



Curing the broken brain model of addiction: Neurorehabilitation from a systems perspective

Reinout W. Wiers^{a,b,*,1}, Paul Verschure^{c,d,e,1}

^a University of Amsterdam, Addiction Development and Psychopathology (ADAPT)-lab, Department of Psychology, Amsterdam, The Netherlands

^b Center for Urban Mental Health, University of Amsterdam, Amsterdam, The Netherlands

^c Laboratory of Synthetic, Perceptive, Emotive and Cognitive Systems (SPECS), Institute for Bioengineering of Catalonia (IBEC), Barcelona Spain

^d The Barcelona Institute of Science and Technology, Barcelona Spain

^e Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

HIGHLIGHTS

- Neurorehabilitation from addiction should be considered from a systems-perspective.
- System-oriented neurorehabilitation is contrasted with a “broken brain” perspective.
- System-oriented neurorehabilitation includes high-level concepts related to context and motivation.
- System-oriented neurorehabilitation emphasizes the dynamic nature of the addicted mind and brain.

ARTICLE INFO

Keywords:

Addiction
Brain disease model
Neurorehabilitation
Systems approach

ABSTRACT

The dominant biomedical perspective on addictions has been that they are chronic brain diseases. While we acknowledge that the brains of people with addictions differ from those without, we argue that the “broken brain” model of addiction has important limitations. We propose that a systems-level perspective more effectively captures the integrated architecture of the embodied and situated human mind and brain in relation to the development of addictions. This more dynamic conceptualization places addiction in the broader context of the addicted brain that drives behavior, where the addicted brain is the substrate of the addicted mind, that in turn is situated in a physical and socio-cultural environment. From this perspective, neurorehabilitation should shift from a “broken-brain” to a systems theoretical framework, which includes high-level concepts related to the physical and social environment, motivation, self-image, and the meaning of alternative activities, which in turn will dynamically influence subsequent brain adaptations. We call this integrated approach *system-oriented neurorehabilitation*. We illustrate our proposal by showing the link between addiction and the architecture of the embodied brain, including a systems-level perspective on classical conditioning, which has been successfully translated into neurorehabilitation. Central to this example is the notion that the human brain makes predictions on future states as well as expected (or counterfactual) errors, in the context of its goals. We advocate system-oriented neurorehabilitation of addiction where the patients' goals are central in targeted, personalized assessment and intervention.

1. Introduction

During the past two decades, the dominant biomedical model of substance use disorders or addictions has been that of chronic brain disease (Leshner, 1997; Volkow, Koob, & McLellan, 2016). From this perspective, now often referred to as the brain disease model of addiction (BDMA, Hall, Carter, & Forlini, 2015; Heather et al., 2018), a

vulnerable brain may get “hijacked” by addictive drugs (Nesse & Berridge, 1997). Various neuroadaptations are thought to make behavior increasingly less voluntary and more compulsive, especially when cues indicate the presence of an opportunity to engage in addiction-related behavior. One proposed underlying mechanism concerns increased dopamine-mediated “wanting”, the neural substrate of subjective craving (Berridge & Robinson, 2003, 2016; Robinson & Berridge,

* Corresponding author.

E-mail addresses: r.wiers@uva.nl (R.W. Wiers), paul.verschure@specs-lab.com (P. Verschure).

¹ Both authors contributed equally to the ms.

<https://doi.org/10.1016/j.addbeh.2020.106602>

Received 10 January 2020; Received in revised form 9 July 2020; Accepted 4 August 2020

Available online 08 August 2020

0306-4603/ © 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1993), combined with reduced frontal cortical control over the effects of the impulses driving addictive behavior (Goldstein & Volkow, 2011; Jentsch & Taylor, 1999). Second, it has been argued that cue-induced responding may become habitual in a strong sense (compulsive), where a stimulus is enough to elicit the response, even in the absence of (expected) reinforcement or when the cue has been associated with punishment (Everitt & Robbins, 2005, 2016). Third, over time, addictive behaviors typically lead to mounting negative affect and associated negative reinforcement driven behavior to alleviate it (Koob & Le Moal, 2008). Dependent on the various pharmacological pathways activated by the substances involved, the dominant mechanisms may differ (Badiani, Belin, Epstein, Calu, & Shaham, 2011). Irrespective of this variability, the overall perspective is that the brain has chronically changed as a consequence of the addictive behavior (to what extent this is also the case in non-substance addictions is subject to debate), which has led to the loss of voluntary control, or “loss of free will” (Volkow, 2015). According to this perspective, there is no road back to controlled use or recovery (in line with ideas from Alcoholics Anonymous or 12 steps programs, Segal, 2017), and neurorehabilitation can only be partially effective (as the brain disease is chronic). The main reason is that the addicted or post-addiction brain would remain hypersensitive to conditioned cues signaling opportunities to engage in addictive behavior, unless the brain could be permanently cured, for example by new medications. Yet, medication development in psychiatry is in a crisis due to the lack of understanding of the pathophysiological mechanisms underlying chronic brain disease (Hall et al., 2015; cf., Hyman, 2013). Other brain-oriented interventions have been proposed, such as varieties of neurostimulation (Gorelick, Zangen, & George, 2014) with as of yet unclear impact.

During the past decade, the BDMA has been criticized for various reasons, including its inability to accommodate sudden recovery, even after severe addiction (Lewis, 2015; Longo & Lewis, 2018), epidemiological data pointing to recovery as the most frequent long-term outcome of addiction (Baumeister, 2017; Heyman, 2010), the limitations of generalizing from animal models to human pathology (de Wit, Epstein, & Preston, 2018; Field & Kersbergen, 2020), and the neglect of psychological and social factors in both the development of addiction and in its recovery (Hart, 2013; Heather et al., 2018). In terms of clinical and social implications, the BDMA has been advanced as beneficial for patients because it counters the moral perspective (the addicted individual is responsible for their self-destructive choices), but the BDMA may also negatively impact confidence in a positive outcome (self-efficacy) in both therapists and patients (Barnett, Hall, Fry, Dilkes-Frayne, & Carter, 2018), and may increase stigma, as a recent meta-analysis indicated (Loughman & Haslam, 2018).

In the present paper, we propose an alternative model to the BDMA, which on the one hand acknowledges that there are neuroadaptations in addiction, but on the other hand emphasizes the dynamic and integrated nature of the human mind and brain. From this systems-perspective, the neural level is merely one of the multiple layers of organization that defines human behavior, including addictive behaviors. This more dynamic multi-scale conceptualization, questions the validity and utility of the *chronic* brain disease concept. Instead, we propose to place addictive behaviors in the broader context of the addicted brain that drives behavior, which is the substrate of an addicted mind situated in a physical and socio-cultural environment. From this perspective, targeted psycho-social, cognitive, and behavioral rehabilitation can mitigate addiction, which in turn will dynamically influence subsequent brain adaptations (Lewis, 2015). We call this integrated approach *system-oriented neurorehabilitation*.

From the present perspective, the treatment of addiction requires research into effective personalized interventions aimed at system-oriented neurorehabilitation. Importantly, this includes high-level concepts and interventions related to the physical and social environment, motivation, self-image, and the meaning of alternative activities. These will, in turn, drive changes at the neuronal level, such as the

desensitization to addiction-associated cues and behaviors. To illustrate the system-oriented neurorehabilitation approach, we build on the Distributed Adaptive Control (DAC) theory, which conceptualizes mind and brain as complementary properties of a multi-layered architecture that controls action (Verschure, 2016). We illustrate how such a system-level and embodied action-oriented perspective can guide new developments in the rehabilitation of addiction.

2. Addiction and choice

The dominant account of human behavior during the past 50 years, has been that behavior is purposeful: people generally choose to do things of which they expect positive outcomes and refrain from doing things from which they expect negative outcomes, hence their behavior can be described as reasoned, rational or goal-driven (Ajzen & Kruglanski, 2019; Kruglanski & Szumowska, 2020; Tolman, 1966). However, addictive behaviors are hard to understand from this perspective, as they show features of irrationality: can people willfully act against their own goals? (the classical problem of *Akrasia*, see Heather, 2017; Wiers, Van Gaal, & Le Pelley, 2020). First, it is important to note that many behaviors can serve multiple goals (Kruglanski & Szumowska, 2020). For example, having dinner not only serves the goal to obtain food, but also social goals, which may make dietary restraints less relevant in a festive context (Stroebe, Van Koningsbruggen, Papies, & Aarts, 2017). Second, the main perspective from which behavior serves goals is egocentric: in many cases where other people would judge a behavior as irrational and self-destructive, the behavior may actually be purposive, because it serves a salient goal for the actor (see Kopetz & Orehek, 2015).

In psychology, one class of models developed to explain seemingly irrational behaviors concerns dual process models (e.g., Gawronski & Bodenhausen, 2006; Kahneman, 2003; Strack & Deutsch, 2004), which have also been developed for addictions (e.g., Bechara, 2005; Wiers et al., 2007). According to these models, a stimulus can automatically trigger an inclination to act (due to reward-learning and/or habit formation, mediated by an impulsive or associative system), which can be overcome by a reflective and deliberate or rule-based system provided that there is enough capacity (and motivation) to do so. However, theoretical problems have been identified with these models, including the motivational homunculus problem (how does the reflective system know which impulses to inhibit? Gladwin, Figner, Crone, & Wiers, 2011). Moreover, the neural substrate of dual process models is ill-defined (Keren & Schul, 2009). For these and other reasons (see Hommel & Wiers, 2017; Melnikoff & Bargh, 2018), many theorists moved to a position where the central problem in addiction is *biased* choice rather than *loss* of choice (Field et al., 2020; Gladwin et al., 2011; Hogarth, 2020; Wiers & Gladwin, 2016; Wiers, Van Dessel, & Köpetz, 2020; Wiers, Van Gaal et al., 2020). Note that from this perspective, the role of brain areas traditionally associated with inhibition may receive a different interpretation, namely biasing the integration of information supporting specific choices towards valuing long-term rather than short-term gains (Berkman, Hutcherson, Livingston, Kahn, & Inzlicht, 2017; Field et al., 2020; Gladwin et al., 2011). Note further that in (computational) neuroscience, model-free (MF) and model-based (MB) processes have been distinguished (Daw, Niv, & Dayan, 2005), where MF mechanisms have been treated as producing automatic stimulus-response habits, contrasting with MB strategies generating goal-directed choices based on a model of the world which generates predictions on outcomes. However, similar conceptual problems have been identified with this distinction (see Hommel, 2019, and Section 3 below).

Although we agree that neuroadaptations in addiction may influence the decision making process (indeed, biased decision-making is central in our account), we would argue that a total loss of choice, as suggested by BDMA, may only happen in extreme cases resulting from severe collateral damage as in severe Korsakov syndrome (Fenton &

Wiers, 2017). Note that many if not most people suffering from addictions recover without formal treatment and the notion that most people relapse has been argued to be a misperception based on studying only clinical samples (Baumeister, 2017). Importantly, researchers are beginning to correct for this bias, by also considering natural trajectories into and out of addictions (for example in a new large German research consortium, Heinz et al., 2020). It has been argued that excessive habit-formation would make addictive behavior compulsive (Everitt & Robbins, 2005, 2016), which would result in addiction-cues eliciting drug use even in the awareness of negative outcomes. First, it should be noted that the evidence for this account has primarily come from animal studies in experimental paradigms where choice is limited (Hogarth, 2020). Strikingly, recent studies show that when social alternatives are present (i.e., social interactions with other rats), drug choice is largely abandoned (Venniro et al., 2018). Second, evidence for habitual behavior, in the strong sense that it has become immune for negative consequences in humans is limited (De Houwer, 2019; Hogarth, 2020; Kruglanski & Szumowska, 2020). Of course, this does not mean that behavior cannot be habitual in a more colloquial sense of frequent and well-rehearsed, which may lead to slips-of-action, but these are typically repaired in line with original goal-pursuit (Kruglanski & Szumowska, 2020). Similarly, neuroadaptations following a reward make cues signaling the potential reward attractive (Berridge & Robinson, 2016; Robinson & Berridge, 1993), but this does not mean that the behavior becomes totally cue-driven and inflexible, merely that one behavioral option (the addictive behavior) becomes more attractive and probable, once primed in the sense of biased choice competition. Finally, there is evidence both from animal and human research that stress and negative affect may promote addictive behaviors, but the evidence supports the case for biased choice in favor of the addictive behavior rather than totally inflexible compulsive habitual behaviors (review: Hogarth, 2020).

The current evidence supports an account where biased goal-directed choice is central in addictive behaviors and sources of bias can include neuroadaptations as a consequence of experience and learning history. These may be further fueled in addictive behaviors by vulnerability factors (e.g., genetics, early life stress), and in case of substance addictions, by effects of the substances on these more general learning mechanisms. For example, many drugs have effects on the mesolimbic dopamine system which may strengthen the motivational significance of associated cues (Berridge & Robinson, 2016). In terms of neurocognitive processing, this may lead to an enhanced attentional salience of drug related cues (Berridge & Robinson, 2016; Franken, 2003), as has been found for reward-cues in general (Le Pelley, Pearson, Griffiths, & Beesley, 2015; Watson, Pearson, Wiers, & Le Pelley, 2019) as well as for drug associated cues (Anderson, 2016; Wiers, Van Dessel et al., 2020; Wiers, Van Gaal et al., 2020). Initial attentional capture by reward cues may be very difficult to control, for example, male volunteers could not prevent looking at nudes, even when they would be highly rewarded when successful (Most, Smith, Cooter, Levy, & Zald, 2007), and the same may be the case in addiction (Childress et al., 2008; Ingjaldsson, Thayer, & Laberg, 2003). However, subsequent responses can be trained, with positive effects on treatment outcomes, even in severely addicted people, as work on attentional bias modification and approach bias modification in addicted patients has demonstrated (outlined further in Section 4). For example, in the first small RCT of attentional training in alcohol-dependent patients, no effect of training was found for attentional engagement (200 ms), but the later response (500 ms) was successfully modified, which was related to later relapse (Schoenmakers et al., 2010). Hence, from this perspective, drug use remains volitional, throughout different stages of addiction, but the volitional choice process becomes biased. This can be modelled, for example, with drift-diffusion models (Field et al., 2020; Lin, Saunders, Fries, Evans, & Inzlicht, 2020; Wiers, Van Gaal et al., 2020). In these models, experience with addictive behaviors (in interaction with vulnerability factors), affect the decision-making space, for example, by

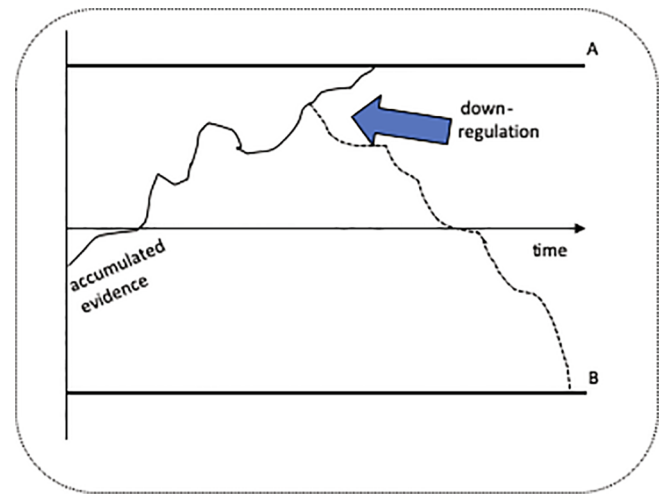


Fig. 1. Decision-making as modelled in drift-diffusion models. In this schematic picture, there is a choice between an addictive behavior (A) and an alternative behavior (B). A is triggered, for example by a conditioned cue, which “pulls” the decision-making toward the threshold of A (once the threshold is passed, an action is initiated). If long-term goals favor B, the decision-making can still be down-regulated toward choice B. The process of becoming addicted, from this model, can change the drift rate (steeper curve toward decision A) and/or lower the decision-threshold. Note that in the example, the chronic brain disease would mean that the decision threshold for A is already reached before downregulation in view of long-term goals can begin, and that this would not change after prolonged abstinence. Fortunately, evidence favors a model in which decision-making is biased, but not in this strong sense.

lowering the boundary value to be reached for a decision and by increasing the drift rate (see Fig. 1). Hence, there is still volitional choice in addiction, but the underlying motivational processes have been affected, favoring the choice to continue the addictive behavior, once triggered by conditioned cues. Note that this bias can be experienced as subjective craving, but this is not necessarily the case (Baumeister, 2017; May, Kavanagh, & Andrade, 2015). These neuroadaptations could only be described as a chronic brain disease if the addiction (and its collateral damage such as in Korsakov syndrome), would make it impossible to overcome this initial action tendency once triggered, which is rarely, if ever, the case, and if this would not revert after prolonged abstinence (see Heather et al., 2018; and see for promising results regarding neurorehabilitation in Korsakov, Loijen et al., 2018).

3. The brain as integrated prediction Machine: Distributed adaptive control

The review of the literature shows that the field of addiction research is facing a number of apparent dilemmas. These can be brought back to dichotomies between compulsive addictive behaviors (varieties of the BDMA) on the one hand, and examples of spontaneous recovery (Heyman, 2010) and responsiveness to small motivational interventions (Miller, 2000), on the other hand. These dilemmas may sometimes be the result from over-focusing on single mechanism-oriented interpretations. Here we advance the view that these putative appositions can be overcome when we place them in the context of the system as a whole. We will develop this perspective by taking a system-level architecture-oriented view.

An increasingly influential perspective on the brain is that it is a hierarchically organized adaptive prediction system (Friston, 2009; Massaro, 1997; Verschure & Althaus, 2003; Verschure & Pfeifer, 1992). This view changes our understanding of human decision making because it allows us to consider behavioral control as resulting from an integrated multi-layered hierarchical architecture (Verschure, 2018). While the highest cognitive level, including propositional reasoning and

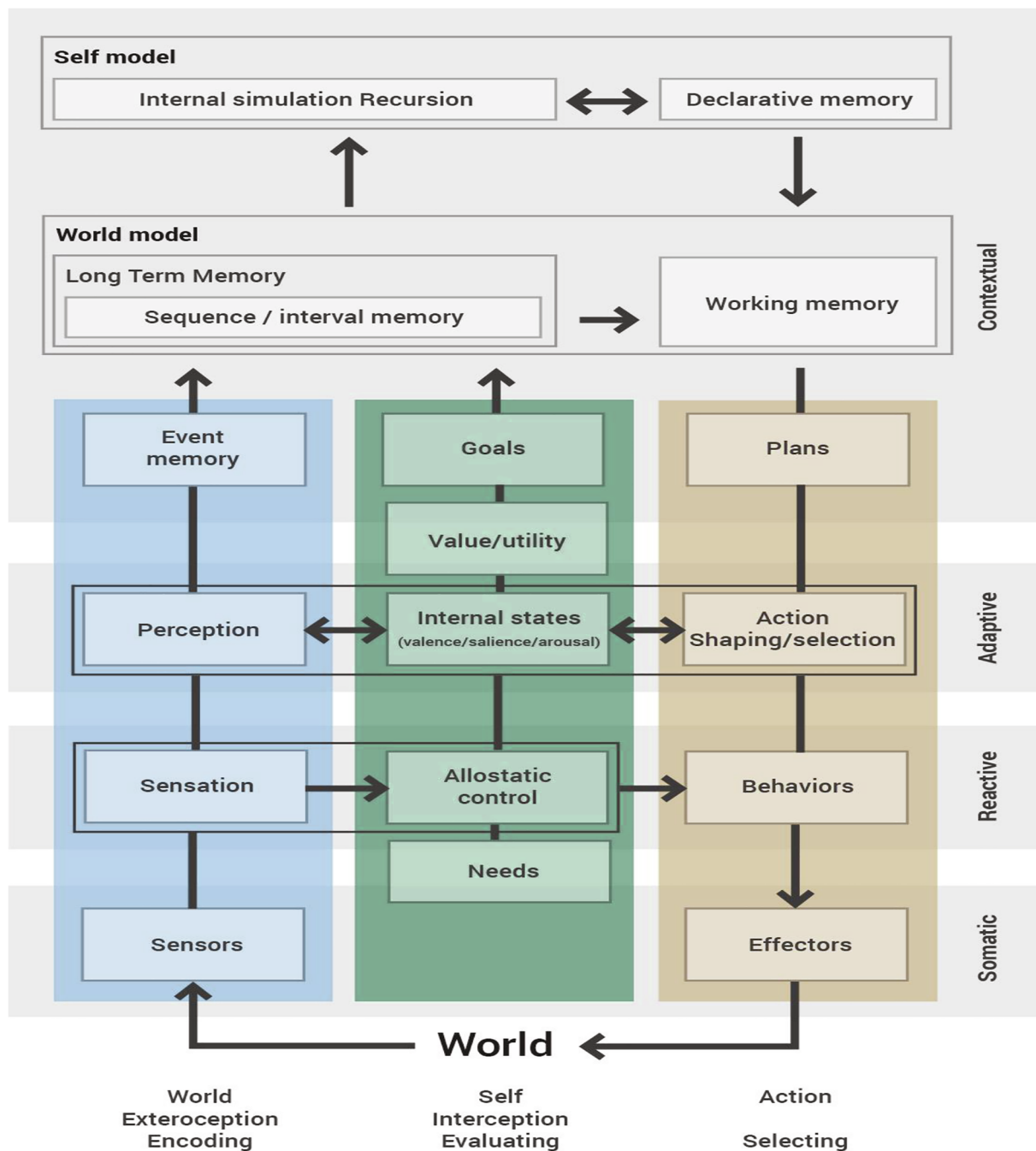


Fig. 2. A highly abstracted representation of the Distributed Adaptive Control (DAC) theory showing its main processes (boxes) and dominant information flows (arrows). DAC is organised along four layers (Soma, Reactive, Adaptive and Contextual) and three columns (World, Self, Action). Across these layers three functional columns of organisation exist: exosensing, the sensation and perception of the *external world* (left, blue); endosensing, detecting and signalling states derived from the physically instantiated *self* (middle, green), and action which establishes the interface between self and the world *action* (right, yellow). The arrows show the primary flow of information, mapping exo- and endosensing into action, defining a continuous loop of interaction with the world. Soma designates the body and its sensors, organs and actuators. It defines the needs, or Self Essential Functions (SEF), the organism must satisfy to survive. The Reactive Layer (RL) comprises dedicated Core Behaviour Systems (CBS) each implementing predefined sensorimotor mappings serving the SEFs. To allow for action selection, task switching and conflict resolution, all BSs are in turn regulated via an allostatic controller that sets their internal homeostatic dynamics relative to overall system demands and opportunities. The Adaptive Layer (AL) acquires representations of the states of the world and the agent and shapes action constrained by the value functions derived from the allostatic control of the RL. Learning by the AL minimises perceptual and behavioural prediction error, building a model free action generation system. The Contextual Layer (CL) further expands the time horizon in which the agent can operate, realising model-based policies, through the use of sequential short and long-term memory systems (STM and LTM respectively). STM acquires conjunctive sensorimotor representations that are generated by the AL as the agent acts in the world. STM sequences are retained as goal-oriented models in LTM when positive value is encountered, as defined by the RL and AL. The contribution of these stored LTM policies to goal-oriented decision-making depends on four factors: perceptual evidence, memory chaining, valence and the expected cost of reaching a given goal state (Verschure & Althaus, 2003). The content of working memory (WM) is defined by the memory dynamics that represents DAC's four-factor decision-making model. The autobiographical memory system allows the restructuring of memory around the unifying notion of Self which DAC proposes is essential to engage with the social world serving a "other like self" social perception model. See text for further explanation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

deliberation, remains essential, it may in specific cases be limited in overriding action tendencies triggered by lower-level mechanisms, resulting in biased decision-making. This view is congruent with models of psychopathology that posit that in human decision making, goal-directed reasoning is guiding, but that in (severe) psychopathology, this may be compromised (e.g., Moors, Boddez, & De Houwer, 2017).

To elaborate the changing perspective on mind and brain and to advance an alternative for the BDMA, we turn to a specific example that combines the view of the predictive brain with that of embodied cognition: the DAC theory of mind and brain (see Fig. 2). In DAC the mind/brain is modelled as a multi-level control system that maintains a multi-stable equilibrium between the embodied agent and its partially predictable and a priori unknown environment through action. Action results from optimizing the core objective functions of why (motivation), what (objects), where (space), when (time) and who (intention of other agents) (Verschure, 2012, 2018). A distinguishing feature of the DAC architecture is that each layer is an integral part of a larger system. For instance, the activation of a reflex at the level of the reactive layer carries critical information on the interaction between the agent and the world. In addition to triggering reflexive behavior, it also generates feedforward control signals for subsequent layers, that in turn drive action, epistemic needs, and conflict resolution. In this way the DAC architecture can bootstrap its knowledge structures and derived plans for action from simple priors. Thus, rather than operating with a fixed configuration, DAC proposes that the brain as a control system continuously re-configures its functional organization, engaging different layers of control to satisfy varying demands. Through this dynamic reconfiguration, a range of trade-offs are addressed such as that between robustness and complexity, speed and stability or exploration and exploitation. The DAC architecture thus follows the general evolutionary principle of being constructed from “constraints that deconstrain” (Doyle & Csete, 2011; Kirschner & Gerhart, 2008). For instance, dopamine affords reward-based learning, which deconstrains the potential task space of the agent, but in turn can be coopted by drugs of abuse which constrains choice, resulting in biased choice. By committing to distinct priors (constraints), the DAC architecture approximates the requirement of general intelligence, where anything can become a task (Newell, 1994). In addition, operations at higher layers of the architecture are progressively performed on virtualized states of the environment whereas the reactive layer responds to analog signals transduced directly from it (e.g., an Unconditioned Stimulus, US). At the cognitive control level of the contextual layer inferences are made over probability distributions of internally represented states, derived from internal models which bias action selection (Verschure, 2016).

The adaptive and contextual layers of DAC explain key features of classical and operant conditioning, respectively. This provides a direct link to our understanding of important aspects of addiction and its underlying biased choice. The adaptive layer as a model of classical conditioning acquires the state space of the environment through interaction, deploying a prediction-based local learning rule (Verschure, Voegtlin, & Douglas, 2003). This learning model directly captures the law of associative competition of classical conditioning (Rescorla & Wagner, 1972): animals only learn when events violate their expectations. More specifically, DAC implements the two-phase model of classical conditioning proposed initially by Konorski (Miller & Konorski, 1928), where classical conditioning is seen as comprising a fast perceptual and slow procedural learning stage. This division has been mapped to the amygdala, basal forebrain, and sensory cortices and the cerebellum, respectively (Medina, Christopher Repa, Mauk, & LeDoux, 2002). Where the former process identifies and represents the Conditioned Stimulus (CS, “what”). The latter shapes the amplitude-time properties of the predefined Unconditioned Response (UR), constructing the Conditioned Response (CR, “how”), driven by explicit peripherally triggered error signals that depend on the US (“when”, Lavond, Kim, & Thompson, 1993). A series of robot-based models of the DAC two-phase model of classical conditioning have shown how the

physiologically observed changes in cortical representations of the CS are required in order to effectively drive the procedural learning by the cerebellum (Inderbitzin, Herreros-alonso, & Verschure, 2010; Giovanni Maffei, Santos-Pata, Marcos, Sánchez-Fibla, & Verschure, 2015). These models also revealed a new contextual component to error-driven motor learning. Ever since Pavlov, the acquired CS/CR association was believed to replace the innate US/UR, an idea still dominant in motor learning, defined by the classical feedback error learning model (Kawato, Furukawa, & Suzuki, 1987). However, in a series of robot-based experiments, it was shown that the overall response was a compound comprising both the CR and UR even after reaching asymptotic levels of learning (Herreros & Verschure, 2013; Herreros, Maffei, Brandi, Sanchez-Fibla, & Verschure, 2013). Hence, rather than replacing the peripheral error-driven UR, the CS generates an acquired predictive error signal, that is shifted in time or *counterfactual error*, that drives the reactive layer feedback system linking the US to the UR, reshaping its amplitude time course informed by forward models (Maffei, Herreros, Sanchez-Fibla, Friston, & Verschure, 2017). This explanation accounts for several anatomical and physiological results, including the observation that during eyeblink conditioning, physiological traces of error signals were found both preceding and co-occurring with the US once learning reached asymptotic levels (Ten Brinke et al., 2015). Interestingly, a direct recurrent pathway exists between the frontal cortex and the cerebellum in the mammalian brain. Cerebellar signals are projected to the forebrain via the thalamus, while the forebrain in turn interfaces to the inferior olive climbing fiber inputs via the mesodiencephalic junction (De Zeeuw, Hoebeek, & Schonewille, 2008). This substrate allows information from advanced task models and cognitive control, represented in the frontal cortex, to be projected onto the error processing structures targeting cerebellar circuits, inferior olive, modulating procedural learning.

A recent study in chronic cannabis users explicitly tested the role of counterfactual error, and demonstrated the clinical relevance of this interaction between adaptive and contextual control where a distinct disruption between rule-based deliberative cognitive control and procedural learning was revealed (Herreros et al., 2019). In the context of the DAC framework, this suggests that in cannabis users, rule-based counterfactual error signals are less effective in modulating the cerebellum, leading to enhanced procedural learning at the expense of diminished rule-based cognitive control. We note that this imbalance between cognitive control and procedural learning is also coupled through the environment and is critically linked to the relationship between processes of error monitoring and overt performance. The concept of counterfactual error as used in this example points to the multi-level nature of internal and external feedback loops that must be considered when diagnosing and treating the effects of addiction. Indeed, suboptimal error-monitoring has also been related to (cocaine) addiction (e.g., Bolla et al., 2004; Moeller et al., 2016).

The DAC architecture is providing a concrete model for the biased decision-making interpretation of addiction. Specifically, the contextual layer of DAC has shown to be Bayesian optimal in decision making tasks based on the organization of action selection around the four factors of perceptual evidence, memory bias, value and goals (Verschure & Althaus, 2003). Neuropathology and drug use can alter the relative contribution of these four factors to action selection, reshaping the trade-off boundaries, the control architecture optimizes. In other words, as opposed to ascribing irrational behavior to an addict, the decision-making process of the contextual layer would operate following the normal principles of Bayesian optimality and integration, but the goal and value systems have changed in their specification of objectives and utility, respectively. Hence, from the perspective of the agent, choice behavior may be fully rational during addiction (Kopetz & Orehek, 2015). This will only change when goals and values change, which can be achieved spontaneously (for example after an impactful experience) or aided by motivational interviewing (Miller, 2000). The four factor decision-making model of DAC also provides a further refinement to the

standard drift diffusion models of decision making and the dominant role the latter attribute to perceptual evidence. Indeed, the elegant explanation drift diffusion models provide of macaque performance and neurophysiology becomes invalid in case tasks become more complex as in case of countermanding where the subject had to withhold an initiated action upon receiving a stop signal (Marcos et al., 2013). A computational model has shown that a key factor underlying the competitive neurodynamics driving decision making is performance monitoring (ibid). In particular, biased choice is dynamically regulated by the active monitoring of performance by the agent itself where decision thresholds and gains depend on errors and success. Changes in monitoring, rather than perceptual evidence, better accounts for biased choice as observed in addiction.

The DAC architecture also provides a different perspective on the relationship between deliberation and habit-driven behaviors. As noted above, in neuroscience, this dichotomy is currently cast in terms of the distinction between model-free and model-based (MF-MB) solutions to the problem of choice (Daw et al., 2005). These solutions each stand at different extremes of the robustness versus flexibility trade-off and are consistent with the dilemma of habit versus choice we encountered in the addiction literature, an example of a dichotomous interpretation of adaptive behavior (Hommel, 2019). Notably, the original proponents of this proposal are either backtracking from the MF-MB distinction and posit a single model-based stage (Daw, 2018) or question the validity of model-free interpretations of simple associative learning (Dayan & Berridge, 2014). In the latter case, the key observation is that a value reversal of a CS (i.e. saltiness of a lever), is either aversive or appetitive depending on the internal motivational state of the animal, and indeed is shown to change after salt deprivation. The observation that drive states modulate the value of a stimulus echoes classical models of conditioning (Hull, 1952). In the MF-MB distinction, the latter can only be achieved by falling back on performing computations on a model of states and outcomes, which contradicts earlier definitions in which classical conditioning was defined as MF. Yet, a multilayer predictive architecture like DAC shows how this dilemma can be resolved. The Adaptive Layer represents MF associations between states and outcomes, but these conjunctive representations in turn become the primitives for the MB Contextual Layer which combines them into expanded goal-associated sequences that in turn are weighted regarding their utility given the current drive state (Duff, Sanchez Fibla, & Verschure, 2011; Verschure & Althaus, 2003). Whether the system relies more on MF or MB mechanisms now depends on the specific task constraints, on the layer of control invoked and the internal motivational state of the agent which transforms associated value into utility. Hence, the MF and MB apposition is illusory in the sense that they are functional realizations of an integrated and dynamic embodied control system which is resolving distinct trade-offs.

4. Neurorehabilitation

The goal of rehabilitation in addiction is two-fold. First, the addiction itself has to be addressed and overcome, reducing the probability of relapse. A second goal can be to address amelioration of the functional deficits incurred by the addiction and its negative effects on the quality of life and health. The challenge of effective rehabilitation in addiction is to enhance the impact of treatment, based on a standard set of principles underlying diagnostics and intervention. We can add to these foundational challenges the more pragmatic one of finding solutions that facilitate scaling to large numbers of patients, including, ideally, those at home, as the large majority of people suffering from addictions are not treated (Alonso et al., 2004). To answer these challenges in the domain of addiction, we can also build on results obtained with the system-oriented neurorehabilitation approach developed within the context of the DAC theory. We first address the current state of affairs in addiction neurorehabilitation and then return to system-oriented neurorehabilitation.

Current (neuro)cognitive training or neurorehabilitation efforts in addiction can be categorized into two broad classes: training of (sub-optimal) general functions, such as working memory (WM), and retraining of abnormally strong cognitive-motivational processes ("cognitive biases") triggered by addiction-related cues, known as cognitive bias modification or CBM (Wiers, 2018). Regarding the first class of neurorehabilitation, there is evidence that training can improve the targeted general function (typically WM), and generalization to other relevant functions has been reported, such as delay discounting (Bickel, Yi, Landes, Hill, & Baxter, 2011), and future episodic thinking (Snider et al., 2018). However, there is little evidence that general cognitive training helps people to control their addictive behaviors better (reduced drinking in one subgroup, Houben, Wiers, & Jansen, 2011; no evidence: Bickel et al., 2011; Snider et al., 2018; Wanmaker et al., 2018). These negative results do not make this type of training useless in a clinical context: for example, the enhanced ability for future episodic thinking can be beneficial when addressed in a therapeutic setting for making post-addiction plans more concrete. Furthermore, feedback about progress in these functions can in turn be motivating to work towards recovery (Bates, Buckman, & Nguyen, 2013). Note that this literature usually assumes that suboptimal cognitive functions are the result of the addiction ("broken brain"). However, evidence for this assumption is typically lacking: often there are no baseline measures from before the addiction (Schulte et al., 2014), and relatively weak cognitive control is one of the most consistent predictors of later addiction (Nigg, 2000). As a consequence, while improvement can be expected and can be motivating, it is questionable whether "normal" performance should be a norm in the specification of intervention outcomes, another variant of the "true" recovery assumption. Moreover, the absence of normative performance does not imply a chronic brain disease when another etiological factor could be at play. Regarding the effects of binge drinking in youth, a recent systematic review and meta-analysis reported mostly small effects on different neuropsychological outcome measures, with low to very low certainty (Lees et al., 2019). Further, a recent mega-analysis on the effects of substance use on behavioral inhibition also reported very limited effects (Liu et al., 2019). The absence of evidence does not mean that there are no detrimental effects of early substance use on brain development, but we would argue that these modest findings should be understood as a factor which is likely to bias future (drug-related) decision-making rather than an indicator of a developing chronic brain disease.

In the second type of cognitive training, CBM, different cognitive biases can be targeted: biases in attention, action tendencies, and memory (Wiers, Gladwin, Hofmann, Salemink, & Ridderinkhof, 2013). All of these biases are triggered by contextual stimuli that are related to the addiction (e.g., a location, object, time of day). The basic idea behind CBM is that these stimuli trigger appetitive reactions (capture attention, trigger memories of pleasant effects and action tendencies to approach the cue) and that these reactions can be systematically retrained. When evaluating the evidence supporting CBM, it is crucial to distinguish between proof-of-principle studies and randomized controlled trials (RCTs, Sheeran, Klein, & Rothman, 2017; Wiers, Boffo, & Field, 2018), which one meta-analysis failed to do (Cristea, Kok, & Cuijpers, 2016). In proof-of-principle studies, a cognitive bias is manipulated (sometimes temporarily increased) in non-addicted subjects with the goal to test its hypothetical causal effect on behavior. In contrast, in clinical RCTs participants consist of patients or volunteers who wish to change their addictive behaviors (Wiers et al., 2018). Proof-of-principle studies typically report short-lived effects, in case the bias is successfully manipulated (Allom, Mullan, & Hagger, 2016; Wiers et al., 2018). Clinical RCTs have shown a small but consistent additive effect on long-term treatment outcomes when combined with standard treatment. For instance, in alcohol use disorders, relapse one year after treatment discharge was found to decrease by approximately 10% across multiple large RCTs (Eberl et al., 2013; Rinck, Wiers, Becker, & Lindenmeyer, 2018; Wiers, Eberl, Rinck, Becker, & Lindenmeyer,

2011). This demonstrates that although choice is biased in addiction, targeted training may help to neutralize this bias. A recent comprehensive Bayesian *meta-analysis*, exclusively including clinical studies, has confirmed a small effect on bias and abstinence, while calling for more studies (Boffo et al., 2019). This effect may be enhanced by optimizing timing: a small study found strong effects of CBM when delivered during detox (Manning et al., 2016). Further, CBM-training may be enhanced from the theoretical perspective of effects on automatic inferences rather than associations (Wiers, Van Dessel et al., 2020). For example, adding consequences to actions could increase the goal-directed nature of training and thereby increase its effectiveness (Van Dessel, Hughes, & De Houwer, 2019). Further, alternative actions can be personalized, especially for other addictions where no generic alternative is present (such as non-alcoholic drinks in AUD) (Kopetz, MacPherson, Mitchell, Houston-Ludlam, & Wiers, 2017). In line with a systems-based approach to neurorehabilitation, these new varieties of CBM address different (personalized) aspects of the situated agent: environmental risk-situations, personally relevant alternative choices and their effects on different (personally relevant) outcomes (Wiers, Van Dessel et al., 2020, Wiers, Van Gaal et al., 2020). However, all these suggestions await further clinical testing in well-designed RCTs.

Starting from the consideration that an adequate theory of mind and brain should provide traction in clinical applications, core principles of the DAC theory have been translated to the treatment of motor, affective and cognitive deficits in several neuro-pathologies. Key features include the organization of training around integrated tasks and goals, as proposed by the contextual layer, to include ecologically realistic sensorimotor contingencies, as defined by the adaptive layer, to individualize task difficulty optimizing effort, fatigue, and motivation as defined by the reactive layer. In addition, all training scenarios are presented in virtual reality (VR), in a first-person embodied perspective, following the DAC predicates of embodiment and situatedness of the somatic layer. Thus, in this approach the recovering brain is asked to take ownership of a virtual body with which to perform tasks in a virtual environment. In general, DAC proposes that the most effective way to retrain a recovering brain is by projecting it in an embodied form into a task-space with well-defined sensorimotor contingencies, goals and feedback.

The application of DAC-based clinical interventions (Rehabilitation Gaming System, RGS), can be used for system-oriented neurorehabilitation. For example, functional rehabilitation after stroke, a pathology often considered one of the most unambiguous examples of a “broken brain”, has been successfully improved with system-oriented neurorehabilitation. As an example we can consider the popular intervention of constraint induced movement therapy, where the use of the paretic arm is promoted through the immobilization of the healthy one (Taub, Uswatte, & Pidikiti, 1999). Recent *meta-analyses* have questioned its effectivity as compared to standard treatment (Kwakkel et al., 2016; Kwakkel, Veerbeek, van Wegen, & Wolf, 2015). Based on the DAC-derived counterfactual error hypothesis, an alternative was proposed where visual error feedback in VR was reduced through intention-compatible enhancement of reaching actions. This intervention restored symmetric arm use in a group of chronic stroke patients in a single session with 100 enhanced trials (Ballester et al., 2015). This is one example of the translation of the principles of DAC-derived system-oriented neurorehabilitation, for which large-scale clinical trials have now shown its effectivity (*meta-analysis*: Rubio et al., 2019).

Building on these examples, we foresee a convergence of principles of diagnostics and training across different neuropathologies that place deficits and their ramifications in the context of the complex linking of the different levels of organization of humans from their genetic and neuronal substrate to their psychological organization and behavioral expression including their specific socio-cultural embedding. As a concrete example we can consider the commonalities between the principles underlying CBM and intentional compatible movement enhancement based on counterfactual error deployed in RGS. In both cases, these interventions link to the fundamental learning paradigm of

classical conditioning as modelled by DAC. The counterfactual error hypothesis predicts that addicted patients will face changes to their error monitoring and processing, and a resulting modulation of choice behavior as demonstrated by the aforementioned study on cannabis users (Herreros et al., 2019). In the stroke rehabilitation example, this principle was used to modulate the controllability of the paretic limb, and CBM for addiction can further elaborate this principle towards the processing of sensory cues and error monitoring. In addition, by elaborating these principles underlying pathological behavior, an additional therapeutic channel is created that will allow patients to develop the *meta-cognition* needed to willfully address the challenges they face as a result of their addiction. Hence, the question is not which of two processes exclusively dominates performance (as in dual-process and MF-Mb models), but rather how the relative contributions of multiple factors to the choice generation process are modulated by a multi-scale embodied control architecture in the face a specific trade-offs, and how this can be influenced in treatment and neurorehabilitation.

5. Conclusions

Addictions are among the most frequent and costly of all mental and brain disorders (Effertz & Mann, 2013). There is no doubt that drugs of abuse and long-term addictions have an impact on the brain. However, we argue that these effects should be understood and treated from a systems-perspective, in line with the multi-layered hierarchical organization of the (human) brain, in which goals and meaning are essential at the highest contextual level, with direct and indirect impact at lower levels of organization. This implies that the human mind/brain, addicted or not, should be considered as a goal-directed dynamic complex system, and its idiosyncratic but rational goal-directedness should guide neurorehabilitation. This perspective differs from the dominant perspective of the chronic brain disease model of addiction and its associated interventions aimed at repairing the broken-brain. *System-oriented neurorehabilitation* takes a dynamic and adaptive hierarchical embodied and situated brain as its starting point. It helps patients to recover by systematic training, addressing multiple levels of understanding, experience, and control. System-oriented neurorehabilitation starts with goals to change (contextual level), the history of experience (adaptive level), which defines individualized and personally relevant training, addressing multiple levels of control (addressing the adaptive and reactive levels). This approach has already shown impact in the treatment of stroke and related problems, early results in cannabis diagnostics and provides a framework in which to elaborate novel approaches such as CBM that can be used to further develop neurorehabilitation for addictions in an efficient and scalable form. Indeed, the example of counterfactual error processing derived from the DAC framework has shown relevance across domains of neuropathology, including addictions. In this way, advances in clinical applications and our fundamental understanding of mind and brain will progress in a complementary and synchronized effort.

Declaration of Competing Interests

The authors declare the following potential conflicts of interest concerning the research, authorship, and publication of this article: PFMJV is the founder and interim CEO of Eodyne s.l. which brings scientifically validated neurorehabilitation technology to society.

Acknowledgements

The authors like to thank Tom Ruiter, and the anonymous reviewers for helpful comments on our manuscript. The work reported in this manuscript is supported by EC project Virtual Brain Cloud (EC H2020 82642), the cognitive Rehabilitation Gaming System (ERC PoC, 840052) and RGS@Home (EIT Health, 20689) to PV.

References

- Ajzen, I., & Kruglanski, A. W. (2019). Reasoned Action in the Service of Goal Pursuit. *Psychological Review*, 126(5), 774–786. <https://doi.org/10.1037/rev0000155>.
- Allom, V., Mullan, B., & Hagger, M. (2016). Does inhibitory control training improve health behaviour? A meta-analysis. *Health Psychology Review*, 10(2), 168–186.
- Alonso, J., Angermeyer, M. C., Bernert, S., Bruffaerts, R., Brugha, T. S., Bryson, H., Girolamo, G., Graaf, R., Demyttenaere, K., Gasquet, I., Haro, J. M., Katz, S. J., Kessler, R. C., Kovess, V., Lepine, J. P., Ormel, J., Polidori, G., Russo, L. J., Vilagut, G., Almansa, J., Arbabzadeh-Bouchez, S., Autonell, J., Bernal, M., Buist-Bouwman, M. A., Codony, M., Domingo-Salvany, A., Ferrer, M., Joo, S. S., Martinez-Alonso, M., Matschinger, H., Mazzi, F., Morgan, Z., Morosini, P., Palacin, C., Romera, B., Taub, N., & Vollebergh, W. A. M. (2004). Prevalence of mental disorders in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica*, 109(s420), 21–27.
- Anderson, B. A. (2016). What is abnormal about addiction-related attentional biases? *Drug and Alcohol Dependence*, 167, 8–14.
- Badiani, A., Belin, D., Epstein, D., Calu, D., & Shaham, Y. (2011). Opiate versus psychostimulant addiction: The differences do matter. *Nature Reviews Neuroscience*, 12(11), 685–700.
- Ballester, B. R., Nirme, J., Duarte, E., Cuxart, A., Rodriguez, S., Verschure, P., & Duff, A. (2015). The visual amplification of goal-oriented movements counteracts acquired non-use in hemiparetic stroke patients. *Journal of NeuroEngineering and Rehabilitation*, 12(1), <https://doi.org/10.1186/s12984-015-0039-z>.
- Barnett, A. I., Hall, W., Fry, C. L., Dilkes-Frayne, E., & Carter, A. (2018). Drug and alcohol treatment providers' views about the disease model of addiction and its impact on clinical practice: A systematic review: Treatment providers' views on the DMA. *Drug and Alcohol Review*, 37(6), 697–720.
- Bates, M. E., Buckman, J. F., & Nguyen, T. T. (2013). A role for cognitive rehabilitation in increasing the effectiveness of treatment for alcohol use disorders. *Neuropsychology Review*, 23(1), 27–47.
- Baumeister, R. F. (2017). Addiction, cigarette smoking, and voluntary control of action: Do cigarette smokers lose their free will? *Addictive Behaviors Reports*, 5, 67–84.
- Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience*, 8(11), 1458–1463. <https://doi.org/10.1038/Nn1584>.
- Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-Control as Value-Based choice. *Current Directions in Psychological Science*, 26(5), 422–428.
- Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, 26(9), 507–513.
- Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, 71(8), 670–679. <https://doi.org/10.1037/amp0000059>.
- Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F., & Baxter, C. (2011). Remember the future: Working memory training decreases delay discounting among stimulant addicts. *Biological Psychiatry*, 69(3), 260–265.
- Boffo, M., Zerhouni, O., Gronau, Q. F., van Beek, R. J. J., Nikolaou, K., Marsman, M., & Wiers, R. W. (2019). Cognitive bias modification for behavior change in alcohol and smoking addiction: Bayesian meta-analysis of individual participant data. *Neuropsychology Review*, 29(1), 52–78.
- Bolla, K., Ernst, M., Kiehl, K., Mouratidis, M., Eldred, D., Contoreggi, C., Matochik, J., Kurian, V., Cadet, J., Kimes, A., Funderburk, F., & London, E. (2004). Prefrontal cortical dysfunction in abstinent cocaine abusers. *JNP*, 16(4), 456–464.
- Childress, A. R., Ehrman, R. N., Wang, Z., Li, Y., Sciortino, N., Hakun, J., & O'Brien, C. P. (2008). Prelude to passion: Limbic activation by "unseen" drug and sexual cues. *PLoS ONE*, 3(1), <https://doi.org/10.1371/journal.pone.0001506>.
- Cristea, I. A., Kok, R. N., & Cuijpers, P. (2016). The efficiency of cognitive bias modification interventions for addictions: A meta-analysis. *PLoS ONE*, 11(9), Article e0162226. <https://doi.org/10.1371/journal.pone.0162226>.
- Daw, N. D. (2018). Are we of two minds? *Nature Neuroscience*, 21(11), 1497–1499.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711.
- Dayan, P., & Berridge, K. C. (2014). Model-based and model-free pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, 14(2), 473–492.
- De Houwer, J. (2019). On How definitions of habits can complicate habit research. *Frontiers in Psychology*, 10, 1–9. <https://doi.org/10.3389/fpsyg.2019.02642>.
- de Wit, H., Epstein, D. H., & Preston, K. L. (2018). Does human language limit translatability of clinical and preclinical addiction research? *Neuropsychopharmacology*, 43(10), 1985–1988.
- De Zeeuw, C. I., Hoebeek, F. E., & Schonewille, M. (2008). Causes and consequences of oscillations in the cerebellar cortex. *Neuron*, 58(5), 655–658.
- Doyle, J. C., & Csete, M. (2011). Architecture, constraints, and behavior. *Proceedings of the National Academy of Sciences*, 108(Supplement 3), 15624–15630.
- Duff, A., Sanchez Fibla, M., & Verschure, P. F. M. J. (2011). A biologically based model for the integration of sensory-motor contingencies in rules and plans: A prefrontal cortex based extension of the distributed adaptive control architecture. *Brain Research Bulletin*, 85(5), 289–304.
- Eberl, C., Wiers, R. W., Pawelczak, S., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2013). Approach bias modification in alcohol dependence: Do clinical effects replicate and for whom does it work best? *Developmental Cognitive Neuroscience*, 4, 38–51.
- Effertz, T., & Mann, K. (2013). The burden and cost of disorders of the brain in Europe with the inclusion of harmful alcohol use and nicotine addiction. *European Neuropsychopharmacology*, 23(7), 742–748.
- Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, 8(11), 1481–1489.
- Everitt, B. J., & Robbins, T. W. (2016). Drug addiction: Updating actions to habits to compulsions ten years on. *Annual Review of Psychology*, 67(1), 23–50.
- Fenton, T., & Wiers, R. W. (2017). Free will, black swans and addiction. *Neuroethics*, 10(1), 157–165.
- Field, M., Heather, N., Murphy, J. G., Stafford, T., Tucker, J. A., & Witkiewitz, K. (2020). Recovery from addiction: Behavioral economics and value-based decision making. *Psychology of Addictive Behaviors*, 34(1), 182–193. <https://doi.org/10.1037/adb0000518>.
- Field, M., & Kersbergen, I. (2020). Are animal models of addiction useful? *Addiction*, 115(1), 6–12.
- Franken, I. H. A. (2003). Drug craving and addiction: Integrating psychological and neuropsychopharmacological approaches. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 27(4), 563–579.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7), 293–301.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: Conceptual, empirical, and metatheoretical issues: Reply to Albarracín, Hart, and McCulloch (2006), Kruglanski and Dechesne (2006), and Petty and Briñol (2006). *Psychological Bulletin*, 132(5), 745–750. <https://doi.org/10.1037/0033-2909.132.5.745>.
- Gladwin, T. E., Figner, B., Crone, E. A., & Wiers, R. W. (2011). Addiction, adolescence, and the integration of control and motivation. *Developmental Cognitive Neuroscience*, 1(4), 364–376.
- Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature Reviews Neuroscience*, 12(11), 652–669.
- Gorelick, D. A., Zangen, A., & George, M. S. (2014). Transcranial magnetic stimulation in the treatment of substance addiction. *Annals of the New York Academy of Sciences*, 1327(1), 79–93. <https://doi.org/10.1111/nyas.12479>.
- Hall, W., Carter, A., & Forlini, C. (2015). Brain disease model of addiction: Misplaced priorities? *The Lancet Psychiatry*, 2(10), 867. [https://doi.org/10.1016/S2215-0366\(15\)00417-4](https://doi.org/10.1016/S2215-0366(15)00417-4).
- Hart, C. (2013). *High price: Drugs, neuroscience and discovering myself*. London: Penguin.
- Heather, N. (2017). Is the concept of compulsion useful in the explanation or description of addictive behaviour and experience? *Addictive Behaviors Reports*, 6, 15–38.
- Heather, N., Best, D., Kawalek, A., Field, M., Lewis, M., Rotgers, F., Wiers, R. W., & Heim, D. (2018). Challenging the brain disease model of addiction: European launch of the addiction theory network. *Addiction Research & Theory*, 26(4), 249–255.
- Heinz, A., Kiefer, F., Smolka, M. N., Endrass, T., Beste, C., Beck, A., Liu, S., Genauck, A., Romund, L., Banaschewski, T., Birmaher, F., Deserno, L., Dolan, R. J., Durstewitz, D., Ebner-Priemer, U., Flor, H., Hansson, A. C., Heim, C., Hermann, D., Kiesel, S., Kirsch, P., Kirschbaum, C., Koppe, G., Marxen, M., Meyer-Lindenberg, A., Nagel, W. E., Noori, H. R., Pilhatsch, M., Priller, J., Rietschel, M., Romanczuk-Seiferth, N., Schlagenhauf, F., Sommer, W. H., Stallkamp, J., Ströhle, A., Stock, A.-K., Winterer, G., Winter, C., Walter, H., Witt, S., Vollstädt-Klein, S., Rapp, M. A., Tost, H., & Spanagel, R. (2020). Addiction Research Consortium: Losing and regaining control over drug intake (ReCoDe)—From trajectories to mechanisms and interventions. *Addiction Biology*, 25(2), <https://doi.org/10.1111/adb.v25.210.1111/adb.12866>.
- Herreros, I., & Verschure, P. F. M. J. (2013). Nucleo-olivary inhibition balances the interaction between the reactive and adaptive layers in motor control. *Neural Networks*, 47, 64–71.
- Herreros, Ivan, Maffei, G., Brandi, S., Sanchez-Fibla, M., & Verschure, P. F. M. J. (2013). Speed generalization capabilities of a cerebellar model on a rapid navigation task. Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference. <https://doi.org/10.1109/IROS.2013.6696377>.
- Herreros, Ivan, Miquel, L., Blithikioti, C., Nuño, L., Rubio Ballester, B., Grechuta, K., & Verschure, P. (2019). Motor adaptation impairment in chronic cannabis users assessed by a visuomotor rotation task. *Journal of Clinical Medicine*, 8(7), 1049. <https://doi.org/10.3390/jcm8071049>.
- Heyman, G. M. (2010). *Addiction: A disorder of choice*. Harvard University Press.
- Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: Translational critique of habit and compulsion theory. *Neuropsychopharmacology*, 45(5), 720–735.
- Hommel, B., & Wiers, R. W. (2017). Towards a unitary approach to human action control. *Trends in Cognitive Sciences*, 21(12), 940–949.
- Hommel, B. (2019). Binary theorizing does not account for action control. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02542>.
- Houben, K., Wiers, R. W., & Jansen, A. (2011). Getting a grip on drinking behavior: Training working memory to reduce alcohol abuse. *Psychological Science*, 22(7), 968–975.
- Hull, C. L. (1952). *A behavior system: An introduction to behavior theory concerning the individual organism*. New Haven: Yale University Press.
- Hyman, S. E. (2013). Psychiatric drug development: Diagnosing a crisis. Retrieved from *Cerebrum: The Dana Forum on Brain Science*, 5, 5. <http://www.ncbi.nlm.nih.gov/pubmed/23720708>0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3662213.
- Inderbitzin, M., Herreros-alonso, I., & Verschure, P. F. M. J. (2010). An integrated computational model of the two phase theory of conditioning. *IEEE World Congress of Computational Intelligence*.
- Ingjaldsson, J. T., Thayer, J. F., & Laberg, J. C. (2003). Craving for alcohol and pre-attentive processing of alcohol stimuli. *International Journal of Psychophysiology*, 49(1), 29–39.
- Jentsch, J. D., & Taylor, J. R. (1999). Impulsivity resulting from frontostriatal dysfunction

- in drug abuse: Implications for the control of behavior by reward-related stimuli. *Psychopharmacology (Berl)*, 146(4), 373–390. <https://doi.org/10.1007/PL00005483>.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>.
- Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, 57(3), 169–185. <https://doi.org/10.1007/BF00364149>.
- Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, 4(6), 533–550.
- Kirschner, M., & Gerhart, J. C. (2008). *The plausibility of life: Resolving Darwin's dilemma*. Yale University Press.
- Koob, G. F., & Le Moal, M. (2008). Addiction and the brain antireward system. *Annual Review of Psychology*, 59(1), 29–53.
- Kopetz, C., MacPherson, L., Mitchell, A. D., Houston-Ludlam, A. N., & Wiers, R. W. (2017). A novel training approach to activate alternative behaviors for smoking in depressed smokers. *Experimental and Clinical Psychopharmacology*, 25(1), 50–60. <https://doi.org/10.1037/pha0000108>.
- Kopetz, C., & Orehhek, E. (2015). When the end justifies the means: Self-defeating behaviors as “rational” and “successful” self-regulation. *Current Directions in Psychological Science*, 24(5), 386–391.
- Kruglanski, A. W., & Szumowska, E. (2020). Habitual behavior is goal driven. *Perspectives on Psychological Science*, in press.
- Kwakkel, G., Veerbeek, J. M., van Wegen, E. E. H., & Wolf, S. L. (2015). Constraint-induced movement therapy after stroke. *The Lancet Neurology*, 14(2), 224–234. [https://doi.org/10.1016/S1474-4422\(14\)70160-7](https://doi.org/10.1016/S1474-4422(14)70160-7).
- Kwakkel, G., Winters, C., van Wegen, E. E. H., Nijland, R. H. M., van Kuijk, A. A. A., Visser-Meily, A., de Groot, J., de Vlugt, E., Arendzen, J. H., Geurts, A. C. H., & Meskers, C. G. M. (2016). Effects of unilateral upper limb training in two distinct prognostic groups early after stroke: The EXPLICIT-stroke randomized clinical trial. *Neurorehabilitation*, 30(9), 804–816.
- Lavond, D. G., Kim, J. J., & Thompson, R. F. (1993). Mammalian brain substrates of aversive classical conditioning. *Annual Review of Psychology*, 44(1), 317–342.
- Le Pelley, M. E., Pearson, D., Griffiths, O., & Beesley, T. (2015). When goals conflict with values: Counterproductive attentional and oculomotor capture by reward-related stimuli. *Journal of Experimental Psychology: General*, 144(1), 158–171. <https://doi.org/10.1037/xge0000037>.
- Lees, B., Mewton, L., Stapinski, L. A., Squeglia, L. M., Rae, C. D., & Teesson, M. (2019). Neurobiological and cognitive profile of young binge drinkers: A systematic review and meta-analysis. *Neuropsychology Review*, 29(3), 357–385. <https://doi.org/10.1007/s11065-019-09411-w>.
- Leshner, A. I. (1997). Addiction is a brain disease, and it matters. *Sciences, New Series*, 278(5335), 45–47. <https://doi.org/10.1126/science.278.5335.45>.
- Lewis, M. (2015). The Biology of Desire: Why Addiction Is Not a Disease. Retrieved from <https://books.google.com/books?hl=en&lr=&id=CRtpCQAAQBAJ&pgis=1>.
- Longo, D. L., & Lewis, M. (2018). Brain change in addiction as learning, not disease. *New England Journal of Medicine*, 379(16), 1551–1560.
- Lin, H., Saunders, B., Fries, M., Evans, N. J., & Inzlicht, M. (2020). Strong Effort manipulations reduce response caution: A preregistered reinvention of the ego-depletion paradigm. *Psychological Science*, 31(5), 531–547.
- Liu, Y., van den Wildenberg, W. P. M., de Graaf, Y., Ames, S. L., Baldacchino, A., Bø, R., Cadaveira, F., Campanella, S., Christiansen, P., Claus, E. D., Colzato, L. S., Filbey, F. M., Foxe, J. J., Garavan, H., Hendershot, C. S., Hester, R., Jester, J. M., Karoly, H. C., Kräplin, A., Kreusch, F., Landrø, N. I., Littell, M., Loeber, S., London, E. D., López-Caneda, E., Lubman, D. I., Luijten, M., Marczinski, C. A., Metrick, J., Montgomery, C., Papachristou, H., Mi Park, S., Paz, A. L., Petit, G., Prisciandaro, J. J., Quednow, B. B., Ray, L. A., Roberts, C. A., Roberts, G. M. P., de Ruiter, M. B., Rupp, C. I., Steele, V. R., Sun, D., Takagi, M., Tapert, S. F., van Holst, R. J., Verdejo-Garcia, A., Vonmoos, M., Wojnar, M., Yao, Y., Yücel, M., Zack, M., Zucker, R. A., Huizenga, H. M., & Wiers, R. W. (2019). Is (poly-) substance use associated with impaired inhibitory control? A mega-analysis controlling for confounders. *Neuroscience & Biobehavioral Reviews*, 105, 288–304. <https://doi.org/10.1016/j.neubiorev.2019.07.006>.
- Loijen, A., Rinck, M., Walvoort, S. J. W., Kessels, R. P. C., Becker, E. S., & Egger, J. I. M. (2018). Modification of automatic alcohol-approach tendencies in alcohol-dependent patients with mild or major neurocognitive disorder. *Alcoholism, Clinical and Experimental Research*, 42(1), 153–161.
- Loughman, A., & Haslam, N. (2018). Neuroscientific explanations and the stigma of mental disorder: A meta-analytic study. *Cognitive Research: Principles and Implications*, 3(1), 1–12. <https://doi.org/10.1186/s41235-018-0136-1>.
- Maffei, G., Herreros, L., Sanchez-Fibla, M., Friston, K. J., & Verschure, P. F. M. J. (2017). The perceptual shaping of anticipatory actions. *Proceedings of the Royal Society*, 284(1869), 20171780. <https://doi.org/10.1098/rspb.2017.1780>.
- Maffei, G., Santos-Pata, D., Marcos, E., Sánchez-Fibla, M., & Verschure, P. F. M. J. (2015). An embodied biologically constrained model of foraging: From classical and operant conditioning to adaptive real-world behavior in DAC-X. *Neural Networks*, 72, 88–108. <https://doi.org/10.1016/j.neunet.2015.10.004>.
- Manning, V., Staiger, P. K., Hall, K., Garfield, J. B. B., Flaks, G., Leung, D., Hughes, L. K., Lum, J. A. G., Lubman, D. I., & Verdejo-Garcia, A. (2016). Cognitive bias modification training during inpatient alcohol detoxification reduces early relapse: A randomized controlled trial. *Alcoholism, Clinical and Experimental Research*, 40(9), 2011–2019.
- Marcos, E., Pani, P., Brunamonti, E., Deco, G., Ferraina, S., & Verschure, P. (2013). Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance. *Neuron*, 78(2), 249–255. <https://doi.org/10.1016/j.neuron.2013.02.006>.
- Massaro, D. W. (1997). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- May, J., Kavanagh, D. J., & Andrade, J. (2015). The Elaborated Intrusion Theory of desire: A 10-year retrospective and implications for addiction treatments. *Addictive Behaviors*, 44, 29–34.
- Medina, J. F., Christopher Repa, J., Mauk, M. D., & LeDoux, J. E. (2002). Parallels between cerebellum- and amygdala-dependent conditioning. *Nature Reviews Neuroscience*, 3(2), 122–131.
- Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, 22(4), 280–293. <https://doi.org/10.1016/j.tics.2018.02.001>.
- Miller, S., & Konorski, J. (1928). Sur une forme particulière des reflexes conditionnels. *Comptes Rendus Des Seances De La Societ'e Polonaise de Biologie*, 49, 1155–1157.
- Miller, W. R. (2000). Rediscovering fire: Small interventions, large effects. *Psychology of Addictive Behaviors*, 14(1), 6.
- Moeller, S. J., Fleming, S. M., Gan, G., Zilverstand, A., Malaker, P., d'Oleire Uquillas, F., Schneider, K. E., Preston-Campbell, R. N., Parvaz, M. A., Maloney, T., Alia-Klein, N., & Goldstein, R. Z. (2016). Metacognitive impairment in active cocaine use disorder is associated with individual differences in brain structure. *European Neuropsychopharmacology*, 26(4), 653–662. <https://doi.org/10.1016/j.euroneuro.2016.02.009>.
- Moors, A., Boddez, Y., & De Houwer, J. (2017). The power of goal-directed processes in the causation of emotional and other actions. *Emotion Review*, 9(4), 310–318. <https://doi.org/10.1177/1754073916669595>.
- Most, S. B., Smith, S. D., Cooter, A. B., Levy, B. N., & Zald, D. H. (2007). The naked truth: Positive, arousing distractors impair rapid target perception. *Cognition & Emotion*, 21(5), 964–981. <https://doi.org/10.1080/02699930600595340>.
- Nesse, R. M., & Berridge, K. C. (1997). Psychoactive drug use in evolutionary psychology. *Science*, 278(5335), 63–66.
- Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.
- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, 126(2), 220.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning. *Classical Conditioning II Current Research and Theory*, 497. <https://doi.org/10.1101/gr.110528.110>.
- Rinck, M., Wiers, R. W., Becker, E. S., & Lindenmeyer, J. (2018). Relapse prevention in abstinent alcoholics by Cognitive Bias Modification: Clinical effects of combining approach bias modification and attention bias modification. *Journal of Consulting and Clinical Psychology*, 86(12), 1005–1016.
- Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews*, 18(3), 247–291. [https://doi.org/10.1016/0165-0173\(93\)90013-P](https://doi.org/10.1016/0165-0173(93)90013-P).
- Rubio, B. B., Maier, M., Duff, A., Cameirao, M., Bermudez, S., Duarte, E., ... Verschure, P. (2019). A critical time window for recovery extends beyond one-year post-stroke. *Journal of Neurophysiology*.
- Schoenmakers, T. M., de Bruin, M., Lux, I. F. M., Goertz, A. G., Van Kerkhof, D. H. A. T., & Wiers, R. W. (2010). Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. *Drug and Alcohol Dependence*, 109(1–3), 30–36. <https://doi.org/10.1016/j.drugalcdep.2009.11.022>.
- Schulte, M. H. J., Cousijn, J., den Uyl, T. E., Goudriaan, A. E., van den Brink, W., Veltman, D. J., Schilt, T., & Wiers, R. W. (2014). Recovery of neurocognitive functions following sustained abstinence after substance dependence and implications for treatment. *Clinical Psychology Review*, 34(7), 531–550. <https://doi.org/10.1016/j.cpr.2014.08.002>.
- Segal, G. M. A. (2017). How an addict's power of choice is lost and can be regained. In N. Heather, & G. Segal (Eds.). *Addiction and Choice*. Oxford: OUP.
- Sheeran, P., Klein, W. M. P., & Rothman, A. J. (2017). Health behavior change: moving from observation to intervention. *Annual Review of Psychology*, 68(1), 573–600.
- Snider, S. E., Deshpande, H. U., Lisinski, J. M., Koffarnus, M. N., LaConte, S. M., & Bickel, W. K. (2018). Working memory training improves alcohol users' episodic future thinking: a rate dependent analysis. *Biological Psychiatry Cognitive Neuroscience and Neuroimaging*, 3, 160–167. <https://doi.org/10.1016/j.bpsc.2017.11.002>.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247.
- Stroebe, W., Van Koningsbruggen, G. M., Papies, E. K., & Aarts, H. (2017). Why most dieters fail but some succeed: A goal conflict model of eating behavior. *The Goal Conflict Model of Eating Behaviour: Selected Works of Wolfgang Stroebe*, 120(1), 181–238. <https://doi.org/10.4324/9781315141817>.
- Taub, E., Uswatte, G., & Pidikiti, R. (1999). Constraint-Induced Movement Therapy: a new family of techniques with broad application to physical rehabilitation—a clinical review. 36(3), 237–251.
- Ten Brinke, M. M., Boele, H. J., Spanke, J. K., Potters, J. W., Kornysheva, K., Wulff, P., ... De Zeeuw, C. I. (2015). Evolving models of pavlovian conditioning: Cerebellar cortical dynamics in awake behaving mice. *Cell Reports*, 13(9), 1977–1988.
- Tolman, E. C. (1966). Gestalt and sign-gestalt. In T. E. C. (Ed.), *Behavior and Psychological Man* (pp. 77–93). Berkeley, Ca.: University of California Press.
- Van Dessel, P., Hughes, S., & De Houwer, J. (2019). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personality and Social Psychology Review*, 23(3), 267–284.
- Venniro, M., Zhang, M., Caprioli, D., Hoots, J. K., Golden, S. A., Heins, C., ... Shaham, Y. (2018). Addiction in rat models. *Nature Neuroscience*, 21. <https://doi.org/10.1038/s41593-018-0246-6>.
- Verschure, P. F. M. J. (2012). Distributed adaptive control: A theory of the Mind, Brain, Body Nexus. *Biologically Inspired Cognitive Architectures*, 1, 55–72. <https://doi.org/10.1016/j.bica.2012.04.005>.
- Verschure, P. F. M. J. (2016). Synthetic consciousness: The distributed adaptive control perspective. *Philosophical Transactions of the Royal Society B*, 371(1701), 20150448. <https://doi.org/10.1098/rstb.2015.0448>.

- Categorization, Representations, and the Dynamics of System-Environment Interaction: a case study in autonomous systems* Vol. 2, 210–217.
- Verschure, P. F. M. J., Voegtlin, T., & Douglas, R. J. (2003). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425(6958), 620–624. <https://doi.org/10.1038/nature02024>.
- Verschure, P. F. M. J. (2018). *The architecture of mind and brain*. Living Machines: A Handbook of Research in Biomimetics and Biohybrid Systems 338.
- Verschure, P. F., & Althaus, P. (2003). A real-world rational agent: Unifying old and new AI. *Cognitive Science*, 27(4), 561–590.
- Volkow, N. D. (2015). Addiction is a disease of free will. Retrieved from <https://www.youtube.com/watch?v=X1AEvKWxblE>.
- Volkow, N. D., Koob, G. F., & McLellan, A. T. (2016). Neurobiologic advances from the brain disease model of addiction. *New England Journal of Medicine*, 374(4), 363–371.
- Wanmaker, S., Leijdesdorff, S. M. J., Geraerts, E., van de Wetering, B. J. M., Renkema, P. J., & Franken, I. H. A. (2018). The efficacy of a working memory training in substance use patients: A randomized double-blind placebo-controlled clinical trial. *Journal of Clinical and Experimental Neuropsychology*, 40(5), 473–486.
- Watson, P., Pearson, D., Wiers, R. W., & Le Pelley, M. E. (2019). Prioritizing pleasure and pain: Attentional capture by reward-related and punishment-related stimuli. *Current Opinion in Behavioral Sciences*, 26, 107–113. <https://doi.org/10.1016/j.cobeha.2018.12.002>.
- Wiers, R. W. (2018). Cognitive Training in Addiction: Does It Have Clinical Potential? *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(2), 101–102. <https://doi.org/10.1016/j.bpsc.2017.12.008>.
- Wiers, R. W., Bartholow, B. D., van den Wildenberg, E., Thush, C., Engels, R. C. M. E., Sher, K. J., Grenard, J., Ames, S. L., & Stacy, A. W. (2007). Automatic and controlled processes and the development of addictive behaviors in adolescents: A review and a model. *Pharmacology Biochemistry and Behavior*, 86(2), 263–283. <https://doi.org/10.1016/j.pbb.2006.09.021>.
- Wiers, R. W., Boffo, M., & Field, M. (2018). What's in a trial? On the importance of distinguishing between experimental lab studies and randomized controlled trials: The case of cognitive bias modification and alcohol use disorders. *Journal of Studies on Alcohol and Drugs*, 79(3), 333–343. <https://doi.org/10.15288/jsad.2018.79.333>.
- Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic Patients' Approach bias for alcohol and improves treatment outcome. *Psychological Science*, 22(4), 490–497.
- Wiers, R. W., & Gladwin, T. E. (2016). Reflective and impulsive processes in addiction and the role of motivation. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.), *Reflective and impulsive determinants of human behavior* (pp. 173–188). Abingdon, UK: Routledge (Reflective).
- Wiers, R. W., Gladwin, T. E., Hofmann, W., Salemink, E., & Ridderinkhof, K. R. (2013). Cognitive bias modification and cognitive control training in addiction and related psychopathology: Mechanisms, clinical perspectives, and ways forward. *Clinical Psychological Science*, 1(2), 192–212.
- Wiers, R. W., Van Dessel, P., & Köpetz, C. (2020). ABC training: A new theory-based cognitive bias modification to foster automatization of alternative choices in the treatment of addiction and related disorders. *Current Directions in Psychological Science* in press.
- Wiers, R. W., Van Gaal, S., & Le Pelley, M. E. (2020). Akrasia and addiction: Neurophilosophy and psychological mechanisms. In J. Harbecke & C. Herrmann-Pillath (Eds.), *Social Neuroeconomics: Integrating the Neurosciences and the Social Sciences*. (p. in press). London: Routledge.