to be influenced by memories that are encoded deeply in

comparison to low intensity emotional experiences. Depth of encoding also refers to the degree to which selection history is populated by habits of conditioned thoughts and behavior. Habit and bias are continually informed by complex feed-forward and feedback between cognitive control processes, sensory-motor experience, and context-relevant cues, triggers, and reward histories. These dynamics continually influence working memory from associative long-term memory stores and impact a predictive model for future thought and behavior. The A2T model predicts the current state of internal selection through the representations in working memory at any one moment (time t) and that content's influence on affective tone and the affective tone's influence on internal attention in any subsequent moment of time (time $t+1$).

The content of negative thoughts, the processes subserving their generation and maintenance, and the short and long-term impact on physiology are likely relevant to attentional selection. For example, some IDC including attention to negative thoughts and feelings may have dissociable influences on mind-brain-body interactions depending on whether they are self- or other-focused. Many negative thoughts and mental sets that are self-ruminative have shown to have greater negative impact on physiology through perseverative stress-mediated modulation in comparison to self-reflective, distanced, or other-focused negative thoughts (Takano, Sakamoto, & Tanno, 2011). Ruminations are characteristically perseverative and involve passive focusing on symptoms of distress, with fixedness and inflexibility (Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008). The consequence of negatively-valenced thoughts, feelings, and behaviors modulate contextual relevance and demand and can amplify the emotional expression and subsequent attention. Once perseveration and magnification overwhelm the current state of the system, the question remains how to model the shift in awareness or break from TUT and return to task-congruent thinking. Although there has been some attempt to tease apart mind-wandering with awareness in comparison to mind wandering without awareness, the A2T model does not clarify how internal attention specifically determines which information rises to awareness spontaneously or in relation to causal factors determined by past associations from memory.

Negatively valenced thoughts may be experienced across individuals with equal intensity, yet different coping strategies and dispositions influence how attention is modulated uniquely across contexts. For example, rumination has differing effects on individuals with unique contributions of cognitive styles, such as negative inferential styles, dysfunctional attitudes, hopelessness, pessimism, self-criticism, or neuroticism (Nolen-Hoeksema et al., 2008). Intensity of negative thoughts may also change future attentional selection depending on any existing psychopathology or pharmacological influence (Verhaeghen, Joorman, & Khan, 2005). Neural and genetic correlates also likely influence the trajectory of such cognitive styles; yet, the influence of predispositions to anxious or depressogenic thoughts or behavior on IDC have not yet been properly characterized in the A2T model. *The A2T model may benefit from a more granular description of maladaptive phenomenology that expands upon existing models of rumination and attentional bias. Specificity in degree of emotional expression will improve through matching thoughts with contextual behaviors/activities, feeling states, and behavior and provide more clarity around the mechanisms supporting dynamic generation and streams of thought that make up IDC.*

To improve sufficiency, there are critical processes underlying IDC that are underemphasized in the A2T model, including interoceptive sensory processing, prospective forms of memory and planning, and constructive forms of IDC like creative idea incubation and analytical problem solving. The relationship between ruminative thinking, psychopathologies, and more adaptive forms of self-reflection (e.g., positive affirmations) can also be further clarified to improve our understanding of IDC. For example, adaptive forms of thinking may influence the frequency or offset of negative thoughts and positive influence on attentional selection. Although an unresolved problem may often be negatively oriented, TUT may also be positive or neutral in valence. For example, an individual may have a creative project in mind or a riddle to figure out that influence's attentional selection, relevant executive functions, and behavior. Mechanisms driving adaptive vs. maladaptive types of TUT would be a helpful clarification for IDC more generally.

Self-evaluation, self-identity, and group-identity are also influenced by IDC. How an individual identifies with his/her own thoughts concerning themselves helps construct a sense of self (Vago, Gusnard, & Silbersweig, 2024; Zahavi, 2003). Self-reflective thinking can be verbal and sub-vocal with clear grammatical syntax, some of which may be non-verbal, symbolic, but temporally constrained to context and motivation (e.g., evaluating one's performance on an exam) and modulated by the frequency and valence of such mental habits. Negatively oriented sensory-symbolic and/or cognitive-linguistic evaluative processes resulting from the tuning of attention toward self-reflection are critical to the ruminative process which has shown to reify a psychopathological sense of identity and bias perception over time (Nolen-Hoeksema et al., 2008). Patterns of negatively-valenced repetitive thinking therefore become the mental habits that define a negatively oriented self-identity narrative. The greater frequency with which self-focused thoughts are the object of one's attention decreases the likelihood that other non-congruent thoughts can compete with and alter self-identity. Cognitive control deficits are described in relation to the tendency to continuously interpret ambiguous information as negative. When negatively oriented information dominates the self-narrative, vulnerability to depression and anxiety increases. *The A2T model suggests that it is this reification process that influences the stickiness of thought and prevents disengagement or "random exits" from repetitive negative thinking. Accounting for these aspects of self-identity may improve the sufficiency by which the model can predict breaking the ruminative cycle.*

Similarly, positive affirmations about oneself, or an optimistic view of oneself in relation to the future or the past,

can also influence self-esteem and improve performance on a task. Research on mindset has shown how positively-oriented beliefs or placebo can influence and increase the likelihood of a positive outcome (Jamieson, Crum, Goyer, Marotta, & Akinola, 2018). How an individual feels judged by others can also have a strong impact on self or group identity. The stereotype threat literature (Spencer, Logel, & Davies, 2016) emphasizes a potential framework for social influence on self, group-identity, and the influence of social context on performance. In this regard, there is a widely replicated social cognition phenomenon that illustrates the power of contextual features to impair rather than enhance performance. Stereotype threat is not about attentional selection of negatively-valenced information into working memory, at least not as it is conventionally described. Instead, stereotype threat operates through largely implicit/unconscious activation of a nocebo effect around performance. The more diagnostic and important the test is seen to be (higher contextual features), the stronger the impact of the stereotype threat. *The A2T model would benefit from the incorporation of self-specific, implicit processes (e.g., stereotype threat) and other contextual features such as mindset into conditions that influence attentional selection.*

Insufficient Characterization of Contextual Demands

A universal framework for understanding internally-oriented attention is well-framed in the opening pages of the A2T target article. As presently defined, the A2T framework pits internal representations against both external ("context") demands, and internal distraction, in the form of attention to task-irrelevant internal representations. This description has good ecological validity for situations where sustained attention is required, and mind-wandering is unequivocally counterproductive. External demands can include any sort of sensory cue (physical or social), whereas internal demands include two classes of examples—valued internal and goal-directed processes (intended targets of attention) and unwanted internal and stimulus-driven processes, with "involuntary" memory activation used as an exemplar. *While spontaneously arising memories are a fair example of the type of distracting internal representation that may capture attention as described in the A2T model, a more comprehensive framework for self-generated cognition would benefit from integration of a broader variety of internal representations across contexts.* For example, Smallwood and Andrews-Hanna (2013) provided a context-regulation hypothesis to describe how the variety of contexts determine the extent and nature of the demand upon our current experience, whether in passive rest or during an ongoing task (Smallwood & Andrews-Hanna, 2013).

There is clear support in the theoretical and research literature for the proposed moderators of internal attention selection as illustrated in Figure 1 of the target article, proposed moderators of internal attention selection as described by Amir and Bernstein in the A2T model, including contextual demands to attend externally (path a), the "autoregressive effect" of working memory to persist across time (path d), and the salience of representations laden with negative emotion (path e). Attention operates on the existing contents of working memory to either help perpetuate them (path d), or to switch away from them when provided with alternatives (context and/or emotion). The A2T model emphasizes the influence of contextual demands to selectively attend to specific information that is relevant to an ongoing task and attenuate influences of other competing representations from long-term, prospective or working memory. High task demands are described as task-oriented contexts and low task demands are described as mind-wandering states. It suggests changes in stimulus-driven signals requiring cognitive resources either active in working memory or retrieved from long-term memory influence the fluctuations of selection more heavily in comparison to internally-motivated signaling that have less task-specific demand on attention. Contextual demands are modulated by cognitive (dys)control, arousal, and emotional reactivity. Simulations of the A2T model demonstrate how demand may contribute to the spontaneous accounts of IDC across time. For example, high contextual demand on neutral information through sustained attention (e.g., focusing on memorizing a list of numbers) decreases the influence of internally-directed attention or affect. When there is low demand for focused attention (e.g., driving down a straight highway with low traffic conditions and in a familiar or monotonous environment), the model predicts more frequent fluctuations and intensity of internal attention and affect, such that mind-wandering is more likely. This common conceptualization emphasizes a dependency on cognitive dyscontrol to favor IDC where there is competition between attention toward goal-relevant representations and away from TUT.

There is mention of competition between internal and goal-directed processes that influence attentional selection through automatic and deliberate constraints as well as contextual and mnemonic constraints, yet very little effort is made to elaborate upon such constraints. Other dynamic constraints and features of context may influence attentional selection beyond the demand on processing resources that fluctuate between internal or external orientation. Christoff et al., 2016 describe the variety of TUT that are either spontaneous, automatically constrained, or deliberate. Fluctuations in the different constraints, as well as sensory-motor, affective, and arousal levels may all help determine what information passes through the attentional bottleneck with different time scales. Each type of thought has unique influences on orienting, filtering relevant information, and the processes subserving monitoring, orienting, engagement and disengagement. For example, if the deliberately constrained goal is to rehearse a particular sequence of numbers for recall later, an individual may still lose (and regain) attentional engagement even as the demand continues to remain high. There may be competing tasks deliberately set earlier with a particular frequency of task switching between rehearing numbers, typing an email, or ordering the objects on a desk. The number of tasks to switch between will shift upon some satisfactory level of completion of each task,

lowering demand for switching with each completion and raising demand as each prospective task operating in the background increases.

Automatically constrained mind-wandering to something highly emotional may also fluctuate with spontaneous daydreaming that draws upon memory retrieval processes. Context demand may be operationalized more appropriately to reflect attention to environmental cues relevant to a specific task. The model assumes that attention to self-reflective or ruminative thoughts reflects low contextual demand; however, cognitive resources are required for both. *For the model to provide the necessary specificity to influence attentional selection through contextual demands, the temporal trajectory for multiple contextual features of sensory-motor, attentional, emotional processes that persist with different time scales may be better characterized. Demand may also be better characterized as effort to reflect the bandwidth necessary for any mental computation. Furthermore, how such effort interacts with attention will also depend on the context of the perceiver. Some individuals who are trained to focus mindfully in the present moment will experience sensory stimuli very differently than someone who does not have such training or has a predisposition to anxiety.*

Limited Characterization of Task-(un)related Thinking

Task-negative or TUT, involves moments in which individuals try to attend to an external task (such as reading a book), but their minds disengage, and *involuntarily* wander to some other object of attention. The A2T framework suggests task relevance at any moment is determined by competition between internal working memory representations of the ongoing task and internal objects of attention that are salient by virtue of its emotional content. Task-relevance implies there is high demand on attention and the information is related to the ongoing task. The A2T model describes that attentional selection at any moment in time will be task-relevant with high task demand and more task-irrelevant at low task demand. One question that arises based on this observation is why is it common experience to have intrusive TUTs during tasks that rely heavily on working memory performance? This formulation of task relevance has considerable ecological validity in a variety of real-world experiences, such as attempting to plan dinner while driving through a construction zone (contextual demands) and worrying about an upcoming deadline (a negatively-valenced internal working memory representation). In such complex situations, it is both necessary to engage internal attention and to resist distraction. Some theories have posited that IDC may not require working memory, but rather working memory may constrain habitual or contextually-irrelevant thoughts and inhibit IDC in a reciprocal fashion (Heitz & Engle, 2007; McVay & Kane, 2010). Although attentional demand may persist across any task, fluctuations in performance of attention are commonly observed, especially when there are multiple competing demands by the continuous onslaught of salient affective or sensory stimuli or prospective plans and executive monitoring. Task-irrelevance or distraction arises when non-target objects of attention win the race for salience facilitating disengagement from the focus and initiating an attentional switch. Some existing similarity-choice theories of attention describe how conceptual distance between targets and distractors influence the probability of choice and speed of decision making (Logan, 2004). The A2T model does not focus on distractor models of attention, but instead describes spontaneous lapses in sustained attention driven by low task demand, personally-relevant chronic goals or unresolved problems. *Further characterization of the role for internal and external distraction, attentional salience, and engagement/disengagement processes would benefit specificity of the A2T model.*

There is some debate in the literature around such a dichotomization between task-negative and task-positive processes and the brain networks supporting them uniquely. The current view suggests that TUT do not activate cognitive processes. Recent studies have challenged this view demonstrating that both TUT and task-related thoughts engage goal-directed cognition, involve executive processes and are both supported by fluctuating interactions of brain networks identified as task-negative and task positive, depending on the nature of the task (Spreng, 2012). Although the spontaneous nature of self-generated thought suggests that it does not require cognitive resources, its priority-driven nature suggests that it is a resource intensive process (Levinson, Smallwood, & Davidson, 2012; Smallwood & Schooler, 2006).

Engagement with TUT is described by the A2T model to occur as a result of cognitive control failures; yet, there is much evidence for TUT as a successful deployment of cognitive resources. Unwanted forms of IDC may include thoughts of urgent personal concern, evaluation of current task performance, thoughts of recent or impending life events, daydreams with fantasies disconnected from reality, problem solving, creative idea incubation, or physiological states like hunger or temperature. Those IDCs of importance or urgent personal concern arguably compete and provide greater demand on resources than other TUTs (Chun et al., 2011; McVay & Kane, 2010). Mechanisms by which the mind tags particular thoughts with urgency may also be related to ruminative processes that tend to disrupt ongoing task-focused cognition.

The A2T model does not elaborate on this idea of urgency, but the paper mentions crisis-related thoughts and unresolved problems as contributing to attentional selection bias. Individuals with deficient control capabilities will more often succumb to TUTs than individuals with better control, who have less urgent personal concerns with which to contend, or those who have more training of attention and inhibition processes (e.g., through meditation practice)(McVay & Kane, 2010). *If indeed, "urgency" is a critical factor, further identifying what makes a thought urgent and how urgency may modulate current or prospective working memory processes should be further clarified. Mechanisms that support the determination of urgency or personal concern remain to be clear. Thus, a focus on level of*

urgency or personal concern can increase the specificity by which TUT influences IDC and such characterization would improve the model.

Definitions of urgency are critical because some criteria are needed to determine which of a panoply of automatic thoughts will be retrieved and presented to consciousness. We suspect the definition of such contextual factors is not trivial, and overlaps with the idea of involuntary memories. Involuntary memory retrieval processes are described to drive the contextual demand competing with goal-oriented, ongoing task related behaviors. Involuntary memories act as distractors that interfere with stable attentional engagement and arise with different frequency and with varied levels of conscious awareness. The tendency for involuntary memories to arise are also described as "automatic" and "reflexive" and characterize spontaneous thoughts, including content related to mind wandering and rumination.

Yet, the A2T model doesn't clarify how thoughts become automatic, the different contexts in which they arise, nor the critical influence of memory systems on automaticity of thought. Automatic and reflexive thoughts or behaviors arise without the influence of intention or present-moment decision making due to a history of repetition in learned behavior. These patterns are what drive top-down predictive processing of the world (Schultz & Dickinson, 2000). Another benefit of IDC is to strengthen associative learning, so anticipation of important outcomes is facilitated. Learning is maximized when there is an error in prediction due to a novel outcome (Schultz & Dickinson, 2000). Past experience creates weighted associations between patterns of thought and behavior modulated by valence, arousal, and reward history. Predictive processing drives top-down goal-directed cognitions and behaviors and is motivated by implicit associations established through skill-based motor memories or manifestations of repeated explicit associations that come from declarative forms of cognition and a specific attentional selection history.

The A2T describes a unique system for selection history; yet, reward history and priming are essential components of the mnemonic system that are not clearly identified. Automatized thinking does not necessarily imply bottom-up sensory processing, but is dependent on motor memories and associated circuitry to drive implicit associations. Retrieval of explicit episodic memories may elicit unanticipated patterns of thought and drive goal-directed behavior. The A2T model would improve its specificity by incorporating *sensory/perceptual, monitoring, and memory systems to better characterize the processes that drive both conscious and nonconscious attentional selection of information. The chronometry of predictively-coded top-down, goal-directed selection of information, stimulus-driven attentional capture, and the automatization of thoughts remains unclear. The modulatory potential of valence, arousal and salience of input would also aid in the further characterization of attentional selection.*

In addition to salience, personal relevance, and urgency, other IDC operations are likely to influence task-relevance in an adaptive fashion. For example, meta-cognitive (e.g., self-reflective, self-knowing) and prospective tasks (e.g., remember to take out the trash), may be necessary for planning and involve setting future-oriented access through a set of background cognitive process with great potential for later interference, but in a meaningful way. These types of thoughts and underlying procedural monitoring of thoughts may have a degree of spontaneity and goal-direction that fluctuate in time and duration in an intentional way (Koriat & Levy-Sadot, 1999). Other self-generated thoughts such as reflection on stimuli for relevance and evaluation of appropriate responses may arise spontaneously with some frequency during goal-directed thought informing the task at hand (e.g., intentional name or word finding or free memory recall). These IDC processes may involve *noetic* feelings that uniquely involve a gut or intuitive feeling that sought-after information, a solution to a problem, or a memory is known and correct, but cannot yet be articulated (Metcalf, 2000). This is commonly experienced as tip-of-the-tongue phenomena or in word-finding puzzles, such as Wordle.

Some work has been done to characterize motivation and intention-driven thoughts that interfere with ongoing demands, but are goal-directed to arise with a volitional frequency matching the relevance during waking consciousness (e.g., setting an internal timer for checking on something cooking on the stove) or even during sleep (e.g., setting an internal alarm to wake up at a certain time)(Aritake et al., 2004; Levinson et al., 2012). Yet, by definition, all TUT are believed to interfere with ongoing task demands through attentional control failures and are generated automatically by salient environmental or mental events (McVay & Kane, 2010). Thus, context may determine the relevance to the task and the demand on IDC as described by Smallwood and Andrews-Hanna (2013), drawing attention and working memory toward context-demanding thought or performance. Context may also contribute significantly to a sense of urgency that involve labeling a thought or group of thoughts for prospective processes and prioritization of effort. Indeed, some thoughts may be relevant for the completion of a task, but also be described as task-unrelated as it has been operationalized. Thoughts associated with anticipatory anxiety, prospective memory, or goal-directed thought may involve meta-cognitive, self-knowing processes that are associated with an ongoing task, but also may arise as distraction. Metacognitive judgments that are deliberately constrained and related to evaluating ongoing performance have been conceptualized in dissociable retrospective (e.g., confidence in past choices) or prospective (e.g., predicting success) processes (Fleming, Massoni, Gajdos, & Vergnaud, 2016). Thoughts relevant to ongoing task demands do not require any prospective process once the task is complete, suggesting relevance to the task and relation to future oriented behavior may contribute to the likelihood for future TUT. When we often refer to "having something on our mind," it refers to some thought or group of thoughts that are unresolved in some way and continue to take up meta-cognitive resources as a form of continued distraction. *There is little emphasis by the A2T model on adaptive forms of reflexive thinking that leads to insight, self-knowing, or creative idea incubation, for example. A more elaborate conceptualization including*

the potential interaction between personal relevance, contextual adaptation, noetic feeling states, and fundamental adaptive value would add to selectivity of the A2T model. Thus, characterizing task-related vs. task-unrelated IDC may need further clarity.

One final point of discussion on task-relatedness involves stochastic/random factors that are described as additional sources of influence on attentional selection and as critical to functioning of the proposed model. These random factors are described as random neural (system) noise and factors that influence task switching. Stochastic factors are described as low in probability, but essential to ensure a degree of *flexibility over time* such that the system does not become "stuck" indefinitely in a biased state due to internal feedback loops. The greater this stochastic element, the more variability in the model. In other words, these factors enable "random exits" from repetitive negative thinking—a potential circuit-breaking function that attenuates ruminative thoughts. This circuit-breaking function could potentially be a critical mechanism for depressogenic thinking and/or inhibition, yet there is little elaboration how a circuit breaking function may operate beyond random noise. *Unfortunately, specifying that random factors are the sole basis for distraction excludes habitual, influential factors and further characterization of stochastic factors for distraction-acknowledging their presence and describing how they manifest more clearly would be helpful.*

Confounding Contributions of Working Memory, Internal & External Attention

One of the main motivations for the creation of the A2T model is described to understand *how* internally-directed attention may dynamically interact with other cognitive processes to facilitate simulation and prediction of future internally-directed cognitive processes that are adaptive in nature. The model highlights the explicit delineation and dynamic interplay of internal versus external stimulus-driven attentional processes as well as interaction with other executive control, mnemonic cognitive processes and data processing systems. The A2T model uses the term, internal attention explicitly to refer to the specific processes of selection and modulation of the objects of attention. Working memory is functionally described to act as short-term storage for conscious detail and to guide behavior.

Early characterization of attentional systems do not emphasize the distinction between internal and external attention, but instead focus on sub-systems of processing related to orienting, detection, monitoring, and maintaining engagement and vigilance (Posner & Petersen, 1990; Raz & Buhle, 2006). Indeed, internal and externally-directed attentional processes are described to have similar computational and anatomical substrates with unique activity supporting the source of attention (external sensory input or long-term memory)(Chun et al., 2011; Lavie, Hirst, De Fockert, & Viding, 2004). More generally speaking, attentional selection is distinct from the computational resources for processing data input from external or internal sources. Computational processing of the objects of attention involves other cognitive functions, including vigilance, inhibition, switching, working memory, and encoding, retrieval, and memory consolidation processes. Regardless of whether attention is externally or internally-driven in any context, modality-specific objects of attention will always have many features of information that are processed by multiple cognitive and neural systems and as such, attentional processes expand and permeate into multiple higher order cognitive processes. Even at low level sensory processes, some aspect of non-conscious orienting, alerting and modulatory sensorimotor activity requires an attentional process. Indeed, attention may operate in a meta-cognitive capacity, but uniquely from the computational process to engage with each thought or perceptually-driven object from the external environment. Thus, attention is often described as a property of multiple, different perceptual and cognitive operations (Chun et al., 2011; Lavie et al., 2004). *The specificity by which IDC is modeled could be improved through more adequate incorporation of different sensory-motor, perceptual, and cognitive operations aside from working memory.*

Working memory involves temporary storage and manipulation of the information necessary for complex cognitive tasks as language comprehension, learning, and reasoning (Baddeley, 1992). This definition has evolved from the concept of a unitary short-term memory system to involve multiple integrated systems and networks supporting the simultaneous storage and processing of information (Baddeley, 1992). The contents of working memory are supplied from competition between a predictive load attributed to learned associations and bottom-up perceptual processes that have survived attentional bias and selective filtering for relevance. One assumption from the A2T model is that the contents of working memory (e.g., an image or thought) and the computation supporting holding that image or thought in a mental scratchpad is an internally-generated process that subserves IDC. This assumption also appears to support the idea that the image or thought can arise spontaneously without any trigger stimulus from the external environment. The assumption is that information is stimulus-independent if it arises from internal sources; yet, there always remains interdependent causal relations between the external environment and internal mental activity with any spontaneous mental activity.

Whether an image or thought arises from content objects previously in working memory or whether the content is derived from long-term memory, attention is stable and the object of attention and its relations to other objects of attention become labile for evaluation and reconsolidation processes—hence a working memory. *No matter the source of input, a comprehensive model of IDC would improve its specificity by addressing competing parallel processes rather than delineating a distinct system for internal and external attention. Furthermore, the focus on decoding working memory processes may provide a unitary system framework for better understanding internally-directed vs. externally-directed attention. Further elaboration of contents vs. processes supporting*

working memory will benefit the development of a model that fully characterizes task relevant or task-irrelevant cognition.

Smallwood (2013) proposes how IDC fluctuates between thoughts coupled and uncoupled from perceptual input (Smallwood, 2013). This process-oriented framework attempts to resolve the difference between the occurrence of IDC and the processes that ensue once the experience has been initiated and perseverate without external sensory input (i.e., spontaneous, stimulus-independent). Even with absence of external sensory input, internally generated cues (e.g., imagery, thought, feeling states) may act as an urgent or "imperative stimulus" that is causally-linked to the generation of a particular thought in the subsequent moment that may be familiar (e.g., repetitive) or novel (e.g., insight). Whether or not stimulus-independence sufficiently characterizes IDC, there is a natural "ebb and flow" between internal and external states that characterizes mind-wandering. No model has yet to disentangle what is truly stimulus-independent or whether sub-vocal thoughts or focused interoceptive concentration without thought and associated internally-directed processes are truly distinct from cognition associated with vocal, expressive language or task-focused attention and performance. As mentioned earlier, the A2T model appears to account for such internal-external fluctuations through a stochastic variable; yet, the timing and nature of fluctuations between stimulus-independent and environmentally-constrained cognition are not adequately addressed. It is likely that IDC may have varying degrees of stimulus-dependence where the causal links for attentional orienting and thoughts are first-order and easy to identify or separated by second or third orders with more obscure causal connections (Smallwood, 2013). There are connectionist models of cognition that base these dynamics on previous mnemonic associations and the temporal contiguity of the concepts that are associated together (e.g., the sight and smell of a rose along with a memory of a rose garden in your neighborhood). Selection history then provides an associative connectionist network of IDC with various weighted associations and modulatory influences from arousal, emotion, and reward circuitry. Weighting of certain causal pathways provide particular mental habits or patterns for priming future thinking that are affect and content-congruent.

TUTs may be responsible for perpetuating further TUTs. The A2T model goes into some detail around the repetitive nature of negatively-oriented thoughts and how the negative thought perpetuates biased selection by increasing negative affect. Attentional biases likely serve personally-relevant functional significance to determine the content of working memory at each successive moment. This network of concepts, thoughts and attentional biases create the tapestry of top-down inferential processes that provide individually distinct responses to stimuli and modulation of a computational system of resource availability. Motivational relevance of sensory input enhances cognitive effort, demands attention-specific resources, and inhibits the arising of TUT or unrelated sensory stimuli (i.e., distraction). As stimuli become repetitive or as individuals systematically train attention, more resources for a broader attentional spotlight become available. This kind of efficiency has been observed in meditators during perceptual discrimination and attentional blink tasks (Brown, Forte, & Dysart, 1984; Carter et al., 2005; Kerr et al., 2008; MacLean et al., 2010; Slagter et al., 2007). The more an individual can automatize future-oriented perception, thought, and behavior, the more efficient the system will be. There is now clear evidence that automatized reflexive and habitual cognition can recruit skill-based mnemonic brain circuitry (e.g., dorsal striatum, cerebellum, thalamus) and increase bandwidth and cognitive resources for focused attention (Ashby, Turner, & Horvitz, 2010).

The A2T model emphasizes a dynamic temporal trajectory of discrete thoughts moment to moment in order to characterize and make predictions of attentional selection. The representation of temporal sequences for thematically similar thought content as well as frequent shifts in thought content across time is a strength of the model. For example, individual differences in threat-related bias have been described in terms of threat value and salience with little interference by conscious awareness within the first 250 ms following sensory experience. A resource allocation mechanism that determines the extent of attentional engagement has been modeled in the next 250 ms when inhibitory control processes are deployed to manage distractions and further elaboration and appraisal processes follow (Gupta, Kujawa, & Vago, 2019). Particular traits like anxiety or mindfulness are also likely to modulate this chronometry of IDC and associated attentional engagement processes. *Further development modeling the temporal chronometry of biased perception and attentional processes will be helpful so that early and late processing stages of orienting, monitoring, alerting, engagement, and evaluation are modeled in a more comprehensive fashion. Incorporating the chronometry of attentional bias and changes in resource allocation will also be informative for modeling attentional selection and IDC* (Gupta et al., 2019).

There is some suggestion that IDC involves executive functioning, including working memory, response selection, and inhibitory control processes. Yet, these processes require cognitive resources or effort, making them incompatible with the A2T model which describes cognitive effort as an inhibitory mechanism for IDC. Much of the description of IDC by the A2T model emphasizes spontaneous thought processes involving rumination, mind wandering that are automatic, negatively valenced, reflexive or automatic, and suggestive of a lack of inhibitory control and awareness. It is arguable that some level of awareness is necessary for rumination and mind wandering, further suggesting inhibitory control deficits are prevalent in those states. Much research in the literature on inhibitory control mechanisms have described the role of inhibitory tone across systems that may operate as a gain function to potentially suppress or enhance the intensity internal and external representations that are contextually relevant. For example, there is a large body of research that has observed cortical alpha-frequency dynamics that support filtering or blocking of external sensory input to primary sensory cortex and supporting

optimization of attentional mechanisms stabilizing on other sources like internal somatic states.

Inadequate Neurobiological Specificity

There has been some progress in identifying the neurobiological specificity of IDC; yet, there remain gaps in the literature and specifically in the proposed A2T model. Internally-directed attention has been described to involve perceptual decoupling; yet, the mechanisms remain unclear. Although multiple neural systems are involved in supporting IDC and can moderate attentional selection, here we briefly describe two putative neural mechanisms as directions for future research to advance the neurobiological specificity of the model.

Neural systems that would be helpful to incorporate in future models of IDC include both exteroceptive and interoceptive sensory-motor networks, supporting cue-driven activation of thought and emotion. Depending upon the extent of surprise and novelty experienced by the stimuli-generated IDC, different sensory-motor patterns of activation are likely to be supported (Allen et al., 2016; Joos, Gilles, Van de Heyning, De Ridder, & Vanneste, 2014). An important caveat is that sensory events may be salient even if they possess no relevance to task demands, such as a car honking or the smell of a cigarette smoker walking in front of you. Internal events possess comparable attentional gravity, such as a sudden pain or a wave of fatigue intruding into consciousness. Memory representations may likewise operate in a pseudo-sensory fashion, cued by contextual associations as reconstructed sensory or semantic events. Attention may then operate on these representations, updating cued associations in light of perceptual and semantic features recalled. Other well-rehearsed goals or concerns that are distinct from both the current context demands or dominant features of working memory are likely competing for attention. Academics are doubtlessly familiar with the creeping awareness of "deep work" like analysis and writing that is left in the background amidst the many urgent (but arguably less important) exigencies of daily life. Even within a framing of describing rumination, notable computational models have already been proposed, (e.g., van Vugt et al., 2018). This model used empirical data to inform the frequency of distracting rumination, but unlike the A2T model, it used a broader memory habit framework to account for intrusive rumination tendencies. To be sure, multiple competing models are needed to compare mechanistic accounts, but an expanded A2T proposal might mention competition between memory availability and emotion for example as two converging influences on attention selection. Positive emotions are also likely attractors of attention, ranging from the wholesome (cuing in on a friend's smiling face) to the problematic (searching for one's drug of choice in substance use disorders). While many forms of life are biased toward attention to negative over positive emotions, positive emotions still possess an indisputable gravity over more neutral representations. There is longstanding evidence that the types of emotions that are most likely to engage in the brain's salience network (most prominently the amygdala for cases of initial detection and attentional orientation), are themselves tuned by contextual demands such as approach (positive) vs. avoidance (negative) goals (Cunningham & Brosch, 2012). It is understandable to start with a simpler model that focuses on being distracted by negative features of experience as an example case. It is important however to acknowledge that involuntary rumination or catastrophizing during a challenging task is very narrow sector of the varieties of human experience, and a broadly defined A2T model must be able to account for a greater variety of attentional selection events, as described above. Interactions among large-scale brain networks have been posited to underlie externally- and internally-directed cognitive processes and the flexible switching between these attentional states (Vago & Zeidan, 2016). Mind Wandering is one of the main internally-directed processes and has been hypothesized to involve the default mode network. Yet, the dynamic interaction between resting state networks such as the fronto-parietal, salience and DMN networks are likely critical. Arousal and modulatory dynamics of the basal forebrain and other aminergic, cholinergic, and glutamatergic activity has yet to be appropriately characterized as part of our understanding of IDC. Predictive top-down vs. interoceptive afferent bottom-up processing (match-mismatch)/active vs. perceptual inferences are all aspects of embodied cognition that would also add value to the development of the A2T model.

A second candidate neural mechanism putatively associated with the core features of the A2T model is the interacting neurotransmitter systems that mediate externally- and internally- directed attention. The nuclei of the locus coeruleus (LC) and basal forebrain (BF) are implicated in cortically mediated attentional processes through extensive noradrenergic (LC) and cholinergic (BF) projections. These systems have been studied extensively in relation to externalized attention (Corbetta, Patel, & Shulman, 2008; Corbetta & Shulman, 2002; Mather & Harley, 2016; Schmitz & Duncan, 2018). Their involvement in mediating internally directed attentional processes is less well understood. LC has been associated with the emergence of off-task, internally generated thoughts and mind-wandering (Mittner, Hawkins, Boekel, & Forstmann, 2016). Tonic activation of LC is thought to broaden attentional focus, reducing external, task-related processing in favor of more off-task, internally directed thoughts mediated by default network brain regions (Smallwood, Brown, Baird, & Schooler, 2012). In contrast, activation of the cholinergic system is associated with deactivation of the default network, and engagement of frontal-parietal systems implicated in externally directed attention and cognitive control processes (Bentley, Driver, & Dolan, 2011). Further, the LC and BF are both integrated with the salience network, which actively suppresses the default network during externally directed attention (Uddin, 2015). Taken together, these early findings suggest that the dynamic interplay of these two sub-cortical regions, and their cortical projection systems, is necessary to flexibly shift between internally and externally directed attentional processes.

Beyond engagement, cessation, and flexibly switching between internally and externally directed attention, these two neurotransmitter systems may also shape the contents of perceptually decoupled, internally generated thought. LC and BF have direct, monosynaptic connections to entorhinal cortex (Theofilas, Dunlop, Heinsen, & Grinberg, 2015; Vago, Bevan, & Kesner, 2007), a critical pathway to the hippocampus, associated with variability in mnemonic thought contents(Kalina Christoff et al., 2016). Further, phasic signaling of the LC biases attention to task-relevant information (Aston-Jones & Cohen, 2005) establishing a more exploitative cognitive mode(R Nathan Spreng & Turner, 2021). An LC-modulated exploitative cognitive mode, anchored in memory and prior knowledge structures and selectively enhanced by cholinergic inputs from the BF (Sarter, Hasselmo, Bruno, & Givens, 2005), may sustain internally directed attentional focus. Time-related variations in attentional selection, encoding and memory retrieval processes correlate with amplitude changes of theta and alpha cortical oscillations, and modulation of those oscillations through cholinergic pathways (Clayton, Yeung, & Kadosh, 2015; Hasselmo, 2006). Other research has demonstrated specific inhibitory and selection mechanisms of attention are correlated with activation of alpha rhythmic activity in brain networks associated with working memory (Clayton et al., 2015; Klimesch, 2012). While speculative, the dynamic interplay between these subcortically mediated systems may provide candidate neural mechanisms of IDC, promoting the transition of off-task thoughts to more intentional, goal-directed mind-wandering associated with planning, incubation of creative ideas, and meditation practice (Mittner et al., 2016; Vago & Zeidan, 2016). Neurophysiological dynamics supporting cholinergic and aminergic modulation of memory encoding, retention, and retrieval processes also provide a temporal window into the role of attention and conscious interpretation of contextual cues that drive attentional selection and enable controlled access to relevant semantic information.

In the A2T model, there is an emphasis on momentary changes in cognitive control that influence emotional reactivity, and further dyscontrol, disinhibition and/or negatively-oriented bias of attention.

However, providing the factors of (external) context and (negative) emotion as an exhaustive set of distractors omits other important and indeed ubiquitous moderating influences on working memory representation. The study of attention selection is a broad area of scholarship; in neural systems theory, this process is perhaps best captured through accounts of the brain's Salience Network (Seeley et al., 2007). While the Salience Network does respond to external contextual demands (Fecteau & Munoz, 2006), activation of this network is not limited to goal-directed cognition. The A2T model does not account for other sources of bias, such as previously conditioned associations or positively-valenced representations. A considerable empirical literature supports the power of affectively-diverse representations to guide attention for constructive self-reflection, simulations of future outcomes, goal setting, and creative idea incubation (Smallwood et al., 2011). Additionally, the formulation of criticality suggests a functional neurobiology of consciousness may dynamically operate on the border of a phase transition between order and chaos (O'Byrne & Jerbi, 2022). *Thus, several other important neurobiological factors that account for the broader experience of context and modulatory power would improve the sufficiency for a comprehensive account of attention to thoughts.*

Expanded Consideration of Internally Directed Cognition in the A2T Model

Does the A2T model sufficiently IDC and the intended (mal)adaptive phenomenology? Many points for improving specificity and sufficiency have been delineated above and help improve our understanding of selective attention processes relevant for common forms of bias and rumination, as well as adaptive aspects of IDC, such as goal setting, creative idea incubation, and memory consolidation. Some of the suggestions may help clarify differential processes supporting internal and external attention, working memory, memory encoding, retrieval, cognitive capacity, or mental habits that contribute to the reification of self-identity. Conceptual clarity may reveal there is a fundamental homunculus argument in which global workspace theory must account for contextually-driven demands on task relevance. Failures to account for attention schema or connectionist models of memory may also provide opportunities for improved computational modeling. Further development of a model for IDC can incorporate a plausible neuroanatomic and neuromodulatory framework. We propose a non-computational model of internally directed attention based on salience and default network interactivity, incorporating noradrenergic and cholinergic inputs. More extensive development of a theoretical model will require integration of empirical data with existing models and clarification of IDC concepts more generally.

Canvasing the realm of human experience more broadly, the A2T framework describes only a limited sector of human experience. Negative valence thus has very shallow influence on the specificity of attentional selection. Several considerations for expanding the constructs used toward a more comprehensive A2T framework and model of IDC are suggested. Although degree of negative affect is a critical factor in influencing attentional selection, models of IDC would benefit greatly from incorporating the broader influence of subjective valence (positive and negative), contextual factors associated with memory, craving/reward, and several antecedent and post-selection processes that influence future attentional selection and selection history. Subjective valence is likely to influence attentional selection in unique ways depending on intensity and context. Further characterization of attentional selection beyond negatively valenced content or degree of emotional intensity using additional constraints may likely increase specificity by which emotional experience influences attentional selection of thought content. Further modeling of IDC may elaborate on the dynamics by which degree of

negative affect may influence negativity biases that may form at the perceptual level and increase the likelihood new information is interpreted as negative, especially when information is ambiguous.

ORCID

David R. Vago http://orcid.org/0000-0002-6269-9027
Norman Farb http://orcid.org/0000-0002-8407-2938
R. Nathan Spreng http://orcid.org/0000-0003-1530-8916

References

Allen, M., Fardo, F., Dietz, M. J., Hillebrandt, H., Friston, K. J., Rees, G., & Roepstorff, A. (2016). Anterior insula coordinates hierarchical processing of tactile mismatch responses. *NeuroImage*, *127*, 34–43.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*, 29–52. doi:10.1111/nyas.12360

Aritake, S., Uchiyama, M., Tagaya, H., Suzuki, H., Kuriyama, K., Ozaki, A., … Okubo, Y. (2004). Time estimation during nocturnal sleep in human subjects. *Neuroscience Research*, *49*(4), 387–393.

Ashby, F. G., Turner, B. O., & Horvitz, J. C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. *Trends in Cognitive Sciences*, *14*(5), 208–215. doi:10.1016/j.tics.2010.02.001

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450.

Axelrod, V., Rees, G., & Bar, M. (2017). The default network and the combination of cognitive processes that mediate self-generated thought. *Nature Human Behaviour*, *1*(12), 896–910.

Baddeley, A. (1992). Working memory. *Science (New York, N.Y.)*, *255*(5044), 556–559. doi:10.1126/science.1736359

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, *23*(10), 1117–1122. doi:10.1177/0956797612446024

Bentley, P., Driver, J., & Dolan, R. J. (2011). Cholinergic modulation of cognition: Insights from human pharmacological functional neuroimaging. *Progress in Neurobiology*, *94*(4), 360–388.

Brown, D., Forte, M., & Dysart, M. (1984). Differences in visual sensitivity among mindfulness meditators and non-meditators. *Perceptual and Motor Skills*, *58*(3), 727–733.

Carter, O. L., Presti, D. E., Callistemon, C., Ungerer, Y., Liu, G. B., & Pettigrew, J. D. (2005). Meditation alters perceptual rivalry in Tibetan Buddhist monks. *Current Biology : CB*, *15*(11), R412–413.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8719–8724. doi:10.1073/pnas.0900234106

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews. Neuroscience*, *17*(11), 718–731. doi:10.1038/nrn.2016.113

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*(1), 73–101.

Clayton, M. S., Yeung, N., & Kadosh, R. C. (2015). The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences*, *19*(4), 188–195. doi:10.1016/j.tics.2015.02.004

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*(3), 306–324.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215.

Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science*, *21*(1), 54–59. doi:10.1177/0963721411430832

Farb, N. A., Segal, Z. V., Mayberg, H., Bean, J., McKeon, D., Fatima, Z., & Anderson, A. K. (2007). Attending to the present: Mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience*, *2*(4), 313–322. doi:10.1093/scan/nsm030

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390.

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, *2016*(1), niw018. doi:10.1093/nc/niw018

Fox, K. C., Andrews-Hanna, J. R., & Christoff, K. (2016). The neurobiology of self-generated thought from cells to systems: Integrating evidence from lesion studies, human intracranial electrophysiology, neurochemistry, and neuroendocrinology. *Neuroscience*, *335*, 134–150. doi:10.1016/j.neuroscience.2016.08.020

Fox, K. C. R., & Christoff, K. (2018). *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming*. Oxford: Oxford University Press. https://books.google.ca/books?id=0mdaDwAAQBAJ.

Gupta, R. S., Kujawa, A., & Vago, D. R. (2019). The neural chronometry of threat-related attentional bias: Event-related potential (ERP) evidence for early and late stages of selective attentional processing. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *146*, 20–42. doi:10.1016/j.ijpsycho.2019.08.006

Hasselmo, M. E. (2006). The role of acetylcholine in learning and memory. *Current Opinion in Neurobiology*, *16*(6), 710–715. doi:10.1016/j.conb.2006.09.002

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*(2), 217–240. doi:10.1037/0096-3445.136.2.217

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception & Psychophysics*, *81*(7), 2288–2303. doi:10.3758/s13414-019-01846-w

Jamieson, J. P., Crum, A. J., Goyer, J. P., Marotta, M. E., & Akinola, M. (2018). Optimizing stress responses with reappraisal and mindset interventions: An integrated model. *Anxiety, Stress, and Coping*, *31*(3), 245–261. doi:10.1080/10615806.2018.1442615

Joos, K., Gilles, A., Van de Heyning, P., De Ridder, D., & Vanneste, S. (2014). From sensation to percept: The neural signature of auditory event-related potentials. *Neuroscience and Biobehavioral Reviews*, *42*, 148–156.

Kerr, C. E., Shaw, J. R., Wasserman, R. H., Chen, V. W., Kanojia, A., Bayer, T., & Kelley, J. M. (2008). Tactile acuity in experienced Tai Chi practitioners: Evidence for use dependent plasticity as an effect of sensory-attentional training. *Experimental Brain Research*, *188*(2), 317–322. doi:10.1007/s00221-008-1409-6

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science (New York, N.Y.)*, *330*(6006), 932. doi:10.1126/science.1192439

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. doi:10.1016/j.tics.2012.10.007

Klinger, E., Marchetti, I., & Koster, E. H. (2018). Spontaneous thought and goal pursuit: From functions such as planning to dysfunctions such as rumination. In *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming* (p. 215). Oxford: Oxford University Press.

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chalken & Y. Trope (Eds.), *Dual-*

processs theories in social psychology (pp. 483–502). New York: Guilford Press.
Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. doi:10.1037/0096-3445.133.3.339
Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The persistence of thought: Evidence for a role of working memory in the maintenance of task-unrelated thinking. *Psychological Science*, *23*(4), 375–380.
Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology*, *55*(1), 207–234. doi:10.1146/annurev.psych.55.090902.141415
MacLean, K. A., Ferrer, E., Aichele, S. R., Bridwell, D. A., Zanesco, A. P., Jacobs, T. L., … Saron, C. D. (2010). Intensive meditation training improves perceptual discrimination and sustained attention. *Psychological Science*, *21*(6), 829–839. doi:10.1177/0956797610371339
Mather, M., & Harley, C. W. (2016). The locus coeruleus: Essential for maintaining cognitive function and the aging brain. *Trends in Cognitive Sciences*, *20*(3), 214–226.
McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin*, *136*(2), 188–189.
Metcalf, J. (2000). Feelings and judgments of knowing: Is there a special noetic state? *Consciousness and Cognition*, *9*, 178–186.
Mittner, M., Hawkins, G. E., Boekel, W., & Forstmann, B. U. (2016). A neural model of mind wandering. *Trends in Cognitive Sciences*, *20*(8), 570–578. doi:10.1016/j.tics.2016.06.004
Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *3*(5), 400–424.
O'Byrne, J., & Jerbi, K. (2022). How critical is brain criticality? *Trends in Neurosciences*, *45*(11), 820–837. doi:10.1016/j.tins.2022.08.007
Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25–42. doi:10.1146/annurev.ne.13.030190.000325
Raz, A., & Buhle, J. (2006). Typologies of attentional networks. *Nature Reviews. Neuroscience*, *7*(5), 367–379.
Sarter, M., Hasselmo, M. E., Bruno, J. P., & Givens, B. (2005). Unraveling the attentional functions of cortical cholinergic inputs: Interactions between signal-driven and cognitive modulation of signal detection. *Brain Research. Brain Research Reviews*, *48*(1), 98–111.
Schmitz, T. W., & Duncan, J. (2018). Normalization and the cholinergic microcircuit: A unified basis for attention. *Trends in Cognitive Sciences*, *22*(5), 422–437.
Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Psychology*, *23*, 473–500.
Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., … Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356. doi:10.1523/JNEUROSCI.5587-06.2007
Slagter, H. A., Lutz, A., Greischar, L. L., Francis, A. D., Nieuwenhuis, S., Davis, J. M., & Davidson, R. J. (2007). Mental training affects distribution of limited brain resources. *PLoS Biology*, *5*(6), e138. doi:10.1371/journal.pbio.0050138.
Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process-occurrence framework for self-generated mental activity. *Psychological Bulletin*, *139*(3), 519–535. doi:10.1037/a0030010
Smallwood, J., & Andrews-Hanna, J. (2013). Not all minds that wander are lost: The importance of a balanced perspective on the mind-wandering state. *Frontiers in Psychology*, *4*, 441.
Smallwood, J., Brown, K., Baird, B., & Schooler, J. W. (2012). Cooperation between the default mode network and the frontal-parietal network in the production of an internal train of thought. *Brain Research*, *1428*, 60–70. doi:10.1016/j.brainres.2011.03.072
Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958.
Smallwood, J., Schooler, J. W., Turk, D. J., Cunningham, S. J., Burns, P., & Macrae, C. N. (2011). Self-reflection and the temporal focus of the wandering mind. *Consciousness and Cognition*, *20*(4), 1120–1126. doi:10.1016/j.concog.2010.12.017
Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, *67*(1), 415–437.
Spreng, N. R. (2012). The fallacy of a "task-negative" network. *Frontiers in Psychology*, *3*, 145.
Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489–510.
Spreng, R. N., & Turner, G. R. (2021). From exploration to exploitation: A shifting mental mode in late life development. *Trends in Cognitive Sciences*, *25*(12), 1058–1071.
Takano, K., Sakamoto, S., & Tanno, Y. (2011). Ruminative and reflective forms of self-focus: Their relationships with interpersonal skills and emotional reactivity under interpersonal stress. *Personality and Individual Differences*, *51*(4), 515–520. doi:10.1016/j.paid.2011.05.010
Theofilas, P., Dunlop, S., Heinsen, H., & Grinberg, L. T. (2015). Turning on the light within: Subcortical nuclei of the isodentritic core and their role in Alzheimer's disease pathogenesis. *Journal of Alzheimer's Disease : JAD*, *46*(1), 17–34.
Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews. Neuroscience*, *16*(1), 55–61. doi:10.1038/nrn3857
Vago, D. R., Bevan, A., & Kesner, R. P. (2007). The role of the direct Perforant path input to the CA1 subregion of the dorsal hippocampus in memory retention and retrieval. *Hippocampus*, *17*(10), 977–987. doi:10.1002/hipo.20329
Vago, D. R., Gusnard, D. A., & Silbersweig, D. A. (2024). Neuropsychiatry of self. In R. Boland (Ed.), *Kaplan & Sadock's concise textbook of clinical psychiatry*. Philadelphia, PA: Lippincott & Williams & Wilkins.
Vago, D. R., & Zeidan, F. (2016). The brain on silent: Mind wandering, mindful awareness, and states of mental tranquility. *Annals of the New York Academy of Sciences*, *1373*(1), 96–113. doi:10.1111/nyas.13171
van Vugt, M. K., van der Velde, M., & ESM-MERGE Investigators (2018). How does rumination impact cognition? A first mechanistic model. *Topics in Cognitive Science*, *10*(1), 175–191. doi:10.1111/tops.12318
Verhaeghen, P., Joorman, J., & Khan, R. (2005). Why we sing the blues: The relation between self-reflective rumination, mood, and creativity. *Emotion (Washington, D.C.)*, *5*(2), 226–232. doi:10.1037/1528-3542.5.2.226
Zahavi, D. (2003). Phenomenology of self. *The Self in Neuroscience and Psychiatry*, *56*, C75.