Perturbance: Unifying research on emotion, intrusive mentation and other psychological phenomena with AI

Luc P. Beaudoin[1]**, Sylwia Hyniewska**[2] **and Eva Hudlicka**[3]

Abstract. Intrusive mentation, rumination, obsession, and worry, referred to by Watkins [1] as "repetitive thought" (RT), are of great interest to psychology. This is partly because every typical adult is subject to "RT". A critical feature of "RT" is of transdiagnostic significance—for example obsessive compulsive disorder, insomnia and addictions involve unconstructive “RT”. We argue that “RT” cannot be understood in isolation but must rather be considered within models of whole minds. Researchers must adopt the designer stance in the tradition of Artificial Intelligence augmented by systematic conceptual analysis [2]. This means developing, exploring and implementing cognitive-affective architectures. Empirical research on "RT" needs to be driven by such theories, and theorizing about “RT” needs to consider such data. We draw attention to H-CogAff theory of mind (motive processing, emotion, etc.) and a class of emotions it posits called perturbance (or tertiary emotions) [3,4], as a foundation for the research programme we advocate. Briefly, a perturbance is a mental state in which motivators tend to disrupt executive processes. We argue that grief, limerence (the attraction phase of romantic love) and a host of other psychological phenomena involving "RT" should be conceptualized in terms of perturbance and related design-based constructs. We call for new taxonomies of "RT" in terms of information processing architectures such as H-CogAff. We claim general theories of emotion also need to recognize perturbance and other architecture-based aspects of emotion. Meanwhile “cognitive” architectures need to consider requirements of autonomous agency, leading to cognitive-*affective* architectures.

In the evenings she peeped out at him from the bookcase, from the fireplace, from the corner — he heard her breathing, the caressing rustle of her dress. In the street he watched the women, looking for someone like her. (“The Lady with the Dog”, Anton Chekhov.)

1 INTRODUCTION

Over 35 years ago, Aaron Sloman and Monica Croucher launched a research programme based on an important , subtle insight that had first been suggested by Herbert Simon [61]: Humans have emotions for the same reason that future robots will, as a *consequence* of interacting information processing mechanisms that address the requirements of autonomous agency [5,6]. This theory was grounded in Artificial Intelligence (AI) and conceptual analysis. From 1990 to 2005 the ensuing Cognition and Affect (CogAff) project (mainly at the University of Birmingham, England) actively pursued this insight, by exploring, implementing, assessing and refining requirements, and software tools, and developing cognitive-affective architectures capable of modelling the hypothesized affective processes[7].

Sloman ultimately proposed three major types of emotion: primary, secondary and tertiary [3]. Tertiary emotion, the focus of [5,6] where it is simply called ‘emotion’, is also the focus of this paper; for reasons explained below, we refer to it as *perturbance* [4]. In a nutshell, in order for a motive to disturb deliberative processes, such as problem solving, it must implicitly or explicitly be assigned a sufficiently high insistence level. A perturbance is a state in which an insistent motivator tends to distract *or alter* deliberative processes in a manner that is difficult for reflective processes to suppress or control. These terms are briefly described below, and more extensively in various CogAff project publications we cite. In short, the concept of perturbance provides a parsimonious, design-based way of understanding the obsessive aspects of emotion-like states, wherein the agent experiences a certain loss of control of attention and hence of management processes.

In this paper, we argue that perturbance is a major feature of the human mind that has been ignored by psychologists but deserves considerable attention. This concept has the potential to unify several areas of study, including fundamental processes such as attention, emotion and emotion regulation, cognitive phenomena such as intrusive thought, and psychopathological conditions such as rumination, obsessive worrying and addictions. Like any theoretical concept, the concept of perturbance does not stand alone. It is meaningful, promising and useful because of the theoretical framework within which it is embedded: a) the CogAff architecture *schema*, and b) H-CogAff, a particular architecture based on CogAff which is aimed specifically at understanding humans [3].

Whereas Sloman made significant attempts to disseminate the design-based approach and H-CogAff to emotion and AI researchers, the impact on the psychology literature so far has been minimal, due to various factors some of which we will allude to here. Meanwhile, affective computing (AC), a discipline of computer science that focuses on emotion, including emotion modelling, is gaining momentum. However, AC currently tends to pursue narrow problems relevant to practical applications focusing on primary emotions (e.g., machine perception of primary emotions). In AC, there is almost no research on automatically detecting perturbance, let alone attempts to produce systems that can experience and monitor perturbance. This is the case despite the fact that Sloman’s work, including the concept of perturbance, was described in Rosalind Picard influential *Affective Computing* [65]. Sloman, who was

[1] Faculty of Education, Simon Fraser Univ. Email: lpb@sfu.ca.
[2] Dept. of Psychology, Univ. of Bath. Email: Sylwia.Hyniewska@gmail.com.
3. Psychometrix Associates, Inc. Email: hudlicka@ieee.org.

one of the first AI researchers to systematically emphasize computational *architectures*, did foresee that AC would be a long road [66]. Still, AI's highly visible progress, and its work on architectures, bode well for AC. We believe that history will prove Sloman's theory of perturbance is a "sleeping beauty". According to [8] these "beauties" tend to 'awaken' when they are discovered by a new community of researchers.

This paper is meant to promote consideration of H-CogAff by indicating its relevance to many phenomena and research communities, while focusing on one of its original concepts, perturbance. However, we only have space for a cursory overview of the theory itself. For more information about it, see [2-7,9-10] and other papers cited below.

2 WHY HUMANS HAVE PERTURBANT EMOTIONS AND HIGHLY AUTONOMOUS ROBOTS WILL TOO

Sloman & Croucher [5,6] claimed emotions will emerge as side-effects in minds designed to meet the requirements of autonomous agency. These challenges include dealing with multiple endogenous sources of motivation with limited physical and processing resources in a rapidly changing, unpredictable, and only partially controllable environment. Autonomous agents require relatively simple mechanisms to generate and activate goals. For various *a priori* reasons their deliberative mechanisms have limited parallelism (see [4] ch. 4 and [61]) .

Not every activated goal can be considered simultaneously by deliberative processes. There must be comparatively simple mechanisms to decide whether the deliberative layer of the architecture may be interrupted (or otherwise influenced) by a given goal. These include insistence assignment, which heuristically reflects the importance and urgency of a goal, and interrupt filtering. For example, if a hungry autonomous agent detects a rare opportunity to consume a source of energy, a new goal to approach the source may be triggered. However, in order for this goal to even be considered, it needs to be sufficiently *insistent* to penetrate the attention filter and interrupt current executive processing and behaviour. If the agent is under attack, its executive processes might not even notice its goal to approach the source because the filter threshold will have been raised higher than the insistence level of the goal to approach. As any good software designer knows, designing software involves trade-offs. It's impossible to design perfect insistence and filtering rules. Sometimes, the robot will tend to be distracted *by its own insistent* goals that it keeps rejecting (e.g., to approach an appealing agent the pursuit of whom would violate its norms or other goals—conflicted robot love.) Thus, not all emotion-like states need be built into a robot; perturbant emotions will emerge.

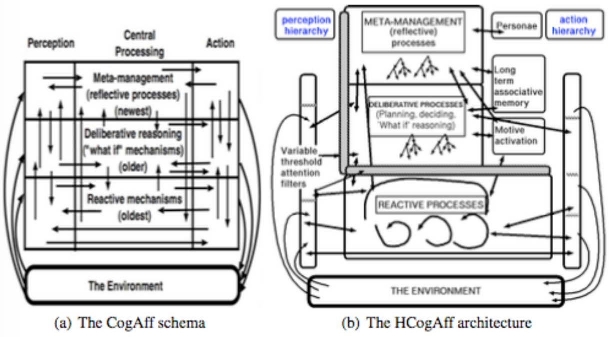
Sloman, Beaudoin and their colleagues on H-CogAff project continued to be challenged by psychologists who insisted that the states they were describing were not really emotions. Meanwhile, psychologists still do not agree on the meaning of the word "emotion" [11-13], a highly polymorphous concept in ordinary language. In order to avoid pointless turf wars over a label and to stimulate progressive research, Beaudoin [4] coined the term "perturbance" for Sloman's original technical concept of emotion [5]. Since then, one has been able to say there are *perturbant* emotions (or perturbant states), while allowing researchers to stipulate other types of emotion. We prefer the term 'perturbance' to 'tertiary emotion' (introduced later by Sloman [3]) because the former denotes a more general concept—e.g., it can be interpreted in terms of architectures with more than three layers.

Perturbance is of considerable adaptive significance because it is an affection of the human brain's *executive* processes, which govern the agent.

Alas, internal attentional disturbance still does not figure prominently in general theories of human emotion (e.g., [13,14]). Ironically, it is in a biological theory of emotion that such disturbance is highlighted, in what Panksepp & Biven [15] also call tertiary emotions. Unfortunately, the architecture-based *concept* of perturbance is still not used widely outside H-CogAff. Yet the loss of control of attention of many emotional episodes needs to be accounted for in such terms. We believe that the concept of perturbance and its label still need to be disseminated. It is our hope that this paper will help the idea gain acceptance and treat this alexithymia in the literature on affect.

3 H-CogAff: AN AUTONOMOUS AGENT ARCHITECTURE

The concept of perturbance is part of a design-based research programme that proposes a class of mental architectures (CogAff) whose particular instance, H-CogAff, is the backdrop of this paper [7]. H-CogAff is a response to human autonomous agency requirements emanating from that programme. They were alluded to above, and elaborated in [4]. A sketch of H-CogAff is presented next to the more generic CogAff schema in Figure 1.


Perception
Central Processing
Action
Meta-management (reflective processes) (newest)
Deliberative reasoning ("what if" mechanisms) (older)
Reactive mechanisms (oldest)
The Environment
perception hierarchy
META-MANAGEMENT (reflective) processes
Personae
action hierarchy
Long term associative memory
DELIBERATIVE PROCESSES (Planning, deciding, 'What if' reasoning)
Motive activation
Variable threshold attention filters
REACTIVE PROCESSES
THE ENVIRONMENT

(a) The CogAff schema (b) The HCogAff architecture

Figure 1. CogAff schema and H-CogAff architecture diagrams from [3]

This highly interconnected architecture assumes mechanisms for perceiving and affecting the environment, generating alarms, and creating and activating goals (and other types of motivators) in real-time, synchronously and asynchronously from executive processes. There are two broad types of executive processes: deliberative processes (manifold base level reasoning, evaluating, planning, scheduling, deciding, and control, also known as management processes) and meta-management processes (reflection and high-level control), localized in the upper two levels of the H-CogAff architecture. The meta-management layer could, for instance, postpone the consideration of a newly activated goal till some juncture — deliberation scheduling. The "reactive" layer is more closely coupled to the environment than the other two. H-CogAff supposes variable-threshold qualitative and quantitative interrupt filters, which protect limited-capacity executive processes. A detailed specification of the structure of goals and the H-CogAff processes operating on them is provided in [4].

On the basis of this architecture, Sloman was able to distinguish three types of emotion [3,16]. *Primary* emotion involves alarms triggered by perceptual information, such as an angry glare or the unexpected appearance of the object of one's infatuation. *Secondary* emotion involves alarms triggered by noticing an executive layer's content (e.g., suddenly realizing a plan of action will or would have a disastrous side-effect). Alarms have *global* effects in the architecture, physiological or exclusively mental. Tertiary emotion—more generally, *perturbance*—involves an interaction, between motive activators, filters, deliberative and meta-management processes, etc. In perturbance, even if the deliberative layer were to postpone consideration of an insistent motivator, the motivator would still tend to penetrate the filter and *divert* deliberative processes; or the motivator might *maintain* control of executive processes. Perturbance is an *emergent* phenomenon; special cases aside, it is not necessarily adaptive or maladaptive. Adaptiveness and function are attributes of the architecture and its constituent mechanisms.

The potential of this theory for psychology derives partly from the research methods, the designer stance [67], that gave rise to it. This stance can address a deep issue that surrounds but has not previously been explicitly linked to psychology's "replication crisis" [17]. Psychology often lacks sufficient theory for the phenomena it empirically investigates. We call for (1) a better explicit characterization of human capabilities, an exploration of mental architectures (designs), and implementations [2]; and (2) empirical research driven by unified theories of mind [18,19]. *Cognitive* architectures, still largely ignored in psychology, are not enough; affective processes deserve equal consideration. H-CogAff is still incomplete; but it is a starting point worth considering.

4 TWO PERTURBANT EMOTIONS

Let us consider two emotions that can last for months, are eminently perturbant, but are insufficiently explained by general theories of emotion: grief and limerence. Perhaps this oversight is because these emotions cannot ethically be manipulated in the laboratory. Empirical psychology needs to be more concerned with explaining observed individual possibilities in detail (Newell, 1973), including detailed case studies, diary studies, correspondence studies (e.g., [20]) and fiction ([21-23]). Researchers could do worse than to try to design minds that support the perturbant emotions depicted in humanity's greatest works of romantic poetry and fiction —such as of the main characters in Shakespeare's Romeo & Juliet and Chekhov's *The Lady with the Dog.*

Grief. When grieving, one tends to be assailed by memories and motives about the deceased. Wright, Sloman & Beaudoin [24,25] offered a design-based explanation of emotion, illustrated by a case study of grief, in which they claimed grief is (often) "an extended process of cognitive reorganization characterized by the occurrence of negatively valenced perturbant states caused by an *attachment structure* reacting to news of the death." That theory addresses important questions such as: Why does grief consume the mourner? Because executive processes have limited capacity and become swamped by highly insistent motivators generated by a structure of attachment to a highly valued individual; in addition, re-learning and detachment require extensive rumination, which can maintain perturbance.

Limerence. The prototypical perturbant state is limerence: The nearly universal attraction phase of romantic love [26,27]. Notably, limerence researchers agree that limerence is characterized by focused attention on and intrusive mentation (IM) about the limerent object (LO) with manifold intense and insistent motives for union with the LO ([26]). Limerence is of great evolutionary and human significance, because it enhances the likelihood of mating—and, in most cultures, attaching to the LO, which helps offspring survive [28]. Yet affective scientists have hardly considered the phenomenon as a generally representative and illuminating emotion, let alone from the designer stance.

A defining *feature* of perturbance is diminishment of the already limited human capacity to control one's own attention. Consider a limerent's diary entry "“This obsession has infected my brain. I cannot shake those constantly intruding thoughts of you. Every thought winds back to you no matter how hard I try to direct its course in other directions.”" [26]). Many, perhaps most, limerent minds are aware of this intrusiveness. This is only possible because (unlike most species) humans can, to a limited extent, monitor and voluntarily control their attention (i.e., execute meta-management functions).

The H-CogAff framework seems to be at least as promising for limerence as it is for grief—two emotions that normally operate in opposite ways on attachment structures. Limerence, the attraction phase, involves establishing attachment structures: motivators, motive generators, insistence assignment rules, other reactive processes, filters, plans, etc. Grief is an extended process of dismantling such attachment structures. Limerence and grief overlap in heartbreak and lovelornness. Also like grief, limerence can loosen prior attachment (facilitating the abandonment of one's current partner for a new one, or forgetting a prior love). Accounting for attachment processes is important given that emotions seem to have evolved in large part to enable individuals to indirectly manage each other via commitments and attachments [29]. Several H-CogAff projects have already examined perturbance in relation to attachment (e.g., [30]).

While it may be tempting to cast limerence as a pathological form of romantic love [31,32], this would distort the original and common academic conception of limerence [33]. This would also overlook the near universality and evolutionary significance of limerence. Like other obsessions and other emotional states, limerence lies on continua [34] and may or may not be pathological. We believe the distorted casting should be resisted by scholars; instead other terms should be used to describe pathological limerence. We also recommend that scientific literature on this phase converge on the term 'limerence' and help shape folk psychology.

There is more to limerence than perturbance, just as there is more to motive processing and emotions than perturbance.

Perturbance is a particularly promising concept partly because it encourages questions to be raised *progressively* about mental states in terms of whole-mind design (motive generators, attachment structures, etc.), leading to further requirement and design specification. Perturbance cannot be understood in isolation. It transcends folk psychology and the intentional stance.

5 REPETITIVE AND INTRUSIVE MENTATION INVOLVE PERTURBANCE

Watkins [1] suggested that an important attentional phenomenon should be conceptualized as "repetitive thought" (RT). He echoed a definition of RT as a "process of thinking attentively, repetitively or frequently about one's self and one's world [forming] the core of a number of different models of adjustment and maladjustment." (p. 163) Under the banner of RT, Watkins included such varied phenomena as cognitive and emotional processing of persistent intrusions, depressive rumination, perseverative cognition, rumination, worry, planning, problem solving, and mental simulation, mind wandering, counterfactual thinking, post-event rumination, defensive pessimism, positive rumination, reflection, habitual negative self-thinking. To this list we would add obsessive and compulsive mentation and *cravings*. Watkins notes that worry, for instance, was defined in [35] as "a chain of thoughts and images, negatively affect-laden and relatively uncontrollable" and as "an attempt to engage in mental problem-solving on an issue whose outcome is uncertain but contains the possibility of one or more negative outcomes" (p. 9).

Watkins's reasons for favouring RT as the overarching concept were that it is more inclusive than the alternatives, atheoretical, clearer, highly correlated with measures of worry and rumination, and non-evaluative (constructive or unconstructive).

We agree that RT phenomena are scientifically significant. For many of them are typical of normal self-regulation—everyone experiences IM, for instance. Furthermore, extreme forms of RT are transdiagnostic [36]. RT plays a critical role in insomnia and depression, for instance [37]. Insomnia also is of transdiagnostic significance [37]—a cause and consequence of RT.

However, Watkins' RT conceptualization is limited. Firstly, the expression "RT" misleadingly suggests that the repetitive content is cognitive in the traditional sense ("thought"), whereas it is often affectively-laden. Moreover, the processes that manage the 'repetitive' mental content serve these motivators, such as assessing and deciding. Repetitive *mentation* (RM) is more inclusive and germane. Further, the atheoretical criterion is unrealistic and counterproductive; it also runs against Watkins's other criterion of being conceptually clear. One needs a general theory, beyond folk psychology, in relation to which intrusions and the executive processes that respond to them are specified. (Compare progress in evolutionary classification based on molecular genetics rather than phenotypic features.) Whether authors are explicit and clear or not about their theory, the concepts at play in RT involve, or at least require, a functional architecture. For something must be generating motivators; something must be interrupting in intrusions; something must be considering goals; something must be prioritizing them; etc. These mechanisms need to be named and specified in relation to an architecture. The theory ought to "cut nature at its joints" and be amenable to a progressive research programme of simulation, further theoretical development and cumulative empirical research [38]. Furthermore, the all inclusive *RT* conceptualization comes at the cost of papering over significant differences, for instance between reflection and rumination. The farrago of "RT" concepts requires conceptual analysis and functional specification, which will lead to much pruning and reclassification.

The phenomena of RM are too global, involving too many diverse wide-ranging mechanisms of mind, to be understood without reference to a broad and explicit theory of mind. Moreover, one must understand the *how* of normal information processing (IP) to assess mentation as constructive or unconstructive. Alas, the RT literature has failed to adopt or develop architectural models of mind. For instance, in describing a highly studied phenomenon of RM, affective biases, Mathews, Mackintosh & Fulcher [39] invoke interrupt signals, attentional vigilance, effortful suppression and intrusions. The concepts of cognitive and attentional 'biases' [68], are currently cast mainly in terms of 'external and internal stimuli' rather than in terms of goal or motive processing (contrast [4-5,61]), i.e., the mechanisms that are being affected. The attentional bias and RM literatures fail to invoke an overall model of mind which, for instance generates motives, prioritizes, them and acts upon them, i.e., that addresses the types of capabilities with which H-CogAff is concerned.

Wells & Mathews published a book length theory, the Self-Regulatory Executive Function (S-REF) model [40], that valiantly attempts to address many phenomena at the intersection of cognition and affect, including RT. The model is explicitly inspired by architecture-based AI. However, the empirical RT literature seems at most to pay lip service to it. For instance, in their extensive book on transdiagnostic processes, Harvey et al. [36] *summarily* reject S-REF. It is not noted that and how S-REF would need to be improved to address more of the requirements of autonomous agency — normal multi-purpose (multi-motive) competence. The main issue, that this promising underdeveloped theory needed AI attention was not mentioned.

Watkins (2008) and others point to control theory as an explanatory framework for "RT" and self-regulation. While some of these models are promising (e.g., [41]), they too need to be integrated with a broader architecture. They need to deal with rich qualitative control states and mechanisms that follow from the requirements of autonomous agency (see [42]).

H-CogAff provides a theoretical framework in relation to which classification and modelling may proceed. This framework has the advantage of being constructed to explore how human minds might solve real world problems of autonomous agency. It is by no means a complete or detailed specification; but it has proven to be useful for generating and exploring models, many of which have already been implemented [7,9].

H-CogAff offers a path towards a deeper conceptualization of "RT". In [1], intrusive thought (IT) is not a category of RT, likely because it is an essential *aspect* of RT. IT is better, and more generally, conceived as intrusive *mentation* (IM), and more deeply as *perturbance*. The concept of perturbance is based on the *dispositional* concept of insistence of mental content: a motivator may be insistent and yet not disrupt processing. To understand IM as perturbance we must specify in terms of an architecture (like H-CogAff) the ways in which insistence assignment, interrupt filtering and attention switching are effected.

This may also help address the need in the RT literature for a design-based taxonomy of patterns of executive processes. [4]

and [25] put forth several categories, such as oscillation between decisions, manifest perturbance, digressions and maundering. Several other patterns have been identified in the CogAffect project (e.g., [25,43]). These, and several types of phenomena labelled by Watkins as RT (such as worry and rumination) need to be systematically characterized in terms of patterns of interaction between management, reflective and reactive processes in H-CogAff.

6 OTHER PSYCHOLOGICAL LITERATURES IN NEED OF PERTURBANCE AND RELATED ARCHITECTURAL CONCEPTS

Several other research problems need to be reinterpreted specifically in terms of perturbance and, more generally, from the designer stance. Motivation in psychology tends to be conceived as the directing and energizing of behaviour [44] (what goals do people choose; when, why, and how intensely do they pursue them), rather than in terms of motive processing (how can motives be processed to evince autonomous agency). For instance, none of the *Behavior & Brain Sciences* peer responses to the Selfish Goal theory [45] noted its lack of explicit architecture nor that its goal specification and processes are bare (e.g., where is insistence? Contrast [4]). Pleasure and avoidance of pain are still normally assumed to be the *final* ends, while the deeper, more subtle and generative possibility of *architecture-based motivation* [46] is ignored even in rare discussions of effectance ([41]; contrast [47]). Stanovich developed a promising theory [48] to explain and improve rationality with a three-level architecture which, although referring to H-CogAff, fails to use motive processing constructs. Yet the perturbance theory was meant to account for breakdowns in rationality [5]. Meanwhile, the recent theory of cognitive energetics [49], which is meant to explain all instances of goal-directed thinking, also lacks an architecture (contrast the related concept of economy of mind in Wright [25]).

Given that perturbance is an underlying construct to explain RT, and RT is transdiagnostic, it stands to reason that the concept of perturbance is relevant to transdiagnostic approaches. For instance, addictions involve motivators that are both insistent (attention grabbing) and intense (control behaviour). Obsessions and compulsions also involve perturbance. More generally, a design-based approach is required for transdiagnostic understanding [50]. Even more generally, to understand abnormal psychology we must understand *normal* psychology in design-based terms.

Pain in its various forms involves aversive perturbance and should be modelled with H-CogAff or related designs.

Beaudoin [47] argued that mindfulness-based therapies, which are either explicitly behaviourist [51] or use architectures detached from AI, could benefit from H-CogAff. Mindfulness therapies assume *direct* experience [51]. But no one has ever built a machine that can directly perceive anything, nor demonstrated the possibility of such a machine—perception is in fact always highly indirect. Mindfulness therapies prescribe awareness of *emotion*, but by this term their authors mainly refer to *affective feelings*. Shouldn't therapists and clients be trained with a rich design-based theory of mind to improve clients' awareness, i.e., models of themselves? Similarly, the acceptance and commitment therapy (ACT) technique of "cognitive defusion" [51] requires an IP ontology of mental states that ACT fails to invoke.

Perturbance is also quite relevant to human memory. Following Anderson's adaptive explanation of memory [62], Beaudoin [47] proposed the heuristic relevance-signaling hypothesis ("HRS") from the designer stance. On a daily basis, humans process enormous amounts of information. The brain cannot deeply interpret it all, nor store all of its interpretations. Nor can the cortex directly signal relevance top down (The direct command "I shall remember this phone number" does not work.) What information should be given precedence? Testing effects are amongst the most well documented findings in empirical psychology: repeatedly recalling information potentiates it. The HRS hypothesis states that deliberative layer recall attempts are implicit cues to the brain's heuristic memory indexing mechanisms to prioritize access to information ('memories') related to the perturbance—information (interpretations, narratives, etc.) that the deliberative layer has at least *attempted* to recall (reconstruct). Perturbances are hijackings of these mechanisms by insistent motivators, potentiating memories related to the perturbant objects (e.g., the limerent object).

Psychology has struggled with the question: in what respect can the experience of music in particular and art more generally be emotional? From the designer stance we might similarly ask how can great art rivet us and reverberate within us, from catchy ear worms to more? We suggest a new answer based on H-CogAff theory, namely that music and fiction may trigger an *illusion of perturbance*: the reflective-layer impression that the agent is experiencing a genuine perturbance (as if self-generated motives were *insistently* being activated, captivating management processes). More obviously, art likely often operates by increasing the insistence of one's own latent motivators (triggering limerence and grief, for instance). To explore and specify these vague hypotheses, we suggest modeling responses to high-calibre, multi-modal art depicting limerence and grief that uses repetition in provocative ways, such as Veda Hill & Amiel Gladstone's musical theatre adaptation of Tchaikovsky opera, *Onegin* [64], itself based on Pushkin's poem, *Eugene Onegin*.

It should be noted that perturbance is not the only type of loss of control in minds. Dean Petters described several other types in relation to H-CogAff [43].

We also believe a theory of perturbance can be used for positive psychology and self-help. For example, Beaudoin (2013) developed the cognitive shuffle a technique to combat insomnia which is meant to work partly by interfering with bedtime perturbance [52]. Focusing and flow are essential to cognitive productivity and hence to knowledge economies. Distraction is largely affective yet theories of attention —and knowledge translation on the subject e.g. [53-54] Levitin (2014), Gallagher (2006) — do not deal with motive processing and fail to invoke perturbance. Theories of learning, expertise and productive practice need to explain how humans can deliberately develop their mental architectures, e.g., creating new goal generators [47,55-56].

In short, previous research phenomena and problems can systematically be revisited from the designer stance as involving perturbance.

7 CONCLUSION

We have called attention to perturbance as a way to understand a broad variety of normal and pathological mental phenomena in IP terms. This concept has the advantage of being firmly rooted in AI and of involving a flexible, extensible architectural framework. This enables research problems to be considered in terms of models of entire minds.

Perturbance and other aspects of H-CogAff are not final explanations. They are part of the beginning of what we believe can be a progressive research programme.

The designer stance also is directly relevant to education and training. Psychology students need to be able to think about themselves, other humans and possible minds in terms of multiple cognitive-affective IP architectures. Psychology and AI students should also graduate well-trained in conceptual analysis [57,58] as they are in empirical research methods. (These would be fitting topics in [59], for example.)

We are not suggesting a one-way flow of influence. Instead, we advocate a progressive theory-driven research programme to improve H-CogAff and related proposals. There is a need for more AI researchers to consider broad, integrative, multi-layered, affective autonomous agency. We believe psychology and AI researchers need to work more closely together, not only on purely cognitive problems but affective ones as well. AI and psychology must blend more. For the opening quotation of Beaudoin's (1994) Ph.D. thesis [4] is still true: "The problem is not that we do not know which theory is correct, but rather that we cannot construct any theory at all which explains the basic facts" [60] (p. 109.)

ACKNOWLEDGMENTS

We would like to thank Dr. Al Sather, Carol Woodworth and two referees for their comments which helped improve this paper.

REFERENCES

[1] E. R. Watkins, "Constructive and unconstructive repetitive thought," *Psychological Bulletin* **134**, 163–206 (2008).

[2] A. Sloman, "Prospects for AI as the general science of intelligence," 1993, Amsterdam, 1–10, IOS Press.

[3] A. Sloman, "How many separately evolved emotional beasties live within us?," in *Emotions in humans and artifacts*, R. Trappl, P. Petta, and S. Payr, Eds. (MIT Press, 2003).

[4] L. P. Beaudoin, "Goal processing in autonomous agents" (Birmingham, England, 1994).

[5] A. Sloman and M. Croucher, "You don't need a soft skin to have a warm heart: Towards a computational analysis of motives and emotions," 004, 1981.

[6] A. Sloman and M. Croucher, "Why robots will have emotions," 1981.

[7] A. Sloman, "The Cognition and Affect project: Architectures, architecture-schemas, and the new science of mind," 2008.

[8] Q. Ke, E. Ferrara, F. Radicchi, and A. Flammini, "Defining and identifying Sleeping Beauties in science," *Proceedings of the National Academy of Sciences* **112**, 7426–7431 (2015).

[9] N. Hawes, "A survey of motivation frameworks for intelligent systems," *Artificial Intelligence* **175**, 1020–1036 (2011).

[10] E. Hudlicka, "Affective BICA: Challenges and open questions," *Biologically Inspired Cognitive Architectures* **7**, 98–125 (2014).

[11] T. Read and A. Sloman, "The terminological pitfalls of studying emotion," 1–8 (1993).

[12] C. E. Izard, "The many meanings/aspects of emotion: Definitions, functions, activation, and regulation," *Emotion Review* **2**, 363–370 (2010).

[13] J. A. Russell, "Emotion, core affect, and psychological construction," *Cognition & Emotion* **23**, 1259–1283 (2009).

[14] K. R. Scherer, "What are emotions? And how can they be measured?," *Social Science Information* **44**, 695–729 (2005).

[15] J. Panksepp and L. Biven, "The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions" (2012).

[16] A. Sloman, R. Chrisley, and M. Scheutz, "The architectural basis of affective states and processes," in *Who needs emotions? The brain meets the robot*, J. M. Fellous and M. A. Arbib, Eds. (New York: Oxford University Press, 2005).

[17] S. E. Maxwell, M. Y. Lau, and G. S. Howard, "Is psychology suffering from a replication crisis? What does 'failure to replicate' really mean?," *The American psychologist* **70**, 487–498 (2015).

[18] A. Newell, *Unified theories of cognition* (Harvard University Press, Cambridge, MA, 1990).

[19] A. Wells and G. Mathews, *Attention and Emotion: A Clinical Perspective* (Lawrence Erlbaum Associates Publishers, Hillsdale, NJ:, 1994).

[20] L. Nys, "Emotional 'counter-practices' in the discipline section of the state re-education institution for female juvenile delinquents (1927-1939)," 30 July 2015, 1–16.

[21] K. Oatley, *Such stuff as dreams: The psychology of fiction* (2011).

[22] K. Oatley, *Best Laid Schemes* (Cambridge Univ Press, Cambridge, 1992).

[23] P. C. Hogan, *What Literature Teaches Us about Emotion* (Cambridge University Press, 2011).

[24] I. Wright, A. Sloman, and L. P. Beaudoin, "Towards a design-based analysis of emotional episodes," *Philosophy, Psychiatry & Psychology* **3**, 101–126 (1996).

[25] I. P. Wright, "Emotional Agents" (1997).

[26] D. Tennov, *Love and Limerence* (Scarborough House, 1979).

[27] S. E. Reynolds, "'Limerence': A new word and concept," *Psychotherapy* **20**, 107–111 (1983).

[28] H. E. Fisher, "Lust, attraction, and attachment in mammalian reproduction," *Human Nature* **9**, 23–52 (1998).

[29] M. Aubé, "Unfolding commitments management: A systemic view of emotions," in *Handbook of research on synthetic emotions and sociable robotics New applications in affective computing and artificial intelligence*, J. Vallverdú and D. Casacuberta, Eds. (New York, NY, 2009).

[30] D. Petters and L. P. Beaudoin, "Attachment modelling: From observations to scenarios to designs," in *Computational Neurology and Psychiatry*, P. Erdi, B. S. Bhattacharya, and A. Cochran, Eds. (2017).

[31] A. Wakin and D. B. Vo, "Love-variant: The Wakin-Vo IDR model of limerence," 2008.

[32] M. Reynaud, L. Karila, L. Blecha, and A. Benyamina, "Is Love Passion an Addictive Disorder?," *The American Journal of Drug and Alcohol Abuse* **36**, 261–267 (2010).

[33] H. van Steenbergen, S. J. E. Langeslag, G. P. H. Band, and B. Hommel, "Reduced cognitive control in passionate lovers," *Motivation and Emotion*, 444–450 (2013).

[34] E. Hatfield and S. Sprecher, "Measuring passionate love in intimate relationships," *Journal of Adolescence* **9**, 383–419 (1986).

[35] T. D. Borkovec, E. Robinson, and T. Pruzinsky, "Preliminary exploration of worry: Some characteristics and processes," *Behaviour Research and Therapy*, 9–16 (1983).

[36] A. G. Harvey, *Cognitive Behavioural Processes Across Psychological Disorders* (Oxford University Press, USA, 2004).

[37] M. R. Dolsen, L. D. Asarnow, and A. G. Harvey, "Insomnia as a transdiagnostic process in psychiatric disorders," *Curr Psychiatry Rep* **16**, 471 (2014).

[38] R. P. Cooper, "The role of falsification in the development of cognitive architectures: Insights from a Lakatosian analysis,"

Cognitive Science **31**, 509–533 (2007).

[39] A. Mathews, B. Mackintosh, and E. P. Fulcher, “Cognitive biases in anxiety and attention to threat,” *Trends in cognitive sciences* **1**, 340–345 (1997).

[40] A. Wells and G. Matthews, *Attention and Emotion* (Psychology Press, 1995).

[41] O. Nafcha, E. T. Higgins, and B. Eitam, “Control feedback as the motivational force behind habitual behavior,” in *Motivation - Theory, Neurobiology and Applications* **229** (Elsevier, 2016).

[42] A. Sloman, *Beyond turing equivalence. Red. P. Millican, A. Clark. Machines And Thought: The Legacy Of Alan Turing, vol I: 179-219* (1990).

[43] D. Petters, “Losing control within the H-Cogaff architecture,” in *From animals to robots and back: Reflections on hard problems in the study of cognition* **22** (Springer International Publishing, Cham, 2014).

[44] K. Danziger, *Naming the Mind* (SAGE, 1997).

[45] J. Y. Huang and J. A. Bargh, “The Selfish Goal: Autonomously operating motivational structures as the proximate cause of human judgment and behavior,” *The Behavioral and Brain Sciences* **38**, 121–135 (2015).

[46] A. Sloman, “Architecture-based motivation vs. reward-based motivation,” 2015.

[47] L. P. Beaudoin, *Cognitive productivity: Using knowledge to become profoundly effective* (CogZest, Pitt Meadows, BC, 2014).

[48] K. E. Stanovich, *Rationality and the reflective mind* (Oxford University Press, USA, 2011).

[49] A. W. Kruglanski, J. J. Bélanger, X. Chen, C. Köpetz, A. Pierro, and L. Mannetti, “The energetics of motivated cognition: A force-field analysis.,” *Psychological Review* **119**, 1–20 (2012).

[50] E. Hudlicka, “Computational modeling of cognition‐emotion interactions: Theoretical and practical relevance for behavioral healthcare,” in *Handbook of Affective Sciences in Human Factors and HCI*, M. P. Jeon, Ed. (Elsevier, Waltham, MA, 2017).

[51] S. C. Hayes, K. D. Strosahl & K. G. Strosahl. *Acceptance and commitment therapy: The process and practice of mindful change*, Guilford Press, New York, 2011.

[52] L. P. Beaudoin, N. Digdon, and K. O'Neill, “Serial diverse imagining task: A new remedy for bedtime complaints of worrying and other sleep-disruptive mental activity,” 2016, A209.

[53] D. J. Levitin, *The organized mind: Thinking straight in the age of information overload* (2014).

[54] W. Gallagher, *Rapt* (Penguin, 2009).

[55] P. H. Winne, “Self-regulated learning viewed from models of information processing,” in *Self-regulated learning and academic achievement: Theoretical perspectives*, 2nd ed., B. J. Zimmerman and D. H. Schunk, Eds. (Lawrence Erlbaum, Mahwah, NJ, 2001).

[56] L. P. Beaudoin, “Developing expertise with objective knowledge: Motive generators and productive practice,” in *From Robots to Animals and Back*, J. Wyatt and D. Petters, Eds. (Springer, 2014).

[57] A. Sloman, *The computer revolution in philosophy: Philosophy, science and models of mind* (Harvester Press, 1978).

[58] A. Ortony, G. L. Clore, and M. A. Foss, “The referential structure of the affective lexicon,” *Cognitive Science: A Multidisciplinary Journal* **11**, 341–364 (1987).

[59] K. E. Stanovich, *How to think straight about psychology*, 9 ed. (Allyn & Bacon, 2004).

[60] R. Power, “The organisation of purposeful dialogues,” *Linguistics* **17**, 107–152 (1979).

[61] H. A. Simon. "Motivational and emotional controls of cognition" *Psychological Review*, **74**, 29–39 (1967).

[62] Anderson, John R. "Is human cognition adaptive?" Behavioral and Brain Sciences 14, 471-485 (1991).

[63] P. N. Johnson-Laird & K. Oatley "Emotions, Music, and Literature" in *Handbook of Emotions* L. F. Barrett, M. Lewis & J. M. Haviland-Jones (Eds). (New York, 2008).

[64] V. Hill & E. Gladstone. *Onegin*. (2016) http://artsclub.com/shows/2015-2016/onegin

[65] R. W. Picard. *Affective Computing*. (MIT Press, 2000).

[66] A. Sloman. "Review of: Rosalind Picard’s affective computing." *AI Magazine*, 20, 127-137 (1997). http://www.cs.bham.ac.uk/research/projects/cogaff/Sloman.picard.review.pdf

[67] J. McCarthy. "The well-designed child." Artificial Intelligence, 172, 2003-2014 (2008).