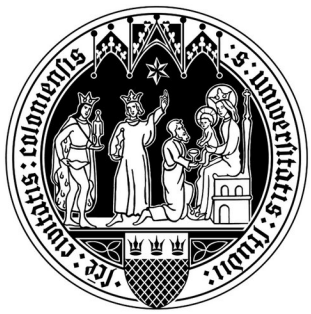
Virtual Reality as a Chance to Reconcile Ecological Validity and Experimental Control in Cue-Reactivity Research



Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 18.12.2018

vorgelegt von:

Luca-René Bruder

Abgabe am 19.09.2021

Acknowledgements

I would like to start by thanking all my great colleagues and friends at the Department of Biopsychology of the University of Cologne. Without the lively discussions and input from you all, I would have never made it to this stage of my dissertation project. I'm grateful to my peer doctoral candidates Ben and Kilian who lent their support while we all struggled through the challenges of our projects and I'm as grateful to our postdocs Julia and David who shared their wisdom on how to overcome these challenges. In addition, I would like to thank my supervisor Professor Dr. Jan Peters for his guidance and support. Your sense for details and perfectionism really helped me to be the best version of myself and perform to the best of my abilities. I had a lot of freedom in choosing the direction in which I wanted to take our research projects and I appreciated that. I wish everyone at the Department of Biopsychology of the University of Cologne the best for their future. I'm off to new adventures but my time with you will not be forgotten.

I also want to express my sincere gratitude to my fiancée Svenja. You really went through the ups and downs with me and always managed to keep me on course, despite my not too infrequent attempts to jump ship. Thank you, from the bottom of my heart. Finally, I would like to conclude by thanking my parents Birgitta and Thomas for their support, be it emotional, financial, or otherwise. I hope to pay you all back some time.

Table of contents

Introduction

Recent advances in the technology of high-performance virtual reality (VR) have enabled researchers to increase the ecological validity of stimuli presented in studies of counterconditioning[1], social gazing[2], equilibrium training[3], cue-exposure[4,5], and gambling behaviors[6,7]. The dissertation project presented here is aimed at making these technological advances accessible for research in the field of decision-making and cue-reactivity. Detrimental changes in decision-making processes form a core characteristic of a wide variety of psychiatric and neurological disorders. Addiction-related disorders such as substance-use-disorders (SUD)[8–10] or gambling disorder (GD)[11–13] form a clear example of this. Disentangling the cognitive processes involved in decision-making and which of these processes are impaired in certain psychiatric and neurological disorders is difficult, because often these processes form a black box even for the participants themselves. With the advance of the field of computational psychiatry[14] an exciting tool to dissect decision-making processes has emerged. Computational psychiatry employs theoretically grounded mathematical models to shed light on cognitive processes involved in decision-making and assesses how decision-making is disturbed in neurological and psychiatric disorders[15]. The main aim behind this is the establishment of transdiagnostic markers for these disorders[16]. This might in turn aid in the development of effective interventions as well as in the identification of treatment targets. Moreover, these markers might help to identify vulnerable individuals.

Over the years, research investigating the role of maladaptive decision-making processes in psychiatric disorders has identified a range of interrelated but distinct processes. The first process of interest in the dissertation presented here is the discounting of reward value over time (temporal discounting). Studies have associated both steeper and shallower discounting with different psychiatric and neurological conditions[8]. Tasks that gauge the temporal discounting tendencies of participants usually ask participants to choose between a constant smaller but immediate reward and a range of larger but temporally delayed rewards[17]. The resulting choices as well as the response times (RT) can then be used to quantify the discounting tendencies via computational models. Evidence is mounting that altered temporal discounting might be a transdiagnostic marker for several neurological and psychiatric disorders[8,16], with addictions and related disorders being a prominent example[16,18].

The second process that is of importance for the dissertation project presented here is reinforcement learning (RL)[19]. RL has produced considerable attention in a variety of fields, one of these being computational psychiatry. One current view is that RL is separated into two

systems, a habitual "model-free" system that learns stimulus-response associations, and a goal-directed "model-based" system that adds a model of contingencies in the environment into the calculation of action consequences[20]. To quantify the degree to which participants employ the two systems in their decision-making, researchers often use the Two-Step task (TST)[21,22]. The TST is a two-stage decision-making task where first stage decisions probabilistically determine the presented second stage choice set and consequently the rewards that can be obtained. Reductions in the usage of "model-based" RL have been associated with a range of subclinical symptoms[23]. Likewise, participants suffering from GD show reduced model-based RL[24].

Given that prominent characteristics of addiction related disorders are compulsive drug seeking and insensitivity to negative consequences[25], these disorders provide an exciting opportunity to study maladaptive decision-making processes. This is especially true for GD, which provides a clearer view on the underlying mechanisms, since GD is not based on addictive substances with direct effects on the neurobiological processes of participants. A theory that has been advanced to explain maladaptive decision-making in addictions is the Incentive-sensitization theory (IST)[26,27]. The IST postulates that neural circuits mediating the incentive motivation to obtain rewards become sensitized to cues predicting these rewards, leading to craving and drug-seeking behavior. The neural circuits thought to mediate this sensitization are assumed to be within the mesocorticolimbic dopamine system[27]. In line with this, there is evidence that exposure to addiction-related cues correlates with the modulation of striatal value signals in temporal discounting tasks[13] and increased striatal dopamine release in humans[28]. In the context of SUDs and behavioral addictions, such effects in response to addiction-related cues are termed *cue-reactivity*[29–31]. Manifestations of cue-reactivity can be observed on a physiological and subjective level[29,30]. Furthermore, exposure to addiction-related cues is assumed to increase temporal discounting[13,32,33], modulate risk-taking[34] and impair cognitive performance[35].

Studies examining cue-reactivity in participants suffering from GD typically either used picture stimuli in a lab setting[11,13,34,36–42] or were conducted in real-life gambling facilities[32,33]. These two paradigms represent two extremes of the trade-off between ecological validity and the control of confounding variables. Studies conducted in a lab grant strong control over confounding variables but lack the ecological validity of field studies. Vice versa, field studies in real-life gambling outlets arguably have a high ecological validity but grant little control over confounding variables and make comprehensive physiological measurements more difficult.

By equipping participants with head-mounted VR-glasses and a safe space to navigate the VR-environment freely, a strong sense of immersion and realism can be created. The main

idea behind this dissertation project is therefore to provide a combination of the strengths of both approaches by creating a VR paradigm to test subjective, physiological, and behavioral cue-reactivity in a lab environment. In our paradigm, participants are placed in two rich and navigable VR environments that either resemble a gambling-related casino environment or a (neutral) café environment. With the immersion taking place within a controlled lab setting we can easily accommodate the measurement of physiological variables like electrodermal activity (EDA)[43] and heart rate, as indicators of physiological cue-reactivity[29,30]. Such a paradigm could not only be applied in research but potentially in therapy to test reactions and learned cognitive strategies in a secure environment.

As VR controllers in 3D space present a fundamentally different response mode compared to typically utilized keyboards, button boxes and computer screens used in standard lab-based behavioral testing, an additional aim of this dissertation was to assess how well decision-making tasks performed in VR are suited for subsequent computational modeling of the gathered response data. Computational modeling approaches assessing latent processes underlying learning and decision-making recently started to include decision RTs in addition to simple binary decisions. One example for this are sequential sampling models such as the drift diffusion model (DDM)[44]. Reasons for this are the improved stability of parameter estimates[45,46] as well as a more complete picture of the latent processes underlying decision-making[47–50] that these models can provide. VR controllers present a fundamentally different response mode than commonly employed keyboards and screens and thus produce different RT patterns. We therefore decided to explore the applicability of drift diffusion modeling for behavioral data obtained in VR.

Taken together, the dissertation project presented here had three major aims. First, we aimed to supplement earlier work on VR applied in the context of cue-reactivity in addiction[5–7] by comprehensively modeling the choices of participants suffering from GD in decision-making tasks performed in gambling-related and neutral VR environments. For this we used a temporal discounting task[17] and the TST[21]. Second, we aimed to develop a paradigm that combines the advantages of lab and field studies with comprehensive measurement of subjective, physiological, and behavioral cue-reactivity. This approach has the potential to advance the field of addiction research and the development of novel treatment opportunities for patients. The paradigm proposed here combines state-of-the-art VR with verbally reported craving with the measurement of EDA and heart rate, and the behavioral tasks described above. Finally, we aimed to assess the feasibility of applying RT-based sequential sampling models to data obtained with VR-compatible controllers. The dissertation at hand will first describe the

background and methods in detail. Subsequently, it will present the studies conducted to achieve the goals described above. To draw a conclusion, it will be discussed how the obtained results can inform and guide further research in this area.

Background

Making rational and informed decision is a core competence for functioning in complex and mentally taxing environments. Addictions, be they SUDs or behavioral addictions, present an intriguing opportunity to study decision processes that have gone awry. Key components of addictions are impaired control or compulsivity in behavioral engagement with the addictive substance or behavior, continued engagement despite negative consequences and a state of craving after abstinence[51]. To uncover what causes these disadvantageous decisions might on the one hand improve treatment and relapse prevention for patients suffering from addictions and on the other hand it might help psychology and neuroscience learn more about the cognitive processes involved in decision-making in general. Over the years, the classification of addictions and the resulting detrimental decisions has shifted more and more towards that of neurological disorders[27]. In this regard, behavioral addictions, with GD as their most prominent example, play a special role. Compared to SUDs there are no direct effects of neuroactive substances that might change the biological mechanisms on which decisions are formed. This project therefore focuses on patients suffering from GD. Nevertheless, this text will first lay the groundwork by introducing the neurobiological basis behind addictions and decision-making in general followed by a more specific discussion of these mechanisms in the light of GD.

Dopamine and dopaminergic systems

Many different neurotransmitter systems influence the dynamics of addictions. Noradrenaline and serotonin have been shown to influence impulse control[52,53]. Opioids are related to feelings of pleasure and influence important systems for motivation and learning (see below)[54]. However, probably the most important neurobiological basis for addictions is the dopaminergic mesocorticolimbic system[11,26,28,55]. To clarify what this system entails, this text will first focus on the neurotransmitter dopamine and its main pathways within the human brain and finally elaborate the proposed mechanisms by which it shapes cognitive processes involved in decision-making, especially in the context of addictions.

Dopamine (DA) is a catecholaminergic compound that plays a key role in several cognitive processes, such as reward learning and incentive motivation[56,57]. It is synthesized from L-dopa within the cell body of dopamine neurons. It simultaneously acts as a precursor to norepinephrine and epinephrine and as a neuromodulator/neurotransmitter of its own accord[58–60]. In its role as a neurotransmitter, dopamine acts as a signal between two neurons that are connected via a synapse[61]. In the common process of synaptic transmission an action potential

in the presynaptic dopaminergic neuron causes a chain of events that leads to the release of dopamine into the synaptic cleft. Within the cleft dopamine diffuses towards the post-synaptic neuron and acts on dopamine receptors located on the membrane. Receptors reacting to dopamine exist in at least five different subtypes[62–64] D1 to D5. All these receptors are G-protein coupled receptors[60]. Thus, dopamine binding to the receptor triggers the intracellular part of the receptor to release a guanine nucleotide binding protein (G-protein) that in turn has different effects on the post-synaptic cell depending on the type of G-protein released. Based on the downstream effects caused by this G-protein, the five different subtypes can more broadly be categorized into the two categories D1-like (D1 and D5) and D2-like (D2, D3, D4). Differentiation between the two categories is commonly based on how the released G-protein modulates the activity of the enzyme adenylyl cyclase and consequently the production of the second messenger cyclic adenosine monophosphate (cAMP)[63–65]. D1-like receptors increase cAMP activity, while D2-like receptors decrease it. The result is that dopamine can have very different effects on the post-synaptic neuron depending on which receptor type is expressed. D1-like receptors are relatively more common in frontal brain regions, while D2-like receptors are generally more prevalent in the caudate nucleus (CN), putamen and nucleus accumbens (NAcc)[65,66]. Importantly however, despite their opposing effects on post-synaptic cells, D1-like and D2-like receptors often act in synergy to enable complex cognitive functioning[67,68]. In addition to its described role as a neurotransmitter, dopamine can cause widespread and temporally prolonged effects on distant neurons, i.e., act as a neuromodulator[69]. Dopamine release can take place in two different forms, tonic and phasic[70–74]. The first form is tonic dopamine release, which is caused by the spontaneous baseline firing of DA neurons. This firing results in fluctuations of extracellular DA concentration in the time range of minutes. These tonic fluctuations are too low to directly cause a change in the rate of action potentials in postsynaptic cells on their own. These changes can however modulate phasic DA firing and thus take up an important role in signal transmission[71]. Phasic firing describes the release of higher concentrations of DA into the synaptic cleft, often triggered by a series of action potentials in the presynaptic cell[72]. These action potentials are usually in response to a stimulus and may cause the postsynaptic cell to fire as well. For a detailed discussion of the role of these forms of DA release we refer the reader to the work by Grace and colleagues[70].

Dopamine neurons are sparse. Only around one percent of neurons are dopaminergic[62,75]. These neurons are mainly localized in the mesencephalon[75]. There are, however, neuronal populations in the diencephalon and the olfactory bulb[61] that are dopaminergic. Nevertheless, most dopaminergic neurons are situated within the substantia

nigra pars compacta (SNc), the ventral tegmental area (VTA) and the retrorubral field (RRF)[62,75] lying in the ventral mesencephalon (Figure 1). From here DA acts on the extrapyramidal motor system (posture and control), limbic and cortical areas (cognitive functions) and the hypothalamus-pituitary system (hormonal regulation)[58]. There are three main dopaminergic pathways that need to be mentioned here. First, the nigrostriatal pathway in which dopaminergic neurons form axons from the SNc to the dorsal striatum (CN and putamen)[76–78]. These connections play a crucial role in voluntary movement and skill learning. Second, the mesolimbic pathway spanning from the VTA to the ventral striatum (VS), especially the NAcc, as well as the septum, amygdala and the hippocampus[27,57,62,75,77]. The mesolimbic pathway plays a key role in RL and motivational behavior[56,77,79,80]. Third, the mesocortical pathway running from the VTA to prefrontal target regions, the cingulate gyrus and the perirhinal cortex[62,81,82]. The main functions influenced by this pathway are working memory and behavioral flexibility[81–83]. Due to the overlap of the mesolimbic and mesocortical pathways they are often commonly referred to as the mesocorticolimbic system[62,83]. Given that most research and theories regarding decision-making and addiction identify the mesocorticolimbic system as one of the key systems[27,84–89] behind these processes, the next paragraph will focus on proposed dopamine functioning in that context.

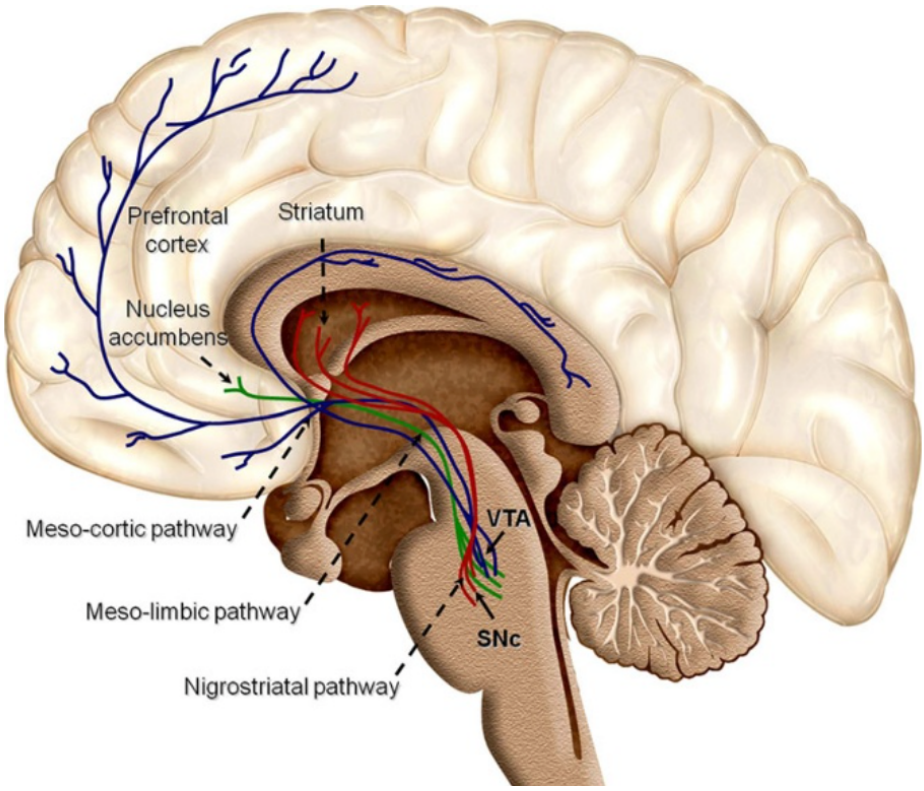


Figure 1. Illustration of the three major dopaminergic pathways in the brain. The meso-cortico/mesocortical pathway running from the VTA to frontal target regions is displayed in blue. The nigrostriatal pathway running from the SNc to the dorsal striatum is displayed in red. The meso-limbic/mesolimbic pathway spanning from the VTA to the ventral striatum is shown in green. Figure adapted from Arias-Carrion and colleagues[62].

Dopamine signaling

Crucial for the understanding of addictive processes is the neurobiology of "natural" rewarding stimuli. Most insights about how rewarding stimuli are processed by the brain come from animal studies. Probably the most important mechanism identified by these studies is DA release in the VS/NAcc[56,83,84]. The first evidence that the mesocorticolimbic system might be crucial for reward or learning was reported by Olds & Milner in 1954[90]. In their study design, often termed intra-cranial self-stimulation, rodents could stimulate areas of the mesolimbic pathway via a button press. The result was that the rodents started to press the button frequently and were increasingly willing to accept aversive stimuli to be allowed to keep pressing the button. Back then this was interpreted as a reward or pleasure signal. Today phasic DA release is more accurately thought of as a reward prediction error (RPE) signal for learning instead of a rewarding signal *per se*[56,83,84]. The agent compares its expectation of a reward that is to be received with the reward that was actually received and the difference between the two (i.e., the RPE) is reflected in the DA signal. The idea behind this dates back to classical conditioning[91] and has been further developed to the concept of RL[19]. It can be formalized mathematically as in equation 1[56], which is often referred to as temporal difference (TD) learning. Here the subjective value of a stimulus at time *t* (V_t) is calculated as the sum of the

value at timepoint *t*-1 (V_{t-1}) and the difference between the received reward at *t* and V_{t-1} (i.e., the RPE) times a learning rate α. The learning rate governs how strong the difference between expected and actual reward at trial *t*-1 changes the expectation about the reward and therefore regulates how strongly past trials are included in the valuation of a stimulus (V_t).

$$V_t = V_{t-1} + \alpha\,[R_t - V_{t-1}] \quad (1)$$

That dopamine release from the VTA into the NAcc resembles a RPE has been repeatedly shown[56,84,92–94]. In line with that, similar results have been reported using less invasive methods of neuroimaging in humans[95,96]. Interestingly, the timing of the phasic DA response shifts towards the reward predicting stimulus if the reward is expected[97]. Thus, as detailed above, an expected reward does not produce a DA response after reception, but a predicting stimulus does. This is attributed to Pavlovian learning of an association between the reward and the predicting stimulus[57]. This mechanism most likely plays an important role in the emergence and perpetuation of addictions (see below). It is important to point out here that this is a somewhat simplistic view of a very complex signal. Different research has pointed out that the phasic DA signal in the VS also varies with stimulus intensity, context and resemblance of other stimuli irrespective of reward[94]. In addition to that, information about the uncertainty of reward is also included in the ramping up of the activity towards the phasic peak[97]. Nevertheless, the view of phasic DA activity as a RPE signal lays important groundwork for current theories about addiction that will be discussed next.

Addiction-related disorders

Basically all drugs of abuse as well as naturally rewarding behaviors (i.e., food and sex) cause DA activity within the mesocorticolimbic system[11,88]. The DA releasing effect of addictive substances however is believed to be much stronger than any natural reinforcer and is also not abolished after being learned[98]. This in turn leads to an escalation by which drug cues come to dominate wanting behavior. Research with human participants revealed that dopamine functioning is altered in participants suffering from SUDs[28,99]. In line with that, early theories of how addiction manifests in the brain stated that the DA releasing effects of addictive substances lead to strong memory formation and neuroplasticity within the mesocorticolimbic system[88,99]. Over time, behavior is increasingly governed by compulsive and habitual patterns. This is mirrored in neural activity in response to addiction-related cues that gradually shifts from mainly ventral striatal areas towards the dorsal striatum[100].

Additionally, salience attribution in the orbitofrontal cortex (OFC) as well as executive functioning in the prefrontal cortex (PFC) and anterior cingulate cortex (ACC) are altered, increasing the chance of relapses[28,101].

A compelling account of how addictions arise precisely and how stimuli associated with the addictive substance become increasingly salient is proposed by the IST[27]. Reward is conceptualized as containing separable components of learning, incentive motivation and pleasure. The central pillar of this theory is that addictions become compulsive when the mesocorticolimbic system becomes sensitized and hyper-reactive to the incentive motivational value of drug-related cues and thereby generates a strong "wanting". The concept of "wanting" or "incentive salience" in the context of the IST is separated from the concept of "liking" which describes the actual pleasurable or hedonic impact of a consumed reward and is mediated by independent neuronal circuits [102,103]. Both concepts are often related in that we often "like" what we "want" and "want" what we "like". However, the two can also be dissociated by manipulations, especially those including DA[104]. On the other hand, "wanting" is less associated with cognitive goals but more linked to reward cues and how these are made attention-grabbing and attractive[105]. This form of "wanting" is mediated by the mesocorticolimbic system and enables cues to trigger urges to obtain and consume the associated rewards. The strength of these urges depends on the associated reward as well as the state of the DA-related brain systems (e.g., stress, emotional excitement, appetite or intoxication)[57]. To distinguish "wanting" from the more ordinary sense of wanting in form of cognitive desires, the terms predicted utility and decision utility can be helpful[106]. Predicted utility is the predicted value of a future reward. Decision utility is the motivational value of that outcome, as revealed by choice, pursuit, or consumption. For cognitive desires, decision utility and predicted utility are the same and stable over time[107]. For the incentive salience or "wanting", there is a situation where the decision utility is greater than the predicted utility for future rewards[55]. To put it differently, it is possible to "want" something that is not expected to be liked and is also not liked when obtained. An example would be patients suffering from SUDs that report a lack of enjoyment after drug use but still report craving it. The mechanism behind this is defined as follows. A reward (natural or addictive substance) acts as an unconditioned stimulus (UCS). A predicting or temporally related stimulus becomes associated (conditioned stimulus, CS) with the UCS via DA signaling and resulting neuroplasticity. The result is a strong sensitivity towards reward predicting CS in the VS. This sensitivity in turn triggers increased "wanting" and therefore motivation to consume the related reward. The obvious problems arising from an increased sensitivity towards addiction-related and therefore

appetitive stimuli and reduced frontal inhibition capacities for addictive behaviors and relapse[28,101] led to a strong research interest in the role these cues play in addictions. The current view on the different aspects of this 'cue-reactivity' process will be the focus of the following section.

Cue-reactivity

Resonating with the ideas from the IST, there has been a considerable amount of research in the field of cue-reactivity in SUDs[29] as well as behavioral addictions[30]. Cue-reactivity describes the phenomenon that participants suffering from addiction-related disorders show an increased reactivity towards addiction related stimuli. This reactivity most likely forms a major risk factor for relapse and other possibly problematic behaviors. Cue-reactivity may manifest on a subjective level by inducing craving for or "wanting" of the addictive goal[5,29,108]. There is also strong evidence that addiction-related cues cause behavioral changes in reward-based decision-making of participants. This has been shown for a range of tasks, with the temporal discounting tasks being the most relevant here[12,13,109,110]. Additionally, it might manifest on a physiological level in peripheral measures such as heart rate and skin conductance[29]. A range of studies has shown cue-reactivity on the neural level in SUDs[111–113]. The results indicate the PFC, ACC, striatum, insula, hippocampus, and amygdala as structures involved in cue-reactivity[40,114–116]. The VS/NAcc as a crucial part of the mesocorticolimbic system and its role in the mechanisms proposed by the ICT is of special interest here[55]. Since addictive substances have direct effects on the neurobiological processes of participants, we will now focus on gambling disorder, which provides a less perturbed view on the underlying mechanisms. A possible explanation for the addictive properties of gambling, that lacks a neuroactive agent, is that uncertainty itself increases DA signaling and is itself rewarding[85,97]. The perpetual uncertainty of gambling is thus seen as UCS in the IST mechanisms described above. This is supported by evidence showing that the reward schedule of commercial gambling machines resembles conditions that lead to increased DA firing during the anticipation of reward[117]. Cue-reactivity on the different levels mentioned above has been demonstrated in participants suffering from GD[13,32,36,41,42,115]. Cues that are related to gambling lead to an increased subjective urge to gamble[118] and cause increases in psychophysiological measures such as EDA and hear rate[119]. On a neurobiological level, participants suffering from pathological gambling show increased reactivity to gambling related cues in neuronal circuits associated with reward, learning, habits and executive control, similar to those shown by participants suffering from SUDs[30,41,42,85,115,116,120,121] (see above). Limbrick-Oldfield[42] and

colleagues were able to show that the subjective urge to gamble induced by gambling related cues was correlated to activity in the VS, lending further credence to the IST. However, it is important to note that the directionality of these effects is still unclear[85]. In line with the reward deficiency theory postulated by Volkow and colleagues[28], participants suffering from GD show decreased responsivity to primary reinforcers like erotic pictures and stronger reactivity to monetary rewards[122]. Furthermore, participants suffering from GD show stronger reactivity to monetary rewards in VS, VTA and ventromedial prefrontal cortex (vmPFC) during realistic gambling tasks[123–125] compared to healthy control participants. Behavioral cue-reactivity has been demonstrated by studies using decision-making tasks, such as temporal discounting[13,32]. Please find a detailed discussion on these results below.

Studies probing cue-reactivity in the context of GD commonly either present visual stimuli on a computer screen within a standardized lab environment or in a scanner[11,13,34,36,38,40–42,115,123] or conduct experiments in real-life gambling facilities[32,33]. These options present two extremes of a trade-off between ecological validity, which is high in a real-life gambling facility, and the control of possibly confounding variables, which is high in a lab surrounding. Moreover, field studies in a real-life gambling facility make it difficult to measure psychophysiological or neurobiological variables.

Virtual reality

The emergence of high-performance VR technology in recent years opens an exciting possibility to combine the best of both worlds described before. In VR paradigms participants are equipped with VR headsets. The VR headset covers the whole visual field of the participant and presents images that are matched to the tracked motion of the participant. By doing so, VR creates a strong sense of immersion in a virtual environment. Modern VR headsets can update the presented images so fast that there is no perceptible lack between movement and image update. This helps to overcome early problems with VR technology, which caused vertigo and nausea in participants. Thus, modern VR technology enables designs that provide a high level of ecological validity via immersion in virtual environments, while taking place in a highly controllable lab setting. This combination of methodological advantages makes VR a very attractive tool in addiction research as well as therapeutical practice. The hope behind research with VR headsets is to create a paradigm that can expose patients, be it patients suffering from GD or any other form addiction-related disorder, to realistic cues in a safe environment and train them to use cognitive behavioral therapy techniques to reduce stress and craving[5]. Naturally, this hope can only come to fruition if research can reliably show that addiction related

cues can induce cue-reactivity on a level sufficient to justify the extensive effort of creating realistic VR designs. Studies assessing cue-reactivity in VR environments demonstrated increases in craving in participants suffering from GD[5,108], participants suffering from nicotine addiction[126] and participants suffering from alcohol addiction[4]. Furthermore, Wang and colleagues[1] were able show that reactivity to metamphetamine-related cues could be reduced by a VR counter-conditioning procedure. In a promising study by Dickinson and colleagues[7] on participants suffering from GD, higher levels of arousal and immersion were found during a gambling task in VR compared to a laptop. Taken together, these results demonstrate the great promise of VR. VR designs targeted at patients and participants suffering from GD so far have focused on one VR environment with gambling related cues and possibly a neutral starting area to accustom participants to VR. To date, there is no study that combines these designs with a systematic evaluation of cue-reactivity on all levels described above. In the dissertation project presented here, we propose a VR design that combines decision tasks from computational psychiatry, measurement of psychophysiological variables and subjective urge to gamble ratings. Please find a detailed description of the VR environments in the studies reported below.

We hope that we can contribute to cue-reactivity research in addiction by overcoming ecological validity shortcomings of earlier lab-based studies and help to advance the use of VR in treatment of patients suffering from GD and addictions in general. Additionally, we hope to provide preliminary evidence that commonly used decision-making and RL tasks can be applied in VR with confidence. The next sections will present a detailed discussion of the different variables measured within the VR environments and which statistical methods were used to analyze the resulting data.

Methods

In the next sections this text will discuss how the different forms of cue-reactivity were operationalized in this dissertation project to overcome the methodological issues of ecological validity shown by earlier studies in cue-reactivity describe above.

Subjective cue-reactivity

The measurement of subjective cue-reactivity in VR was accomplished by asking participants suffering from GD to report their urge to gamble on a 10-point likert scale. The experimenter verbally asked the participants to verbally report their urge to gamble during five time points during the VR exposure. Verbal reports of craving have been used successfully in previous cue-reactivity research[127] and for comparable designs[108].

Behavioral cue-reactivity

Analyzing the learning and decision-making behavior of participants suffering from addictions is often accomplished with cognitive models developed in the field of computational psychiatry. Computational psychiatry employs theoretically grounded mathematical models to disentangle behavioral abnormalities and develop improved diagnostic criteria as well as treatments. The following paragraph will discuss the two behavioral tasks used in the two studies presented here and describe how the resulting data was analyzed using methods from the field of computational psychiatry[14].

Temporal discounting

As mentioned above, temporal discounting has been widely used for cognitive modelling in the context of GD[12,32,109,128]. In temporal discounting tasks, participants are repeatedly confronted with a choice between a smaller immediate reward (SS) and larger but delayed rewards (LL)[17]. It is commonly accepted that the degree to which participants discount the value of a delayed reward can be seen as a direct measure of impulsivity[129] (Figure 2). This is supported by studies showing a strong association between temporal discounting measures and personality questionnaires assessing trait impulsivity[12,130]. Furthermore, the individual parameters obtained via temporal discounting measures show a strong temporal stability. Studies assessing the test-retest-reliability of temporal discounting measures consistently attest good to excellent reliability[131–133]. There is even evidence for the heritability of impulsivity measured with temporal discounting[134]. Generally, alterations in reward based decision-making, such as heightened temporal discounting, are proposed to be diagnostic marker of addictions[18]. Participants suffering from GD haven been shown repeatedly to have an increased tendency to discount future rewards[17,39,128]. But the degree to which discounting of future rewards is related to the severity of the GD symptoms is still unclear. There are results that show that South Oaks Gambling Screen (SOGS) scores are related to the steepness of the discounting curve[110] (see below). However, other studies failed to identify a strong directional relationship[17,39]. Despite the unclear association between gambling severity and impulsivity, Miedl and colleagues[17] showed that gambling severity and the representation of values in the VS, OFC and midbrain dopaminergic regions were correlated. These results make sense in the light of addiction research discussed above and resonate well with other results that identify activity within a circuitry of VS, vmPFC and posterior cingulate cortex as the representation of SS and LL values in temporal discounting[101]. The concept of cognitive control also plays an

important role here as Figner and colleagues[135] showed that modulation of the dorsolateral prefrontal cortex (dlPFC) activity via transcranial magnetic stimulation can increase the tendency to discount values of future rewards. The authors concluded that foregoing a smaller but immediate reward necessitates more self-control than foregoing larger but delayed rewards. There is also evidence that participants suffering from GD show impaired cognitive control in a stroop task that correlated with activity changes in the vmPFC[136]. How increased reactivity to gambling related cues in the VS might interact with these findings to influence temporal discounting has also been assessed. Miedl and colleagues[13] showed that gambling cues presented in the background during a temporal discounting task modulate striatal value representation to a degree proportional to induced subjective craving. In the mentioned study, activity related to modeling-based subjective value in the VS was down regulated by cues causing higher craving. These results are reminiscent of other studies assessing temporal discounting in appetitive conditioning contexts[137,138], supporting the idea of incentive sensitization of gambling related cues. The neurobiological observations are supported by an increased discounting rate in the presence of gambling related cues[13,32]. Again, this fits well into the addiction frameworks discussed above, as modulated value representation in key areas in combination with impaired frontal control of impulses are also thought to play a role in temporal discounting[135]. Consequently, temporal discounting tasks represent an attractive way to investigate behavioral effects of addiction related stimuli. Next, the precise way of how these tasks can be analyzed using computational psychiatry methods is discussed.

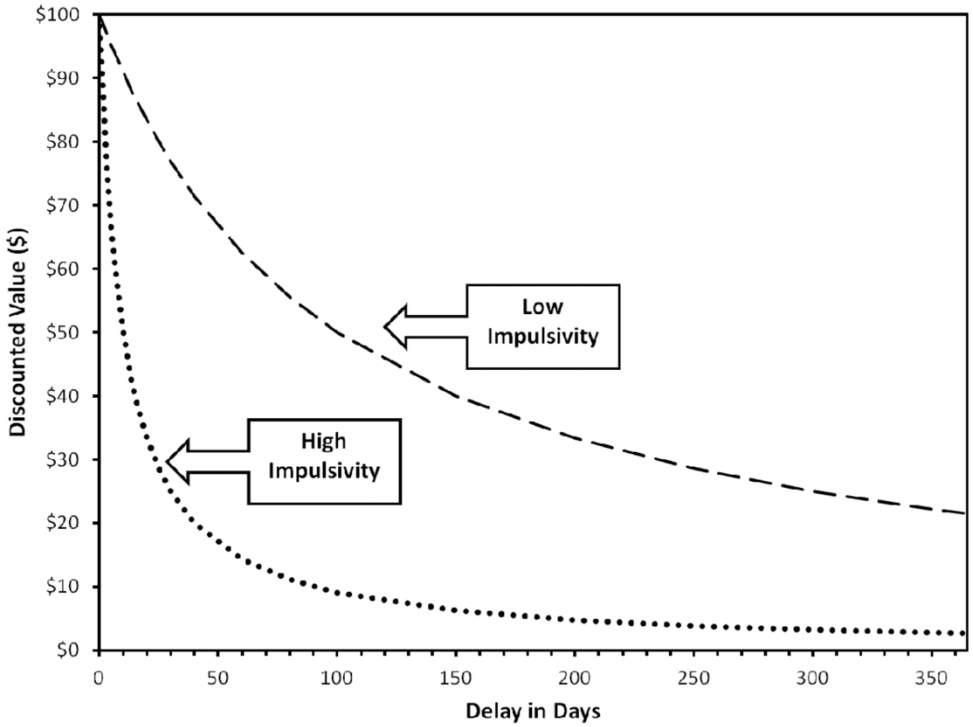


Figure 2. Temporal discounting visualized with a hyperbolic discounting function. The y-axis displays the discounted subjective value of a monetary reward of 100$. The x-axis represents the temporal delay until the reward is paid out. In the hyperbolic model (Eq. 2) high values of *log(k)* indicate a steep curve, which is thought to be associated with higher impulsivity. Low values of *log(k)* lead to a more gradual decline of value over time and are thus thought to reflect less impulsive choices. The figure was adapted from MacKillop and colleagues[139].

Analysis of temporal discounting tasks

Model agnostic analysis. A simple way to analyze the choices made during a temporal discounting task, that does not include complex cognitive modeling, is the area under the curve (AUC)[140]. To obtain AUC values a logistic function is fit to the choices for each delay. The LL value at which this logistic function assigns a probability of .5 is termed the indifference point, because here the larger value has been delayed so much that its discounted subjective value is equal to the value of the SS. The resulting indifference points are then plotted against their corresponding delays. Finally, the area under the resulting curve is calculated and noted as the AUC.

Hyperbolic model. Cognitive modelling approaches to the analysis of temporal discounting choices and/or RTs usually propose that the value decay of the delayed options resembles a hyperbolic function[141,142] (Figure 2). In the this dissertation we therefore chose to use a hyperbolic discounting model in combination with two different choices rules. The hyperbolic discounting model defines the subjective value of a delayed option at trial *t* ($SV(LL_t)$) as the objective value at trial *t* (A_t) divided by a discounting term (Eq. 2).

$$SV(LL_t) = \frac{A_t}{(1+\exp(k)*D_t)} \quad (2)$$

The discounting term consists of the parameter k which governs the steepness of the discounting and D_t representing the associated delay at trial t. To increase the numerical stability of the model for values close to zero we include *exp(k)* in our model. This is later reversed by reporting *log(k)* in the results. By adding choice rules to the calculation of subjective values, we introduce decision noise and create a more plausible model of decision-making[19]. The different choice rules will be discussed next.

Softmax choice rule. The softmax choice rule is commonly used in the context of RL, machine learning and decision-making[19]. For the softmax action selection, the probability of choosing the LL option on trial *t* is given by equation 3.

$$P(LL_t) = \frac{\exp(SV(LL_t)* \beta)}{\exp(SV(SS_t) * \beta)+ \exp(SV(LL_t) * \beta)} \quad (3)$$

The stochasticity of the choices with respect to the calculated $SV(LL_t)$. is determined by the β -parameter. A β of zero would indicate that choices are random, whereas higher β values indicate a higher dependency of choices on option values. We extended this hyperbolic model with a softmax choice rule with shift parameters that describe the shift in the parameters β (s_β) and k (s_k) (Eq. 4 and 5) from a neutral condition to a gambling related condition.

$$SV(LL_t) = \frac{A_t}{(1 + (\exp(k + s_k * I_t)) * D_t)} \quad (4)$$

$$P(LL_t) = \frac{\exp\left(SV(LL_t) * (\beta + s_\beta * I_t)\right)}{\exp\left(SV(SS_t) * (\beta + s_\beta * I_t)\right)+ \exp\left(SV(LL_t) * (\beta + s_\beta * I_t)\right)} \quad (5)$$

I is a dummy variable identifying to which condition (neutral or gambling related) a trial belongs.

DDM choice rule. Another possible choice rule to simulate decision processes is the DDM. The DDM is a cognitive model that describes simple binary choice tasks as diffusion processes between two boundaries[44]. In recent years interest in the DDM has skyrocketed and it has been applied to various forms of temporal discounting and RL tasks[47–49,133,143,144]. There

are two main reasons for this. First, by including the full RT distribution together with the choices made by the participants, the stability of parameter estimates given by the model can be drastically improved[45,46]. Second, by modeling the decision process in this way it is possible to enable the assessment of underlying latent decision processes[48,144] (see below). Therefore, the modelling of DDM parameters was included in all projects described here.

Within the framework of the DDM it is assumed that decisions are formed by a noisy evidence accumulation process over time that approaches one of two boundaries[44] (Figure 3 top panel). The boundaries represent the two possible choices. In the context of correct or incorrect decision the upper boundary represents correct and the lower boundary wrong decision. However, in the realm of value-based decisions the two options are simply assigned to the boundaries. As soon as one of the boundaries is crossed by the diffusion process the corresponding choice is executed. The DDM allows to disentangle the different parameters that describe the evidence accumulation process. By adjusting these parameters, the model fits the empirically obtained RT distributions to the RT distributions produced by the DDM. The four most important parameters are the drift-rate v, the boundary separation α, the non-decision time τ and the decision bias z. v describes the rate of evidence accumulation. The more positive/negative v is the faster one of the two boundaries is reached (Figure 3 top panel). A highly positive/negative v indicates stronger evidence and hence an easier and faster decision. Faster decisions lead to a more positively skewed RT distribution and a shorter tail (Figure 3 middle panel). Slower decisions lead to a less skewed RT distribution and a longer tail. The boundary separation parameter α determines the distances between the two boundaries (Figure 3 top panel). It can be seen as decision threshold, because the further the two boundaries are apart the more evidence needs to be accumulated to form a decision. Consequently, noise inherent in the process has less influence on the decision and decisions are executed slower. α is thus often described as governing the speed/accuracy trade off. The starting point parameter z governs at which point between the two boundaries the diffusion process starts (Figure 3 top panel). It can bias decisions towards a certain boundary and therefore change how responses are distributed across the two choice options. Finally, the non-decision time τ picks up how long the sensory encoding of the evidence (u) and the execution of a response after a decision is made (w) take (Figure 3 bottom panel). In the models of the temporal discounting task presented here, we replaced the softmax function as a decision rule with the DDM to gain additional insights into how decisions are formed in the different contexts.

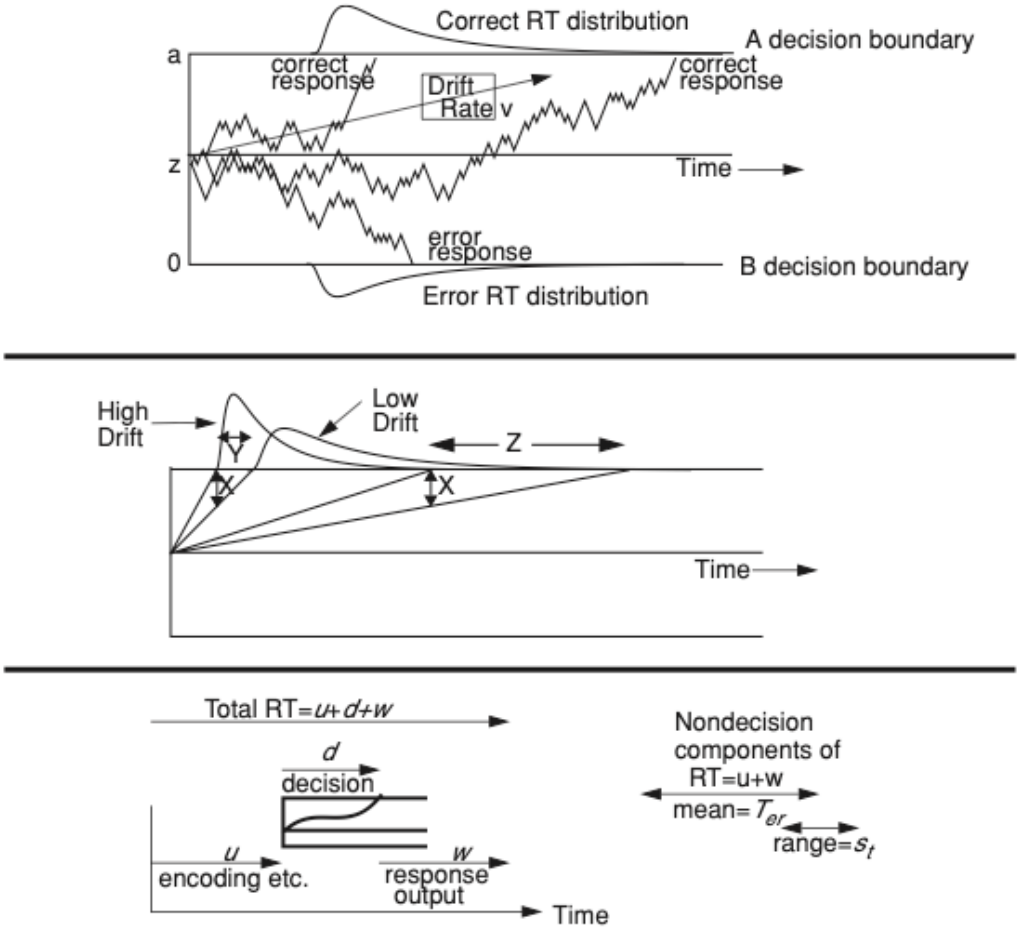


Figure 3. Illustration of the dynamics of modelling choices based on the DDM choice rule. Top: Illustration of what the different latent choice parameters describe within the DDM framework, and how they change the estimated RT distributions. Middle: Detailed view on how the drift rate influences the estimated RT distributions. Very positive/negative drift rates lead to fast decisions, which in turn lead to a more positively skewed RT distribution and a shorter tail. Less extreme drift rates lead to broader, less skewed estimated RT distributions. Bottom: Illustration of the composition of RTs within the DDM framework. The total RT in each trail is made up from the diffusion process (d) and the non-decision components encoding/perception of information (u) and response output/motor execution (w). Within the DDM the parameter τ models the non-decision components u and w.

In the projects described here we used the Wiener first passage time to model the distribution of the RT on trial *t* (Eq. 6).

$$RT_t \sim \text{wfpt}(\alpha, \tau, z, v) \quad (6)$$

We tested and compared different extensions of the DDM to see if the standard DDM null model (DDM_0) described above could be improved. The DDM_0 does not include the value difference between the two options and consequently assumes the drift rate v is the same in each trial. We extended this null model with two different temporal discounting DDMs in which the subjective value difference between the two options was included to modulate the trial-wise

drift rates. First, we tested a linear linking function[50] (DDM_L). The drift-rate v in trial t depends linearly on the scaled value difference between the LL option and the SS option in the temporal discounting task (Eq. 7)[48]. The additional parameter v_{coeff} maps the value differences onto v and scales them accordingly.

$$v_t = v_{coeff} * (SV\ (LL_t) - SV(SS_t)) \quad (7)$$

One major problem with the DDM_L is that the linear representation of the relationship between v and the value differences is that v increases infinitely with high value differences. That might lead to an underprediction of RTs for high value differences[144]. Previous work has shown that a successful way of changing this is by implementing a non-linear sigmoidal relationship between the v and the trial-wise value difference of the options[47,49,144]. This model, called DDM_S, constrains the growth of v with the value difference to the asymptote v_{max}. To accomplish this the linear mapping function from the DDM_L is passed through the sigmoid function S (Eq. 8 and 9).

$$v_t = S\ [v_{coeff} * (SV(LL_t) - SV(SS_t))] \quad (8)$$

$$S(m) = \frac{2 * v_{max}}{1 + \exp(-m)} - v_{max} \quad (9)$$

Equivalent to the softmax models described earlier we included a version of the model in which we added a condition-based shift parameter to the DDM_S model (Eq. 10 to 12) in one of the projects.

$$RT_t \sim wfpt\ (\alpha + s_\alpha * I_t, \tau + s_\tau * I_t, z + s_z * I_t, v + s_v * I_t) \quad (10)$$

$$v_t = S\ [(v_{coeff} + s_{v_{coeff}} * I_t) * (SV(LL_t) - SV(SS_t))] \quad (11)$$

$$S(m) = \frac{2 * (v_{max} + s_{v_{max}} * I_t)}{1 + \exp(-m)} - (v_{max} + s_{v_{max}} * I_t) \quad (12)$$

The additional s-parameters capture the shift in value from one condition to the other, while I is still a dummy variable indicating which condition the trial t belongs to. Next, we will discuss the second task included in the project presented here.

Two-Step Task

Performance on this second behavioral task employed here has also been proposed as a possible marker for psychiatric diseases in computational psychiatry[145]. It is called the Two-Step Task (TST). Theories describing decision-making emphasize the distinction between two systems that work in concert to control behavior[146–148]. The two systems are differentiated along the lines of the trade-off between accuracy and computational demands. One system is fast and automatic but sometimes inaccurate, while the other is slow and effortful but more accurate. The fast system is often compared to Thorndike's law of effect, which states that an action that was rewarded is more likely to be repeated[149]. This resonates well with the RPE signal represented by the DA signaling in the VS and it has been proposed that the two are one and the same[149,150]. Consequently, recent research has formalized both systems within the framework of RL[19,21,22]. The fast and automatic system is formalized as model-free (MF) TD learning described above, whereas the slow and deliberate system is formalized as model-based (MB) learning. MF learning is called model-free because it lacks a representation of causal structures in the environment. It learns the value of an action by a weighted average of past rewards (i.e., TD learning). This produces fast but inflexible response tendencies or habits, which may form the basis of addictions[151]. MB learning on the other hand stores an explicit causal model of the environment and uses this model to formulate plans for future actions. While the computational costs of MB learning are thus much higher it also has a higher potential for accurate choices because changes in the environment can be quickly incorporated into the decision-making processes. The favored method of dissociating these mechanisms of learning and value-based decision-making is the TST[21]. The TST is a sequential decision-making task in which participants must make a series of choices between two abstract picture stimuli (Figure 4). Each trial consists of two stages. In the first stage participants must choose between two stimuli. In each trial the first stage stimuli are always the same two abstract fractals. These two first stage stimuli probabilistically determine which of two possible second stages are presented subsequent to choice execution. To be precise, one first stage stimulus leads to one second stage with a probability of seventy percent (common transition) and with a probability of thirty percent (rare transition) to the other possible second stage. This is reversed for the other first stage stimulus. Each possible second stage has its own pair of abstract picture stimuli. Traditionally, the second stage stimuli are associated with a probability of returning a constant reward. The reward probability associated with each of the four (two for each stage) second stage stimuli is slowly and independently changing in form of a predetermined quasi-random walk. The rare transitions allow for a dissociation between MF and MB choice behavior. Since

the MF strategy is insensitive to the underlying causal structure of the task it will simply increase the probability of a choice based on the reception of reward independent of the transition preceding it (Figure 5 A). Choices guided by MB learning reflect an interaction between the transition type and the reward. A choice that is rewarded after a rare transition will lead to a decrease of the probability to repeat the associated choice (Figure 5 B). The behavior of participants usually reflects a mixture of these two types (Figure 5 C). This version of the task has been criticized for creating situations in which MB strategies do not perform best and thus not being optimal for dissociating the two systems[22]. A potential remedy for the problems with the original task version is replacing the reward probabilities with continuous rewards that also vary in the form of predefined quasi-random walks[22]. This version of the TST was used in the projects presented here. The TST has proven to be a valuable tool to characterize the neurobiological implications of the two systems within this RL framework.

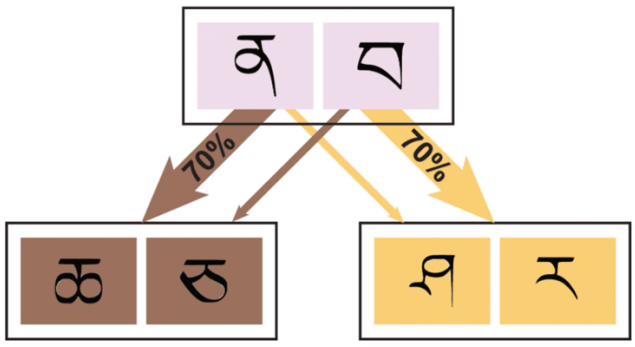


Figure 4. Schematic representation of the Two-Step task. In each trial participants start in the first stage (pink). Their decision in this stage probabilistically determines which of two possible second stages will be presented next. One first stage stimulus (left side) leads to the brown second stage with a probability of 70% and with 30% to the yellow second stage. The other first stage stimulus (right side) does the same but with reversed probabilities. Second stage choices are rewarded based on predetermined quasi-random walks. Figure adapted from Kool and colleagues[22].

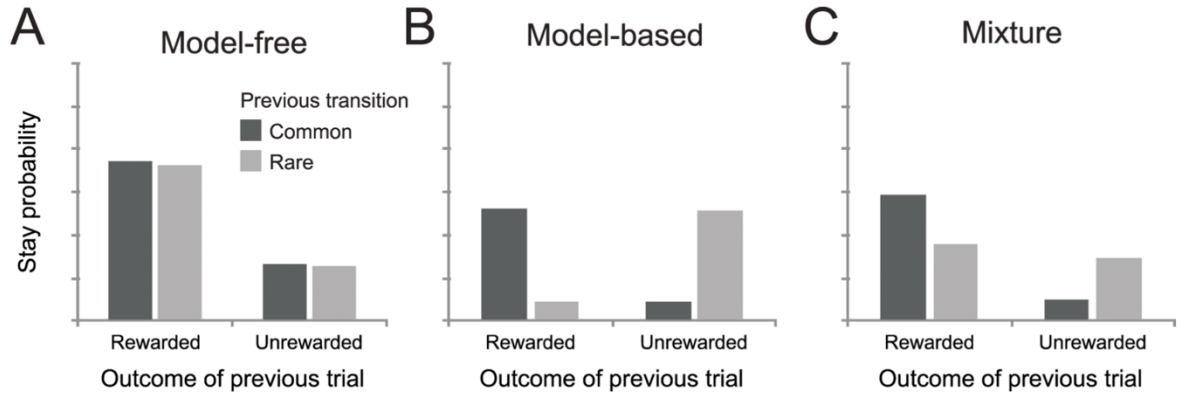


Figure 5. Visualization of the behavioral patterns reflecting the different systems within RL. The y-axis represents the probability of repeating the first stage choice of the previous trial. A) Model-free behavior simply guided by the reception of a reward in the previous trial. B) Model-based behavior guided by the reception of reward as well as the transition type in the previous trial. C) A mixture of both types of RL. Adapted from Kool and colleagues[22].

As mentioned before, the MF decision-making is supposedly mediated by DA signaling in the mesocorticolimbic system[150]. Further studies also identified a correlation between MB value representations and activity in the vmPFC/OFC[152,153]. This fits nicely with earlier research that showed expected future rewards are represented in the VS and vmPFC[154]. That frontal areas involved in the mesocorticolimbic DA system play an important role in processes supposed to contribute to MB learning, such as forward planning, mental simulations and representing real world contingencies, further highlights the importance of frontal executive control[21,81,155–157]. Importantly, there is evidence that DA signaling in the VS has also been shown to represent both MF and MB learning to a degree proportional to the choice behavior. In line with this, presynaptic DA levels in the VS were found to correlate with MB choice behavior in healthy control participants[155,157]. As many current views of addiction emphasize the disbalance between a hypersensitized reward system and a reduction in frontal executive control over the developed habits and action impulses[55,135] (see above), the TST presents an intriguing option to explore changes in decision-making processes in GD.

In the context of addictions, the TST has produced mixed results. A range of studies failed to identify any difference in MF and MB learning between control and alcohol dependent participants[145,158,159]. In contrast, other results indicated impaired MB learning in participants suffering from alcohol dependence[160], binge drinking[161] and metamphetamine dependence[158]. Up to this date there is little research about the balance between MF and MB learning in GD. Wyckmans and colleagues[24] demonstrated that participants suffering from GD show reduced MB learning when compared to healthy control participants. This effect was driven by trials in which no rewards were received. A similar effect could be seen in the RTs of participants suffering from GD. The researchers observed a significantly faster RT after unrewarded trials and hypothesized that this could be caused by reduced loss aversion and less flexible habit-like behavior. In line with the results from Wyckmans and colleagues[24], reduced reactivity to losses and stronger loss anticipation have been demonstrated in participants suffering from GD before[162]. Because addiction related cues are suspected to increase impulsivity[13,32,33] and dependence on pathological habits[26,163], it is natural to assume that participants suffering from GD might decrease their reliance on MB systems[23,164] in favor of a more MF choice behavior. How the choice patterns produced by the TST can be analyzed using computational psychiatry[14] will be discussed next.

Analysis of the Two-Step Task

Model agnostic analysis. Identical to the analysis of the temporal discounting data, analysis of the TST can be accomplished with approaches that are less modeling heavy. The first option is to simply analyze the final score participants achieve during the task, as higher scores are usually associated with a stronger MB influence. A second approach that utilizes less cognitive modeling is assessing the likelihood of making the same first stage choice in the current trial as in the previous trial (pStay) with a hierarchical generalized linear model (HGLM)[21]. The main effects of reward (representing if the last trial was rewarded), transition (common or rare) and the reward x transition interaction are included in the model. A significant effect of reward would indicate MF choices, while a significant interaction between reward and transition would indicate MB behavior. In this analysis both terms can be significant, demonstrating that both systems influence the pStay. To disentangle cue-reactivity and group effects, additional factors like group and condition as well as their interactions can be added to the model. As described above, the version of the TST used in the project at hand did use continuous rewards, making the interpretation of the reward factor difficult as it is not yes (1) or no (0). To overcome this, we included an analysis of second stage RTs. There is evidence from previous research that suggests that longer second stage RTs after a rare transition are associated with stronger MB control, because participants employing more MB control expect a common transition but are startled by a rare transition[165]. We thus additionally constructed a HGLM which included the factors transition (common or rare), condition ($VR_{gambling}$ or $VR_{neutral}$) and group (gambling or non-gambling control) to model second stage RTs.

Softmax choice rule. Like the temporal discounting task, there is a range of possible ways to include computational modeling in the analysis of the TST. One possible cognitive computational modeling approach is based on the analyses from Otto and colleagues[166]. The so called 'hybrid model' combines an RL model that learns state-action values (Q-values) for the options in the different stages with a softmax or a DDM choice rule. For the first stage both the MB and MF Q-values are tracked, whereas there are only MF Q-values (Q_{MF}) in the second stage. The update of Q_{MF}-values in both stages on trial *t* is achieved via the prediction error δ (Eq. 13 to 16). The subscript j (j ϵ [1,2]) denotes the two possible actions in each stage, while *i* denotes the stage presented.

$$Q_{MF}(s_{2j,t}, a_{j,t}) = Q_{MF}(s_{2j,t-1}, a_{ij,t-1}) + (\eta_2 + s_{\eta_2} * I_t) * \delta_{2,t} \quad (13)$$

$$Q_{MF}(s_{1i,t}, a_{j,t}) = Q_{MF}(s_{1,t-1}, a_{j,t-1}) + (\eta_1 + s_{\eta_1} * I_t)\delta_{1,t-1} + (\eta_2 + s_{\eta_2} * I_t)\delta_{s2} (14)$$

$$\delta_{1,t} = Q_{MF}(s_{2,t-1}, a_{j,t-1}) - Q_{MF}(s_{1,t-1}, a_{j,t-1}) (15)$$

$$\delta_{2,t} = r_{2,t-1} - Q_{MF}(s_{2i,t-1}, a_{j,t-1}) (16)$$

In the equations above, r represents the second stage rewards of the previous trial (*t*-1). As there are no first stage rewards, first stage Q_{MF}-values are updated based on the second stage Q_{MF}-values and the second stage prediction error (δ_{s2}). η_1 and η_2 represent the first and second stage learning rates (i.e., the impact of RPEs on future reward expectation). To help model convergence, the learning rates were transformed into standard normal space [-4, 4]. The model described here is a shift model that models condition effects on parameter estimates by shift parameters that are added onto the main parameters. *I* again represents the session presented at trial *t*. By removing s terms, the model can be used for single conditions, like the first hyperbolic temporal discounting model described above.

Model based Q-values (Q_{MB}, see Eq. 17) are calculated based on the transition probabilities between the first and the second stage and the second stage Q_{MF}-values, as there are no rewards in the first stage.

$$Q_{MB}(s_1, a_j) = P(s_{21}|s_1, a_j)\, max\, Q_{MF}(s_{21}, a) + P(s_{22}|s_1, a_j)\, max\, Q_{MF}(s_{22}, a) (17)$$

Since there are no further stages to the task, for the second stage, $Q_{MF} = Q_{MB}$. The softmax choice rule governing second stage choices (Eq. 18) is analogous to the temporal discounting softmax choice rule described above. Here, β_2 governs the choice stochasticity.

$$P(a_{i,t} = a|s_{2,t}) = \frac{\exp\left((\beta_2 + s_{\beta_2} * I_t)Q_{MF_{S2}}(a)\right)}{\sum_{a'} \exp\left((\beta_2 + s_{\beta_2} * I_t)Q_{MF_{S2}}(a')\right)} (18)$$

The softmax choice rule governing first stage choices (Eq. 19) adds separate weighting parameters for Q_{MF} and Q_{MB} (β_{MF} and β_{MB}) that model the impact of these Q-values on first stage choices. The selection of actions in the first stage is thus modelled by weighing the influences of MB and MF Q-values.

$$P(a_{i,t} = a|s_{1,t}) = \frac{exp\left((\beta_{MB} + I_t * s_{\beta MB})Q_{MB_{S1}}(a) + (\beta_{MF} + I_t * s_{\beta MF})Q_{MF_{S1}}(a) + (\rho + I_t * s_\rho) * rep(a)\right)}{\sum_{a'} exp\left((\beta_{MB} + I_t * s_{\beta MB})Q_{MB_{S1}}(a') + (\beta_{MF} + I_t * s_{\beta MF})Q_{MF_{S1}}(a') + (\rho + I_t * s_\rho) * rep(a')\right)}, \quad (19)$$

Furthermore, we included the parameter ρ that models the tendency to repeat the action from the previous trial (perseveration). *Rep* takes a value of 1 if the corresponding action was taken on trial *t*-1, and 0 otherwise. Based on previous work on RL[167,168] we also included a term that decayed the Q-values of unchosen options in both stages (Eq. 20). The decay rate η_{decay} determines the speed of this decay towards the center of the reward walks (.5).

$$Q_{unchosen}(s_{i,t}, a_{i,t}) = Q_{unchosen}(s_{i,t-1}, a_{i,t-1}) * \left(\eta_{decay} + s_{\eta decay} * I_t\right) + \left(1 - \left(\eta_{decay} + s_{\eta decay} * I_t\right)\right) * 0.5 \quad (20)$$

DDM choice rule. Parallel to the analysis of the temporal discounting, it is possible to add a range of models that employ a DDM choice rule (see above) to model participant choices and RTs in the TST. Again, the three different DDMs replaced the softmax choice rule as described above. However, the trial wise drift-rates depended on the difference in Q-values between options ($\Delta_{Q_{MB}}$ and $\Delta_{Q_{MF}}$ respectively). For second stage options, this is only the case for Q_{MF} value differences (Eq. 21):

$$v_{S2,t} = \left(v_{coeffS2} + s_{vcoeff_{S2}} * I_t\right) * \Delta_{Q_{MF_{S2}}} \quad (21)$$

The DDM choice rule for the first stage includes Q_{MF} and Q_{MB} values and separate drift-rate coefficients ($v_{coeffMB}$ and $v_{coeffMF}$) for both (Eq. 22):

$$v_{S1,t} = \left(v_{coeff_{MB}} + I_t * s_{vcoeff_{MB}}\right) * \Delta_{Q_{MB}} + \left(v_{coeff_{MF}} + I_t * s_{vcoeff_{MF}}\right) * \Delta_{Q_{MF_{S1}}} + (\rho + I_t * s_\rho) \quad (22)$$

The DDM_L formulation of the model only used the drift rates derived from equations 20 and 21. For the DDM_S version the drift-rates are additionally passed through a sigmoid function S to prevent the underestimation of RTs for high value differences (see Eq. 23):

$$S(m) = \frac{2 * \left(v_{max_{Si}} + s_{v_{max_{Si}}} * \ I_t\right)}{1 + exp\ (-m)} - \left(v_{max_{Si}} + s_{v_{max_{Si}}} * \ I_t\right) (23)$$

In all the described models, separate DDM parameters are estimated for the first and second stage of the task. Next the process of parameter estimation itself will be discussed.

Hierarchical Bayesian modelling

Estimating the best fitting parameter values for these cognitive models is a complex optimization problem. The fields computational psychiatry and cognitive modeling have recently focused more and more on hierarchical Bayesian parameter estimation to fit cognitive models to behavioral data. The process and background of hierarchical Bayesian modelling will be concisely described here. For a broader and more fundamental discussion the reader is referred to Farrell and Lewandowsky[15]. The idea behind Bayesian parameter estimation and modelling is based on the Bayes Theorem of conditional probabilities (Eq. 24).

$$P(a|b) = \frac{P(b|a) * P(b)}{P(a)} (24)$$

Here the $P(a|b)$ is the conditional probability of a given that b was observed. $P(b|a)$ is the same but vice versa (i.e., the conditional probability of b if a was observed). $P(a)$ and $P(b)$ in turn represent the probabilities of observing a and b respectively. In the context of parameter estimation with Bayesian cognitive modelling this rule is translated to give a probability distribution over possible parameter value (Eq. 25).

$$P(\theta|y) = \frac{P(y|\theta) * P(\theta)}{P(y)} (25)$$

In this equation, $P(\theta|y)$ is the probability of parameter values given the data. It represents the probability distribution over the possible parameter values called the posterior. The mode of this distribution gives us the most likely value of a parameter in our sample. Having a distribution over parameter values is one advantage Bayesian cognitive modelling has over

frequentists approaches that only deliver point estimates. $P(\theta)$ is the prior and represents the prior beliefs we have about the probability distribution over possible parameter values. This can either be an uninformative prior in which every possible parameter value has the same probability or an informative prior, which reflects prior knowledge about the possible shape of the posterior. An informative prior could for instance be the posterior obtained in a similar study. The application of a prior is the other fundamental difference from frequentist statistics where prior knowledge about the probability of parameter values is not included in the estimation process. $P(y|\theta)$ is the likelihood and describes the probability of having obtained the data y considering the prior knowledge about the parameter values. Finally, the term $P(y)$ represents the probability of the data that was obtained, irrespective of parameter values. It is called marginal likelihood or evidence. For the purposes of Bayesian cognitive modelling this term is negligible as it does not impact the relative values of the posterior and is often left out

As the posterior cannot be derived analytically it is numerically approximated by drawing a sufficiently large sample from it. By doing so it is possible to infer the main properties of the distribution even without knowing the explicit formula. To put it differently, Bayesian parameter estimation replaces the unknown posterior distribution with a large number of samples that can be analyzed as if they were the true posterior. Drawing samples from the posterior distribution is possible because we have access to the right side of equation 25. The prior is either a theoretically founded assumption or uninformative, while the likelihood of the observed data given the parameter values can be calculated from our data and the cognitive model. Given these terms, we can sample from the posterior using a class of algorithms called Markov Chain Monte Carlo (MCMC)[169]. In Bayesian cognitive modelling, algorithms start off at a plausible starting value that is provided. From this starting value a random walk is generated that is biased towards parameter values with a higher posterior probability. At each step the algorithm takes a guess based on a gaussian distribution (noise). Then this guess is compared to the current value. If the probability of the guess is higher or at least equal to the current value, the guess is accepted as the new value. If it is not, the guess is either discarded or still accepted with a small probability. After enough samples are drawn in this way, the algorithm will converge on the target distribution i.e., the posterior. To ensure that the posterior is correctly approximated, all samples drawn in the early phase of the random walk are discarded. This is called the burn-in period. There are however at least two issues this sampling algorithm might run into. The first problem is convergence. It might happen that the algorithm gets stuck in a parameter subspace it cannot break out of (local optimum) and therefore creates imperfect samples. One way to address this is to have the MCMC run several times from

different starting values and compare if the resulting posteriors match. The different runs are called chains and the convergence of these chains is an important diagnostic tool in Bayesian cognitive modelling. Chain convergence is commonly expressed as the R-hat statistic[170] comparing within-chain variance with between-chain variance. Values between 1 and 1.01 are mostly deemed acceptable. The second possible problem is autocorrelation. As the algorithm steps from sample to sample, two samples that are drawn close together will not be independent. Only samples that are far apart are essentially uncorrelated. A way to combat this autocorrelation is to only keep every x-th sample. However, in practice this problem can often be safely ignored.

To make the process of sampling from the posterior more efficient and stable with lower amounts of data it is possible to employ hierarchical models[15]. In hierarchical models there are multiple levels of parameters. Group level parameters (hyperparameters) describe distributions from which single-participant parameters are drawn. Estimating these together enables the data points from the different participants to inform each other and create more stable parameter estimates from fewer trials. More extreme single-participant parameter estimates are drawn towards the group mean by a degree governed by the width of the hyperparameter distribution determined by the variance between participants. This shrinkage towards the group mean reduces the effects of noise and stabilizes the parameter estimates. Moreover, hierarchical modelling enables meaningful group comparisons based on the hyperparameter estimates. Comparing the posterior distribution for the group mean, for instance, enables a qualified assessment of the evidence in favor or against a possible group difference. The evidence can be quantified in by the directional Bayes Factor (dBF). To achieve this, the posteriors of the hyperparameter of interest can be overlayed and the dBF defined as the probability mass above or below zero of the difference distribution between the two[171]. Thus, a dBF of 3 would imply that a more positive value for one of the groups is three times more likely than the reverse. When groups or conditions are compared this is a liberal value because it is based on testing against zero and not for directionality. Directional Bayes Factors are usually categorized similar to effect sizes[172]. A dBF of 3 is seen as moderate evidence and a value above 12 is seen as strong evidence. Directional Bayes Factors above 100 are considered extreme evidence.

Hierarchical Bayesian cognitive modelling in the projects presented here was done with the JAGS software package (version 4.3.0)[173] and the programming language R[174] as well as with python and the pystan toolbox and STAN (version 2.27)[175]. A final crucial part of the computational modeling of decision-making processes is the comparison between different

models describing the same data. Throughout the dissertation projects described here we used two different measures of model fit. First, we used the Deviance Information Criterion (DIC)[176] to compare relative model fits. Second, we employed the Watanabe-Akaike Information Criterion (WAIC)[177] using the loo-package (version 2.41) in the R programming language. In both measures, lower values represent better model fits.

Physiological cue-reactivity

Assessing cue-reactivity on a physiological level using brain imaging techniques such as functional magnetic-resonance imaging (fMRI) is at least difficult in the context of VR. There are VR headsets that allow VR presentation in a fMRI scanner, however, those are not commonly available yet. VR within scanners is usually presented via monitors and thus lacks proper immersion. Further impedance of the ecological validity of VR environments presented in a scanner is the lack of free exploration of the environment. Participants must stay still to not disturb image acquisition. A fairly easy and cheap method to quantify psychophysiological responses to stimulation is the measurement of physiological variables such as the EDA[43] and heart rate. Measuring these physiological indicators enables participant to fully explore a virtual environment without restraints on movement and immersion, while still enabling some conclusions about physiological processes involved. In the projects described here EDA and hear rate were obtained to quantify the level of physiological cue-reactivity. In the following section this text will discuss both in turn.

Electrodermal activity

The term EDA is an umbrella term for measures stemming from the assessment of the changes of the conductive properties of the skin in response to a constant current applied to the skin[43,178]. Biologically, the signal originates from the eccrine sweat glands located within the skin[179]. What makes these sudomotor glands so interesting is that they are only innervated by the sympathetic axis of the autonomous nervous system (ANS)[43,178,179]. The ANS regulates the body equilibrium and prepares the body for action or rest[180]. The two main axes of the ANS are the parasympathetic and the already mentioned sympathetic axis. Parasympathetic activity is mostly coupled to vegetative functions and rest. Sympathetic activity in contrast prepares the body for action and is consequently often linked to fight-or-flight responses. Central control of the ANS and the sympathetic system specifically is relayed over thermoregulatory posterior part of the hypothalamus and through the tegmental areas of the pons and the reticular nuclei of the medulla oblongata[178]. Over the years, electrodermal

responses have been shown to be closely linked to emotional states and appetitive conditioning[178,181–184]. Consequently, brain areas involved in these processes have been linked to the EDA. Most commonly found areas are frontal areas like the vmPFC, the dlPFC, the ACC as well as limbic areas including the amygdala, hippocampus and striatum[178,182–185]. In particular, the tight circuitry involving the NAcc, amygdala and vmPFC important for appetitive conditioning are relevant for the context of cue-reactivity as these two concepts are tightly related. Unsurprisingly, there is an abundance of research that reported electrodermal responses after addiction related stimuli were presented [29]. As mentioned above the EDA signal is comprised out of several semi-independent quantities (Figure 6)[43]. It can be roughly separated into the slowly varying tonic component and a fast phasic component. The tonic component consists mainly out of the skin conductance level (SCL), which is a slow indicator of general sympathetic arousal. The phasic component consists of skin conductive responses (SCRs). These SCRs can either be coupled to specific events or occur spontaneously without being tied to a specific event.

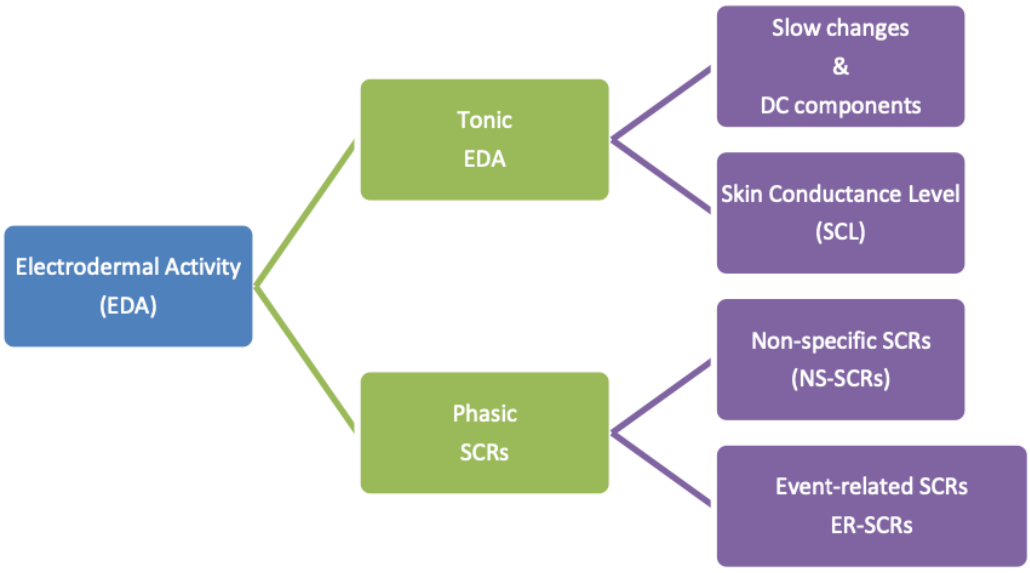


Figure 6. The electrodermal activity signal consists out of several semi-independent quantities. These are roughly grouped into slowly varying tonic components (SCL and electric components) and the fast-changing phasic components (SCR and non-specific SCRs/nSCRs). Figure adapted from Braithwaite and colleagues[43].

Both components of the signal can be quantified in different ways. The tonic general arousal of the sympathetic axis of the ANS can either be quantified as the part of the EDA signal that has been cleaned from SCRs and forms a baseline level of activity (i.e., the SCL) or the number of spontaneous non-specific SCRs (nSCR) within a specified time frame[186] (see below). Similarly, there are several possible measures of event-related SCRs of the phasic component[43] (Figure 7). Usually, an event related average (ERA) of a defined time window

around the stimulus is created. The resulting SCR curve (see Figure 7 for an example) can then be decomposed into a range of interesting parameters. First, the amplitude describes the difference between the peak of the SCR and the underlying baseline conductance. Second, the latency describes the time between the stimulus onset and the start of the SCR. Third, the rise time captures the time between the onset of the SCR and its peak. Finally, there is the half recovery time which is the time between the SCR's peak and the reduction to a value half the way to the baseline conductance.

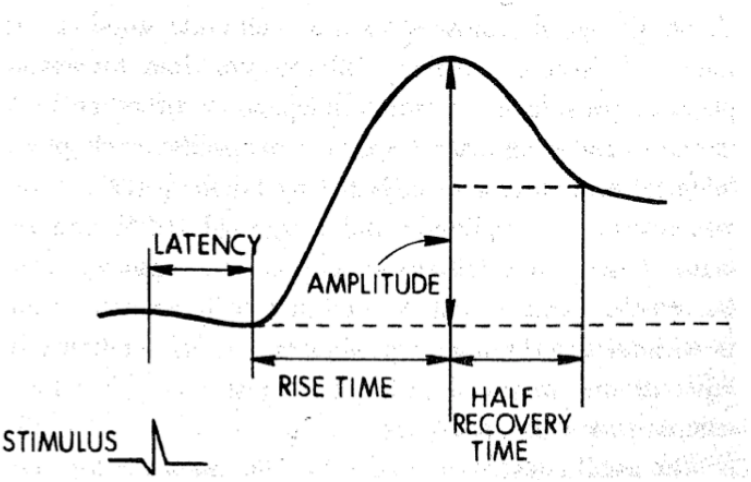


Figure 7. Illustration of an idealized SCR and its components. The latency describes the time from stimulus presentation to the onset of the SCR. The amplitude describes the distance between the peak of the SCR and the baseline activity, while the rise time describes the time from the onset of the SCR until the peak is reached. Finally, the half recovery time describes the time needed until the signal has recovered halfway back to the baseline activity. Figure adapted from Braithwaite and colleagues[43].

In the projects described in this thesis, we were mainly interested in the SCL as we had no time locked stimulus presentation but broader time windows in which participants were presented with different environments. To extract the signals that we need from the EDA signal we employed a technique called continuous decomposition analysis (CDA)[186]. In CDA it is assumed that an EDA signal is comprised out of a driver for the phasic activity, a driver for tonic activity and a so-called impulse response function (IRF). An IRF represents the biological response profile of the sweat gland to activity in the sympathetic fibers. The drivers present the nerve activity underlying the process of sweat secretion (Eq. 26). For the phasic driver this consists out of short bursts of many sympathetic fibers in concert, while the tonic driver reflects weaker but consistent firing.

$$EDA = (Driver_{Tonic} + Driver_{Phasic}) * IRF \quad (26)$$

CDA first deconvolves the EDA signal and identifies tonic and phasic drivers in turn with the help of a first estimate of an IRF. The resulting drivers are then assessed to check if the

deconvolution was successful. Criteria for this assessment are measures of mathematical plausibility of the resulting drivers. The process is repeated with an improved estimate of the IRF based on gradient descent until an acceptable goodness of fit is reached. Finally, the resulting drivers can be analyzed with common statistical methods. SCRs can for instance be detected with common peak detection algorithms. Within the scope of this project all CDA analyses were implemented with the ledalab toolbox[186] in Matlab (MathWorks).

Data acquisition was performed with a BioNomadix-PPGED wireless remote sensor together with a Biopac MP160 data acquisition system (Biopac Systems, Santa Barbara, CA, USA). A GSR100C amplifier module with a gain of 5V, low pass filter of 10 Hz and a high pass filter DC were included in the recording system. The system was connected to the acquisition computer running the AcqKnowledge software. Triggers for the events within the VR-environments were send to the acquisition PC via digital channels from the VR-PC. Disposable Ag/AgCl electrodes were attached to the thenar and hypothenar eminences of the non-dominant palm. Isotonic paste (Biopac Gel 101) was used to ensure optimal signal transmission. The signal was measured in micro-Siemens units (mS).

Heart rate

Like the EDA, the pulse frequency or heart rate is regulated by the ANS[180]. Contrary to the eccrine sweat glands forming the source of the EDA signal, however, the heart rate is regulated by both the sympathetic and parasympathetic axes of the ANS. Consequently, it forms a less clear measure of sympathetic physiological arousal than the EDA. Nevertheless, it has often been used in the context of cue-reactivity in addiction related disorders[29]. Acquisition of the heart rate signal was obtained with a photoplethysmogram (PPG)[187] optical transducer run with the BioNomadix-PPGED wireless remote sensor together with a Biopac MP160 data acquisition system (Biopac Systems, Santa Barbara, CA, USA). The setup used the same parameters as described for the acquisition of the EDA signal. PPG measures the heart rate by exploiting the reflective properties of blood to near infrared light. A PPG transducer placed on top of the skin close to capillary beds (e.g., fingertips, the ear lobe or thenar and hypothenar eminences of a hand) emits near infrared light and measures the reflected light via a detector. The amount of reflected near infrared light changes in accordance with capillary blood volume. In this way the waveform of the signal produced by the PPG transducer peaks when the capillary blood volume is at a maximum, i.e., when a heartbeat pumps blood into them. By counting the peaks occurring in a specific time window the signal can be converted to the standard measure beats per minute or short BPM.

Studies

Study 1

Study 1 summary. The recent emergence of high-performance VR technology makes it possible to create experimental designs that can examine the effects of contexts on cognitive processes realistically in a lab environment. This opportunity to create ecologically valid stimulation in highly controllable environments is extremely relevant for studies of psychiatric disorders, and especially in addiction-related disorders. However, before it is possible to confidently apply VR methods widely it must be established that commonly used behavioral tasks generate reliable data within VR environments. The aim of the first study of this dissertation project was therefore to establish the reliability and validity of data obtained within our VR environments tailored towards cue-reactivity in GD. Furthermore, we aimed to check if there are systematic effects caused by our VR environments in a group of healthy control participants. Finally, we aimed to assess whether sequential sampling models can be used to model behavioral data obtained in VR. The study was designed in a way to replicate the design of the second study presented here as much as possible, while adding a non-VR condition. The results showed good to excellent test-retest reliability estimates for the discount rate log(k), whereas they were poor to moderate for additional DDM parameters. Differences in the parameter estimates between standard lab testing and VR were mostly numerically small and of inconclusive directionality, indicating no systematic effects caused by VR exposure. Finally, exposure to VR generally increased tonic skin conductance measures irrespective of which VR environment participants explored. Taken together, *Study 1* demonstrated the reliability of temporal discounting parameters obtained in VR, and thereby setting the stage for *Study 2*. *Study 1* has been published in Scientific Reports in 2021[133].

scientific reports

Check for updates

OPEN

Reliability assessment of temporal discounting measures in virtual reality environments

Luca R. Bruder✉, Lisa Scharer & Jan Peters

In recent years the emergence of high-performance virtual reality (VR) technology has opened up new possibilities for the examination of context effects in psychological studies. The opportunity to create ecologically valid stimulation in a highly controlled lab environment is especially relevant for studies of psychiatric disorders, where it can be problematic to confront participants with certain stimuli in real life. However, before VR can be confidently applied widely it is important to establish that commonly used behavioral tasks generate reliable data within a VR surrounding. One field of research that could benefit greatly from VR-applications are studies assessing the reactivity to addiction related cues (cue-reactivity) in participants suffering from gambling disorder. Here we tested the reliability of a commonly used temporal discounting task in a novel VR set-up designed for the concurrent assessment of behavioral and psychophysiological cue-reactivity in gambling disorder. On 2 days, thirty-four healthy non-gambling participants explored two rich and navigable VR-environments (neutral: café vs. gambling-related: casino and sports-betting facility), while their electrodermal activity was measured using remote sensors. In addition, participants completed the temporal discounting task implemented in each VR environment. On a third day, participants performed the task in a standard lab testing context. We then used comprehensive computational modeling using both standard softmax and drift diffusion model (DDM) choice rules to assess the reliability of discounting model parameters assessed in VR. Test–retest reliability estimates were good to excellent for the discount rate log(k), whereas they were poor to moderate for additional DDM parameters. Differences in model parameters between standard lab testing and VR, reflecting reactivity to the different environments, were mostly numerically small and of inconclusive directionality. Finally, while exposure to VR generally increased tonic skin conductance, this effect was not modulated by the neutral versus gambling-related VR-environment. Taken together this proof-of-concept study in non-gambling participants demonstrates that temporal discounting measures obtained in VR are reliable, suggesting that VR is a promising tool for applications in computational psychiatry, including studies on cue-reactivity in addiction.

Recent research has exploited the development of high-performance virtual reality (VR) technology to increase the ecological validity of stimuli presented in studies of cue-exposure[1–3], counterconditioning[4], equilibrium training[5], social gazing[6] and gambling behavior in healthy control participants[7]. Furthermore, it has been shown to increase immersion and arousal during gambling games[8]. However, before VR can be widely applied with confidence it is important to establish that commonly applied behavioral tasks still yield reliable data in a VR context. Research focusing on psychiatric disorders, where one goal is to create reliable diagnostic markers based behavioral tasks and model-based computational approaches, would benefit from behavioral tasks that produce reliable parameters on a single participant level in VR.

A core characteristic of many psychiatric and neurological disorders is a detrimental change in decision-making processes. This is especially evident in addiction-related disorders such as substance abuse[9–11] or gambling disorder[12–14]. One approach to study such changes in decision making is computational psychiatry[15], which employs theoretically grounded mathematical models to examine cognitive performance in relation to psychiatric disorders. Such a model-based approach allows for a better quantification of the underlying latent processes[16].

One process that has been implicated in a range of psychiatric disorders is the discounting of reward value over time (temporal discounting): both steep and shallow discounting is associated with different psychiatric conditions[9]. In temporal discounting tasks, participants make repeated choices between a fixed immediate reward

Department of Psychology, Biological Psychology, University of Cologne, Cologne, Germany. ✉email: lbruder@uni-koeln.de

and larger but temporally delayed rewards[17]. Based on binary choices and/or response time (RT) distributions, the degree to which participants discount the value of future rewards based on the temporal delay provides a measure of individual impulsivity. Increased temporal discounting is thought to be a trans-diagnostic marker with relevance for a range of psychiatric disorders[9], with addictions and related disorders being prominent examples[18,19].

There is preliminary evidence that temporal discounting might be more pronounced when addiction related cues are present. Participants who suffer from gambling disorder for instance tend to exhibit steeper discounting[12,20] and increased risk-taking[21] in the presence of gambling-related stimuli or environments. These findings resonate with theories of drug addiction such as incentive sensitization theory[22] which emphasize a prominent role for addiction-related cues in the maintenance of drug addiction (see below). Identifying the mechanisms underlying such behavioral patterns and how they are modulated by addiction-related cues is essential to the planning and execution of successful interventions that aim to reverse these changes in decision-making[23,24].

Accordingly, the concept of cue-reactivity plays a prominent role in research on substance use disorders[25], but has more recently also been investigated in behavioral addictions such as gambling disorder[26]. Cue-reactivity refers to conditioned responses to addiction-related cues in the environment and is thought to play a major role in the maintenance of addiction. Cue-reactivity can manifest in behavioral measures, as described above for temporal discounting and risk-taking, but also in subjective reports and/or in physiological measures[25]. Incentive-Sensitization Theory[22,27] states that neural circuits mediating the incentive motivation to obtain a reward become over-sensitized to addiction-related cues, giving rise to craving. These motivational changes are thought to be mediated by dopaminergic pathways of the mesocorticolimbic system[28–30]. In line with this, craving following cue exposure correlates with a modulation of striatal value signals during temporal discounting[12], and exposure to drug-related cues increases dopamine release in striatal circuits in humans[30]. While studying these mechanisms in substance use disorders is certainly of value, it is also problematic because substances might have direct effects on the underlying neural substrates. Behavioral addictions, such as gambling disorder, however, might offer a somewhat less perturbed view on the underlying mechanisms.

Studies probing cue-reactivity in participants suffering from gambling disorder have typically either used picture stimuli[12,13,21,31–38] or real-life gambling environments (i.e. gambling facilities)[20]. Both methods come with advantages and disadvantages. While presenting pictures in a controlled lab environment enables researchers to minimize the influence of noise factors and simplifies the assessment physiological variables, it lacks the ecological validity of real-life environments. Conversely, a field study in a real gambling outlet arguably has high ecological validity but lacks the control of confounding factors and makes it difficult to obtain physiological measures.

By equipping participants with head-mounted VR-glasses and sufficient space to navigate within the VR-environment, a strong sense of immersion can be created, which in turn generates more realistic stimulation. In this way VR also offers a potential solution for the problem of ecologically valid addiction-related stimuli for studies in the field of cue-reactivity[7,8]. For example, Bouchard et al.[2] developed a VR-design that is built to provide ecologically valid stimuli for participants suffering from gambling disorder by placing them in a virtual casino. The design can be used in treatment in order to test reactions and learned cognitive strategies in a secure environment. The present study builds upon this idea to create a design that allows assessment of behavioral, subjective and physiological cue-reactivity in VR-environments. Participants are immersed in two rich and navigable VR environments that either represent a (neutral) café environment or a gambling-related casino environment. Within these environments, behavioral cue-reactivity can be measured via behavioral tasks implemented in VR. Given that immersion in the virtual environment takes place in a controlled lab setting, the measurement of physiological variables like electrodermal activity[39] and heart rate, as indicators of physiological cue-reactivity[25,26], is also easily accommodated.

Studies using computational modeling to asses latent processes underlying learning and decision-making increasingly include not only binary decisions, but also response times (RTs) associated with these decisions, e.g. via sequential sampling models such as the drift diffusion model (DDM)[40]. This approach has several potential advantages. First, leveraging the information contained in the full RT distributions can improve the stability of parameter estimates[41,42]. Second, by conceiving decision making as a dynamic diffusion process, a more detailed picture of the underlying latent processes emerges[43–47]. Recent studies, for instance, applied these techniques to temporal discounting, where they revealed novel insights into effects of pharmacological manipulation of the dopamine system on choice dynamics[46]. Likewise, we applied these techniques to examine the processes underlying reinforcement learning impairments in gambling disorder[48] and decision-making alterations following medial orbitofrontal cortex lesions[45]. Importantly, most standard lab-based testing settings use keyboards, button boxes and computer screens to record responses and display stimuli during behavioral tasks. In contrast, in the present study we used VR-controllers in a 3D virtual space. This represents a fundamentally different response mode, because in VR, participants have to physically move the controller to the location of the chosen option and then execute a button press to indicate their choice, adding additional motor complexity. In particular in the context of RT-based modeling, a crucial question is therefore whether responses obtained via VR-controllers allow for a comprehensive RT-based computational modeling, as previously done using standard approaches. Therefore, we also explored the applicability of drift diffusion modeling in the context of behavioral data obtained in VR.

Besides validating our VR-design with a healthy cohort of participants, the study at hand investigated the stability of parameters derived from temporal discounting tasks, in particular the discount rate $\log(k)$. Recently, the reliability of behavioral tasks as trait indicators of impulsivity and cognitive control has been called into question[49,50], in particular when compared to questionnaire-based measures of self-control[49]. It has been argued that the inherent property that makes behavioral tasks attractive for group-based comparisons renders them less reliable as trait markers[51]. Specifically, Hedge et al.[51] argue that tasks having a low between participant variability produce robust group effects in experimental studies and are therefore employed frequently. However, some of these tasks suffer from reduced test–retest-reliability for individual participants due to their low between-participant variability. Notably, Enkavi et al.[49] reported a reliability of 0.65 for the discount rate k, the highest

of all behavioral tasks examined in that study, and comparable to the reliability estimates of the questionnaire-based measures. This is in line with previous studies on the reliability of *k*, which provided estimates ranging from 0.7 to 0.77[52,53]. Importantly, as outlined above, both the actual response mode and the contextual setting of VR-based experiments differ substantially from standard lab-based testing situations employed in previous reliability studies of temporal discounting[49,52–55]. Therefore, it is an open question whether temporal discounting measures obtained in VR exhibit a reliability comparable to the standard lab-based tests that are typically used in psychology.

Taken together, by examining healthy non-gambling participants on different days and under different conditions (neutral vs. gambling-related VR environment, standard lab-based testing situation), we addressed the issue of reliability of temporal discounting in virtual versus standard lab environments. We furthermore explored the feasibility of applying the drift diffusion model in the context of RTs obtained via VR-compatible controllers. Finally, we also examined physiological reactivity during exploration of the different virtual environments. The specific virtual environments employed here are ultimately aimed to examine these processes in gambling disorder (e.g. the setup includes a gambling-related and a neutral cafe environment). However, the present study has more general implications for the application of behavioral and psychophysiological testing in virtual environments by examining the reliability of model-based analyses of decision-making in lab-based testing versus testing in different VR environments in a group of young non-gambling controls.

We hypothesized that the data produced on different days and under different conditions would yield only little evidence in favor of systematic shifts in temporal discounting behavior within a group of healthy non-gambling participants, suggesting only insubstantial effects caused by the different environments in our VR-design. Furthermore, we hypothesized that temporal discounting would show a strong reliability, adding further strength to the case that temporal discounting is stable over time and can be applied in VR. Finally, we hypothesized that we could capture latent decision variables in a VR context with the DDM.

Methods

Participants. Thirty-four healthy participants (25 female) aged between 18 and 44 (mean = 26.41, std = 6.44) were invited to the lab on three different occasions. Participants were recruited via flyers at the University of Cologne and via postings in local internet forums. No participant indicated a history of traumatic brain injury, psychiatric or neurological disorders or severe motion sickness. Participants were additionally screened for gambling behavior using the questionnaire Kurzfragebogen zum Glückspielverhalten (KFG)[56]. The KFG fulfills the psychometric properties of a reliable and valid screening instrument. No participant showed a high level (> 15 points on the KFG) of gambling affinity (mean = 1.56, std = 2.61, range: 0 to 13).

Participants provided informed written consent prior to their participation, and the study procedure was approved by the Ethics Board of the Germany Psychological Society. The procedure was in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

VR-setup. The VR-environments were presented using a wireless HTC VIVE head-mounted display (HMD). The setup provided a 110° field of view, a 90 Hz refresh rate and a resolution of 1440 × 1600 Pixel per eye. Participants had an area of about 6 m^2 open space to navigate the virtual environment. For the execution of the behavioral tasks and additional movement control participants held one VR-controller in their dominant hand. The VR-software was run on a PC with the following specifications: CPU: Intel Core i7-3600, Memory: 32.0 GB RAM, Windows 10, GPU: NVIDIA GeForce GTX 1080 (Ti). The VR-environments themselves were designed in Unity. Auditory stimuli were presented using on-ear headphones.

VR-environments. The two VR-environments both consisted of a starting area and an experimental area. The starting area was the same for both VR-environments. It consisted of a small rural shopping street and a small park. Participants heard low street noises. The area was designed for familiarization with the VR-setup and the initial exploration phase. The experimental area of the environments differed for the two environments. For the $VR_{neutral}$ environment it contained a small café with a buffet (Fig. 1a–c). Participants could hear low conversations and music. The gambling-related environment ($VR_{gambling}$) contained a small casino with slot machines and a sports betting area (Fig. 1d–f). The audio backdrop was the sound of slot machines and sports. The floorplan of both of these experimental areas was identical but mirrored for the café (Fig. 1a, d). Both experimental areas additionally included eight animated human avatars. These avatars performed steady and non-repetitive behaviors like gambling and ordering food for the gambling-related and neutral environments, respectively. Both experimental areas (café and casino) had entrances located at the same position within the starting area of the VR-environments, which were marked by corresponding signs.

Experimental procedure. Participants were invited to the VR lab for three different sessions on three different days. The time between the sessions was between one day and nineteen days (mean = 3.85, std = 3.36). During the three sessions participants either explored one of two different VR environments (VR-sessions) followed by the completion of two behavioral tasks, or simply performed the same two behavioral tasks in a standard lab-testing context (Lab-session). If the session was a VR-session, electrodermal activity (EDA)[39] was measured during a non-VR baseline period and the exploration of the VR-environments. The order of the sessions was pseudorandomized. At the first session, not depending on if VR was applied or not, participants arrived at the lab and the behavioral tasks were explained in detail. If the session was a Lab-session, participants proceeded with the two behavioral tasks. If the session was the first of the VR-sessions, participants were subsequently familiarized with the VR-equipment and handling. Participants were seated and a five-minute EDA baseline was measured (baseline phase). For both VR-sessions participants were then helped to apply the VR-equipment

Figure 1. Experimental areas of the VR-environments. (**a**) Floorplan of the café within the VR-neutral environment. (**b**) View of the main room of the café. (**c**) View of the buffet area of the café. (**d**) Floorplan of the casino within the VR-gambling environment. (**e**) View of the main room of the casino. (**f**) View of the sports bar within the casino.

and entered the VR-environments. Within the VR-environments participants first explored the starting area for 5 min (first exploration phase). After these five minutes participants were asked to enter the experimental area of the environment (either the café or the casino) (Fig. 1). Participants were instructed to explore the interior experimental area for five minutes (second exploration phase). Each of the three phases was later binned into five one-minute intervals and labeled as B (1 to 5) for the baseline phase, F (1 to 5) for the first exploration phase and S (1 to 5) for the second exploration phase. During the exploration the experimenter closely monitored the participants and alerted them if they were about to leave the designated physical VR-space. After the second exploration phase participants were asked to proceed to a terminal within the VR-environment on which the behavioral tasks were presented.

Physiological measurements. EDA was measured using a BioNomadix-PPGED wireless remote sensor together with a Biopac MP160 data acquisition system (Biopac Systems, Santa Barbara, CA, USA). A GSR100C amplifier module with a gain of 5 V, low pass filter of 10 Hz and a high pass filter DC were included in the recording system. The system was connected to the acquisition computer running the AcqKnowledge software. Triggers for the events within the VR-environments were send to the acquisition PC via digital channels from the VR-PC. Disposable Ag/AgCl electrodes were attached to the thenar and hypothenar eminences of the non-dominant palm. Isotonic paste (Biopac Gel 101) was used to ensure optimal signal transmission. The signal was measured in micro-Siemens units (mS).

Behavioral tasks. Participants performed the same two behavioral tasks with slightly varied rewards and choices in each of the three sessions: a temporal discounting task[17] and a 2-step sequential decision-making task[57,58]. Results from the 2-step task will be reported separately. In the temporal discounting task participants had to repeatedly choose between an immediately available (smaller-but-sooner, SS) monetary reward of 20 Euros and larger-but-later (LL) temporally delayed monetary rewards. The LL options were multiples of the SS option (range 1.025 to 3.85) combined with different temporal delays (range 1 to122 days). We constructed three sets of six delays and 16 LL options. Each set had the same mean delay and the same mean LL option. Combining each delay with every LL option within each set resulted in three sets of 96 trials. The order of presentation of the trial sets was counter balanced across participants and sessions. All temporal discounting decisions were hypothetical[59,60]. In the VR-version of the task two yellow squares were presented to the participants (Fig. 2). One depicted the smaller offer of 20 Euros now, while the other depicted the delayed larger offer. For the lab-based testing session were presented in the same way except that the color scheme was white writing on a black background. Offers were randomly assigned to the left/right side of the display and presented until a decision was made. The next trial started 0.5 to 1 s after the decision. Participants indicated their choice either by aiming the VR-controller at the preferred option and pulling the trigger (VR-sessions) or by pressing the corresponding arrow key on the keyboard (Lab-session).

Model-free discounting data analysis. The behavioral data from the temporal discounting task was analyzed using several complementary approaches. First, we used a model-free approach that involved no a priori hypotheses about the mathematical shape of the discounting function. For each delay, we estimated the LL

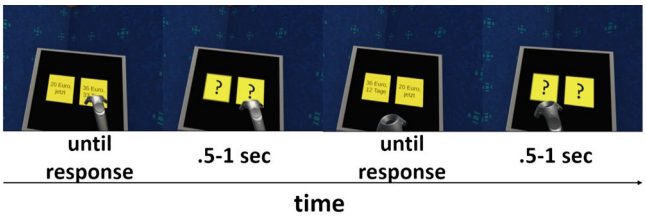


Figure 2. Presentation of the temporal discounting task in VR. Participants had to repeatedly decide between a small but immediate reward (SS) and larger but temporally delayed rewards (LL). Amounts and delays were presented in yellow squares. During the inter-trial intervals (.5–1 s) these squares contained only question marks. Participants indicated their choice by pointing the VR-controller at one of the yellow squares and pulling the trigger.

reward magnitudes at which the subjective value of the LL reward was equal to the SS (indifference point). This was done by fitting logistic functions to the choices of the participants, separately for each delay. Subsequently, these indifference points were plotted against the corresponding delays, and the area under the resulting curve (AUC) was calculated using standard procedures[61]. AUC values were derived for each participant and testing session, and further analyzed with the intra-class correlation (ICC) and the Friedman Test, a non-parametric equivalent of the repeated measures ANOVA model.

Computational modeling. Previous research on the effects of the delay of a reward on its valuation proposed a hyperbolic nature of devaluation[62,63]. Therefore, the rate of discounting for each participant was also determined employing a cognitive modeling approach using hierarchical Bayesian modeling[16]. A hierarchical model was fit to the data of all participants, separately for each session (see below). We applied a hyperbolic discounting model (Eq. 1):

$$\mathrm{SV(LL_t)} = \frac{\mathrm{At}}{\left(1 + \exp(\mathrm{k}) * \mathrm{Dt}\right)} \tag{1}$$

Here, SV(LL) denotes the subjective (discounted) value of the LL. A and D represent the amount and the delay of the LL, respectively. The parameter k governs the steepness of the value decay over time, with higher values of k indicating steeper discounting of value over time. As the distribution of the discount rate k is highly skewed, we estimated the parameter in log-space (log[k]), which avoids numerical instability in estimates close to 0.

The hyperbolic model was then combined with two different choice rules, a softmax action selection rule[64] and the drift diffusion model[44]. For softmax action selection, the probability of choosing the LL option on trial t is given by Eq. (2).

$$\mathrm{P(LL_t)} = \frac{\exp\left(\mathrm{SV_{LL_t}} * \beta\right)}{\exp\left(\mathrm{SV_{SS_t}} * \beta\right) + \exp\left(\mathrm{SV_{LL_t}} * \beta\right)} \tag{2}$$

Here, the β-parameter determines the stochasticity of choices with respect to a given valuation model. A β of 0 would indicate that choices are random, whereas higher β values indicate a higher dependency of choices on option values. The resulting best fitting parameter estimates were used to test the ICC and systematic session effects via comparison of the posterior probabilities of group parameters.

Next, we incorporated response times (RTs) into the model by replacing the softmax choice rule with the drift diffusion model (DDM)[43–46]. The DDM models choices between two options as a noisy evidence accumulation that terminates as soon as the accumulated evidence exceeds one of two boundaries. In this analysis the upper boundary was set to represent LL choices, and the lower boundary SS choices. RTs for choices of the immediate reward were multiplied by − 1 prior to model estimation. To prevent outliers in the RT data from negatively impacting model fit, the 2.5% slowest and fastest trials of each participant were excluded from the analysis[44,45]. In the DDM the RT on trial t is distributed according to Wiener first passage time (wfpt) (Eq. 3).

$$\mathrm{RTt} \sim \mathrm{wfpt}\,(\alpha, \tau, \mathrm{z}, \nu) \tag{3}$$

Here α represents the boundary separation modeling the tradeoff between speed and accuracy. τ represents the non-decision time, reflecting perception and response preparation times. The starting value of the diffusion process is given by z, which therefore models a potential bias towards one of the boundaries. Finally, rate of evidence accumulation is given by the drift-rate ν.

We first fit a null model ($\mathrm{DDM_0}$), where the value difference between the two options was not included, such that DDM parameters were constant across trials[45,46]. We then used two different temporal discounting $\mathrm{DDM_S}$, in which the value difference between options modulated trial-wise drift rates. This was done using either a linear ($\mathrm{DDM_L}$) or a non-linear sigmoid ($\mathrm{DDM_S}$) linking function[47]. In the $\mathrm{DDM_L}$, the drift-rate ν in each trial is linearly dependent on the trial-wise scaled value difference between the LL and the SS options (Eq. 4)[44]. The parameter ν_{coeff} maps the value differences onto ν and scales them to the DDM:

$$\nu_t = \nu_{\mathrm{coeff}} * (SV(LL_t) - SV(SS_t)) \tag{4}$$

Parameter	Prior for group mean
log(k)	*Uniform*(− 20, 3)
softmax β	*Uniform*(0, 10)
ν	*Uniform*(− 100, 100)
τ	*Uniform*(.1, 6)
α	*Uniform*(.01, 5)
z	*Uniform*(.1, .9)
ν_{coeff}	*Uniform*(− 100, 100)
ν_{max}	*Uniform*(0, 100)

Table 1. Ranges for the uniform priors of group-level parameter means. Ranges were chosen to cover numerically plausible values. Parameters included in multiple models are only listed once.

One drawback of a linear representation of the relationship between the drift-rate ν and trial-wise value differences is that ν might increase infinitely with high value differences, which can lead the model to under-predict RTs for high value differences[45]. In line with previous work[45,46] we thus included a third version of the DDM, that assumes a non-linear sigmoidal mapping from trial-wise value differences to drift rates (Eqs. 5 and 6)[43]:

$$\nu t = S\left(\nu_{coeff} * (SV(LL_t) - SV(SS_t))\right) \tag{5}$$

$$S(m) = \frac{2 * \nu_{max}}{1 + \exp(-m)} - \nu_{max} \tag{6}$$

Here, the linear mapping function from the DDM_L is additionally passed through a sigmoid function S with the asymptote ν_{max}, causing the relationship between ν and the scaled trail-wise value difference m to asymptote at ν_{max}.

We have previously reported detailed parameter recovery analyses for the DDM_S in the context of value-based decision-making tasks such as temporal discounting[45], which revealed that both subject-level and group-level parameters recovered well.

Hierarchical Bayesian models. All models were fit to the data of all participants in a hierarchical Bayesian estimation scheme, separately for each session, resulting in independent estimates for each participant per session. Participant-level parameters were assumed to be drawn from group-level Gaussian distributions, the means and precisions of which were again estimated from the data. Posterior distributions were estimated via Markov Chain Monte Carlo in the R programming language[65] using the JAGS software package[66]. For the DDM's the Wiener module for JAGS was used[67]. For the group-level means, uniform priors over numerically plausible parameter ranges were chosen (Table 1). Priors for the precision of the group-level distribution were Gamma distributed (0.001, 0.001). The convergence of chains was determined by the R-hat statistic[68]. Values between 1 and 1.01 were considered acceptable. Comparisons of relative model fit were performed using the Deviance Information Criterion (DIC), where lower values reflect a superior model fit[69].

Systematic session effects on model parameters. Potential systematic session effects on group level posterior distributions of parameters of interest were analyzed by overlaying the posterior distributions of each group level parameter for the different sessions. Here we report the mean of the posteriors of the estimated group level parameters and the difference distributions between them, the 95% highest density intervals (HDI) for both of these as well as directional Bayes Factors (dBF) which quantify the degree of evidence for reductions versus increases in a parameter. Because the priors for the group effects are symmetric, this dBF can simply be defined as the ratio of the posterior mass of the difference distributions above zero to the posterior mass below zero[70]. Here directional Bayes Factors above 3 are interpreted as moderate evidence in favor of a positive effect, while Bayes Factors above 12 are interpreted as strong evidence for a positive effect[71]. Specifically, a dBF of 3 would imply that a positive directional effect is three times more likely than a negative directional effect. Bayes Factors below 0.33 are likewise interpreted as moderate evidence in favor of the alternative model with reverse directionality. A dBF above 100 is considered extreme evidence[71]. The cutoffs used here are liberal in this context, because they are usually used if the test is against a H_0 implying an effect of 0. In addition, we report the effect size (Cohen's d) based on the mean posterior distributions of the session means, the pooled standard deviations across sessions and the correlation between sessions.

ICC analysis. The test–retest reliability of the best fitting parameter values between the three sessions was analyzed using the intra-class correlation coefficient (ICC). The ICC-analysis was done in the R programming language[65] and was based on a mean-rating of three raters, absolute agreement and a two-way mixed model. ICC values below 0.5 are an indication of poor test–retest reliability, whereas values in the range between 0.5 and 0.75 indicate a moderate test–retest reliability[72]. Higher values between 0.75 and 0.9 indicate a good reliability, while values above 0.9 suggest an excellent test–retest reliability.

Session	Log(k)					β				
	Mean	HDI		dBF	d	Mean	HDI		dBF	d
Lab	−4.083	−4.643	−3.530	–	–	.417	.355	.489	–	–
$VR_{neutral}$	−4.348	−4.912	−3.797	–	–	.577	.461	.714	–	–
$VR_{gambling}$	−4.274	−4.882	−3.687	–	–	.448	.363	.547	–	–
Lab-$VR_{neutral}$	.266	−.520	1.054	2.712	.38	−.16	−.31	−.024	.01	.9
Lab-$VR_{gambling}$	.191	−.620	1.01	2.162	.3	−.03	−.148	.081	.446	.18
$VR_{gambling}$-$VR_{geutral}$	.074	−.746	.885	1.264	.1	−.129	−.29	.023	.048	.56

Table 2.. 95% HDIs for the two parameters of the hyperbolic discounting model. HDIs are described by the min. value first and the max value second. Directional Bayes Factors (dBF) are calculated as BF = i/(1-i), with i being the probability mass of the difference distributions above zero. Effect sizes are given as Cohen's d.

Analysis of physiological data. A frequently used index of sympathetic activity is electrodermal activity, i.e. changes in skin conductance (SC)[73]. Here the physiological reactivity to the VR-environments is measured as the slowly-varying skin conductance level (SCL)[39]. Thus, the SCL was extracted from the EDA signal using continuous decomposition analysis (CDA) via the Ledalab toolbox[74] for Matlab (MathWorks). For the deconvolution, default settings were used. The resulting signal was then transformed into percentage change from the mean signal of the five minutes baseline phase at the beginning of the experiment. Subsequently, five one-minute bins were constructed for each phase of the VR-session (baseline phase, the first exploration phase and the second exploration phase). An alternative way of classifying tonic sympathetic arousal can be the number of spontaneous phasic responses (SCR) in the EDA signal[74]. Again, the signal was divided in one-minute bins and the number of spontaneous SCRs during each bin was calculated from the phasic component of the deconvoluted EDA signal using the Ledalab toolbox. The resulting values were similarly transformed into percentage change from the mean number of SCRs during the five baseline bins. To test whether entering the VR-environments had a general effect on sympathetic arousal, we compared the values for the last time point of the base line phase (B5) with the first time point of the first exploration phase (F1) for both sessions using a non-parametric Wilcoxon Signed-Rank Test. To test whether there was a differential effect of entering the different experimental areas of the VR-environments on sympathetic arousal, for both measures the differences between the last time point of the first exploration phase (F5) and the first time point of the second exploration phase (S1) were compared across VR-sessions using a non-parametric Wilcoxon Signed-Tanks Test[75]. Effect sizes are given as r[76], computed as the statistic Z divided by the square-root of N. Effect sizes between 0 and 0.3 are considered small and effect sizes between 0.3 and 0.5 are considered medium and r values > 0.5 are considered large effects.

Data and code availability. Raw behavioral and physiological data as well as JAGS model code is available on the Open Science Framework (https://osf.io/xkp7c/files/).

Results

Temporal discounting AUC. The analysis of the AUC values revealed no significant session effect across participants (Friedman Test: Chi-Squared = 1.235 df = 2 p = 0.539). Furthermore, the ICC value was 0.93 (95% confidence interval (CI): 0.89–0.96) (p < 0.001) indicating an excellent test–retest reliability of temporal discounting AUC values over the three sessions (Table 2). Pairwise correlations between all sessions can be found in the supplementary materials (Supplementary Fig. S1).

Softmax choice rule. For the hyperbolic model with softmax choice rule, the group level posteriors showed little evidence for systematic effects of the different sessions on log(k) (all BFs < 3 or > 0.33) (Fig. 3a, c and Table 2). In contrast, the softmax β parameter was higher (reflecting higher consistency) in the $VR_{neutral}$ session compared to the other sessions (vs. Lab: dBF = 0.01 and vs. $VR_{gambiling}$: dBF = 0.048) (Fig. 3b, d and Table 2). This indicates that a higher β in the $VR_{neutral}$ session was approximately 100 (Lab) or 20 ($VR_{gambling}$) times more likely than a lower β. There was little evidence for a systematic effect between the Lab and $VR_{gambling}$ sessions (dBF = 0.446).

The ICC value for the log(k) parameter indicated an excellent test–retest reliability of 0.91 (CI: 0.86–0.96) (p < 0.001) (Table 3). For the β-parameter of the softmax choice rule the ICC value was 0.34 (CI: 0.17–0.53) (p < 0.001) indicating a poor test–retest reliability (Table 3). The pairwise correlations of estimated parameter values between all sessions can be found in the Supplement (Supplementary Figs. S2 and S3). Pairwise correlations between all sessions for both parameters can be found in the supplementary materials (Supplementary Figs. S2 and S3).

Drift diffusion model choice rule. Model comparison revealed that the DDM_S had the lowest DIC in all conditions (Table 4) replicating previous work[45,46,48]. Consequently, further analyses of session effects and reliability focused on this model. For the log(k) parameter, the 95% HDIs showed a high overlap between all sessions indicating no systematic session effects, however the BFs showed moderate evidence for a reduced log(k) in the $VR_{neutral}$-session (Fig. 4a, d and Table 5). A lower value in the $VR_{neutral}$-session was about seven (Lab-session dBF = 6.756) or four times ($VR_{gambling}$ dBF = 3.86) more likely than a lower value. Similarly, the posterior

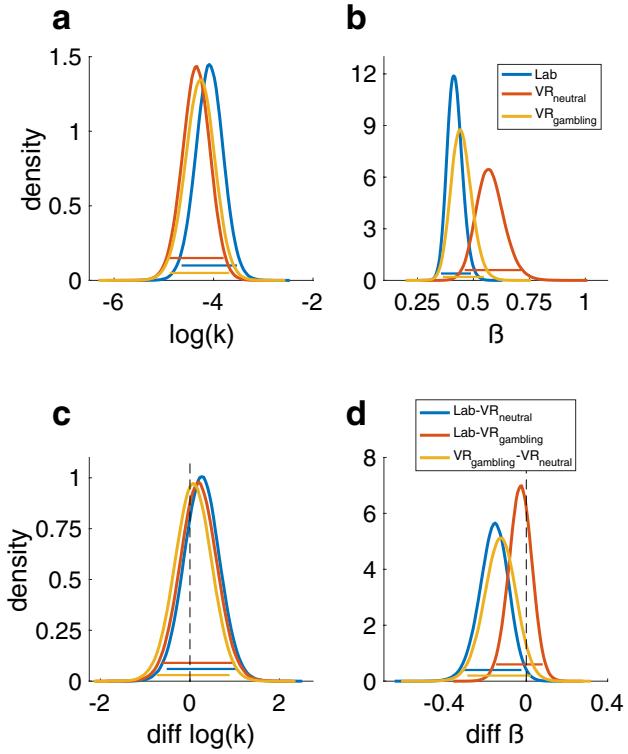


Figure 3. Posterior distributions of the parameters of the hyperbolic discounting model. Colored bars represent the corresponding 95% HDIs. (**a**) Posterior distribution of the log(k) parameter (reflecting the degree of temporal discounting) for all three sessions. (**b**) Posterior distribution of the β or inverse temperature parameter (reflecting decision noise). (**c**) Pairwise difference distributions between the posteriors of the log(k) parameters of all three sessions. (**d**) Pairwise difference distributions between the posteriors of the β parameters of all three sessions.

Parameter	ICC	*p*	Lower bound	Upper bound
AUC	.93	< .001	.89	.96
log(k)	.91	< .001	.86	.95
β	.34	< .001	.17	.53

Table 3. Summary of the results of the ICC analysis for the AUC values as well as the two parameters of the hyperbolic discounting model with a softmax choice rule. Lower and upper bound describe the 95% confidence interval.

Model	Lab	$VR_{neutral}$	$VR_{gambling}$	Rank
DDM_0	9275.7	9569.8	9225.7	3
DDM_L	7558.9	7921.4	7663.0	2
DDM_S	6992.3	7327.2	7033.1	1

Table 4. Summary of the DICs of all DDM models in all sessions. Ranks are based on the lowest DIC in all sessions.

distributions of v_{max}, v_{coeff} and α were highly overlapping, whereas some of the dBFs gave moderate evidence for systematic directional effects within these parameters (Figs. 4b, c, e, f, and 5b, e, Table 5). v_{coeff}, mapping trial-wise value difference onto the drift rate, was lowest in the Lab-session and highest in $VR_{neutral}$ (Lab-$VR_{neutral}$ dBF = 0.074, Lab-$VR_{gambling}$ = 0.2, $VR_{gambling}$-$VR_{neutral}$ = 0.228). Thus, an increase in v_{coeff} in $VR_{neutral}$ compared to the Lab-session was approximately thirteen times more likely than a decrease. Likewise, it was approximately five times more likely that there was an increase in the $VR_{neutral}$ compared to the $VR_{gambling}$-session. For v_{max},

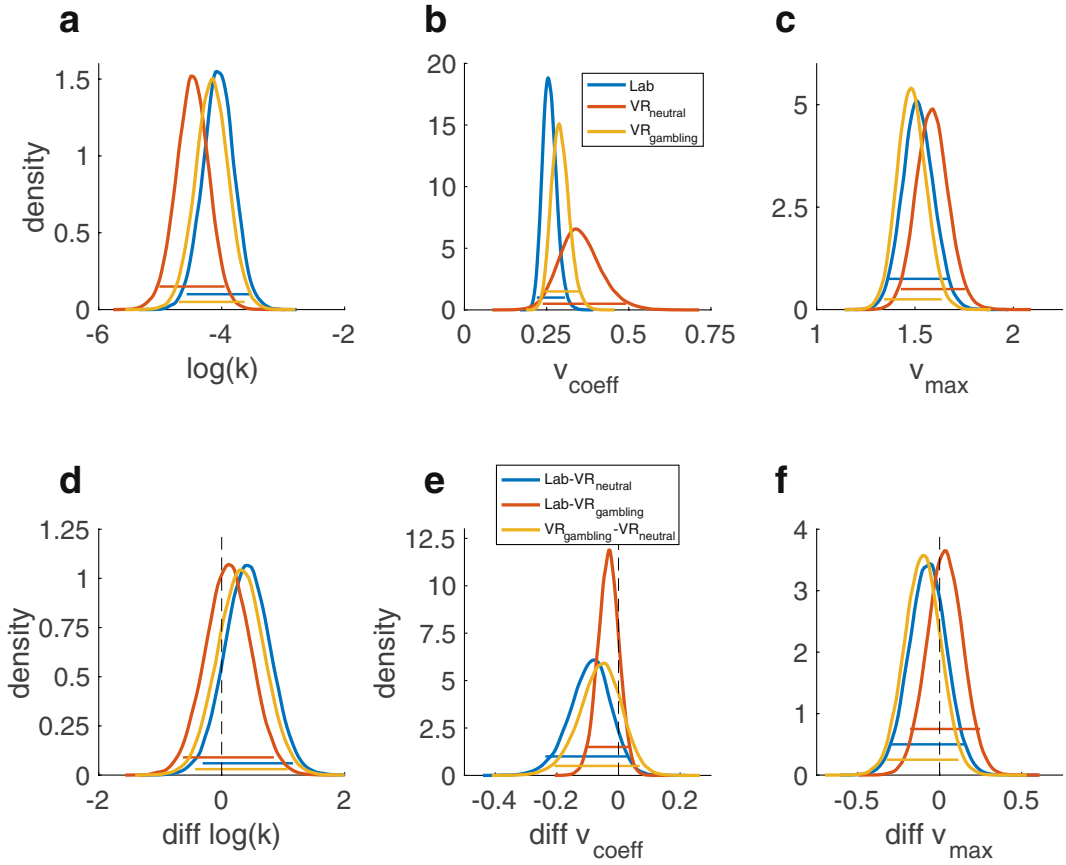


Figure 4. Posterior distributions of the parameters of the DDM_S model. Colored bars represent the corresponding 95% HDIs. (**a**) Posterior distributions of the log(k) parameter for all three sessions. (**b**) Posterior distributions of the v_{coeff} parameter (mapping the drift rate onto the trial wise value difference). (**c**) Posterior distributions of the v_{max} parameter (setting an asymptote for the relation between the trial wise value difference and the drift rate). (**d**) Pairwise difference distributions between the posterior distributions of the log(k) parameters of the three sessions. (**e**) Pairwise difference distributions between the posterior distributions of the v_{coeff} parameters of the three sessions. (**f**) Pairwise difference distributions between the posterior distributions of the v_{max} parameters of the three sessions.

	log(k)		v_{coeff}		v_{max}		τ		α		z	
Contrast	**dBF**	**d**	**dBF**	**d**	**dBF**	**d**	**dBF**	**d**	**dBF**	**d**	**dBF**	**d**
Lab-$VR_{neutral}$	6.756	.37	.074	.37	.377	.2	> 100	1.2	.255	.224	.530	.2
Lab-$VR_{gambling}$	1.679	.19	.200	.59	1.573	.09	> 100	1.5	.358	.160	1.118	.04
$VR_{gambling}$-$VR_{neutral}$	3.860	.29	.228	.27	.203	.34	3.413	.17	.629	.070	.458	.2

Table 5. Directional Bayes factors (dBF) and effect sizes (Cohen's d) for all between session comparisons for all parameters of the DDM_S. Means and HDIs of the posteriors and difference distributions are summarized in the supplementary materials (Supplementary Table S1). BFs are calculated as BF = i/(1 − i), with i being the probability mass of the difference distributions above zero.

the upper boundary for the value difference's influence on the drift rate, the dBFs indicated that a positive shift from $VR_{gambling}$ to $VR_{neutral}$ was five times more likely than a negative shift (dBF = 0.203) but there was only very little indication of a systematic difference between both of them and the Lab-session. Finally, a reduction of the boundary separation parameter α was five times more likely than an increase when comparing the $VR_{neutral}$ to the Lab-session (dBF = 0.255). There was little evidence for any other systematic differences. The bias parameter *z* displayed high overlap in HDIs and little evidence for any systematic effects between sessions (all dBFs > 0.33 or < 3) (Fig. 5c, f and Table 5). For the non-decision time parameter τ there was extreme evidence for an increase in the VR-sessions compared to the Lab-session (both dBFs > 100), reflecting prolonged motor and/or perceptual components of the RT that was more than 100 times more likely than a shortening of these components (Fig. 5a, d and Table 5).

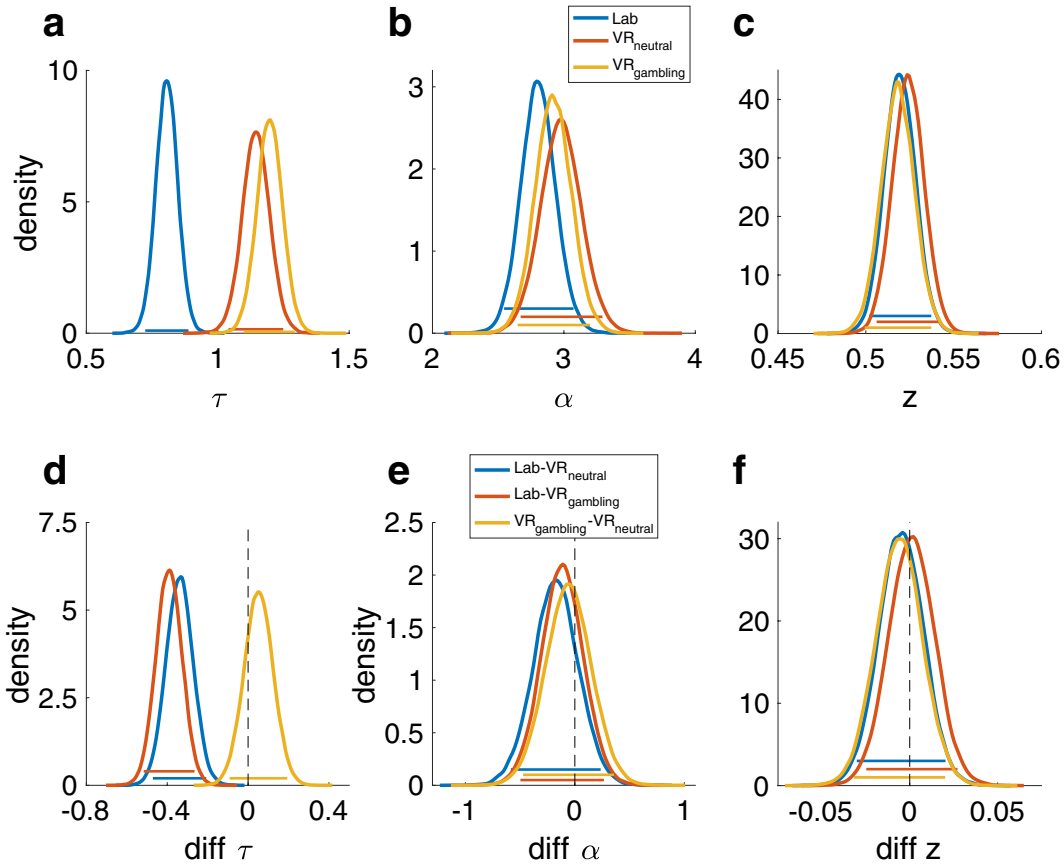


Figure 5. Posterior distributions of the remaining parameters of the DDM_S model. Colored bars represent the corresponding 95% HDIs. (**a**) Posterior distributions of the τ parameter (non-decision time) for all three sessions. (**b**) Posterior distributions of the α parameter (separation between decision boundaries). (**c**) Posterior distributions of the z parameter (bias towards one decision option). (**d**) Pairwise difference distributions between the posterior distributions of the τ parameters of the three sessions. (**e**) Pairwise difference distributions between the posterior distributions of the α parameters of the three sessions. (**f**) Pairwise difference distributions between the posterior distributions of the z parameters of the three sessions.

Parameter	ICC	*p*	Lower bound	Upper bound
log(k)	.7	< .001	.56	.8
v_{coeff}	.11	.14	− .053	.3
v_{max}	.33	< .001	.16	.52
τ	.19	.033	.019	.38
α	.42	< .001	.24	.59
z	.4	< .001	.22	.58

Table 6. Summary of the results of the ICC analysis of the DDM_S parameters.

The ICC value for the log(k) parameter was 0.7 (CI: 0.56–0.8) indicating a moderate test–retest-reliability (Table 5). For the other DDM_S parameters, ICC values were substantially lower (Table 6). Pairwise correlations between all sessions for all parameters can be found in the supplementary materials (Supplementary Figs. S4–S9).

Split-half reliability control analyses for DDM parameters. In light of the lower ICC values for the DDM_S parameters beyond log(*k*), we ran additional analyses. Specifically, we hypothesized that these lower ICC values might be attributable to fluctuations of state factors, e.g. mood, fatigue or motivation, between the different sessions. Therefore, we explored within-session reliability of these parameters, separately for each session. Trials where split into odd and even trials and modelled separately using the DDM_S, as described above. In general, within-session split-half reliability was substantially greater than test–retest reliability, and mostly in a good to excellent range (range: − 0.1 for v_{coeff} in $VR_{gambling}$ to 0.94 for τ in $VR_{neutral}$). The lower test–retest reliabilities of some of the DDM_S parameters are therefore unlikely to be due to the specifics of the parameter estimation procedure. Rather, these findings are compatible with the view that the parameters underlying the evidence accumula-

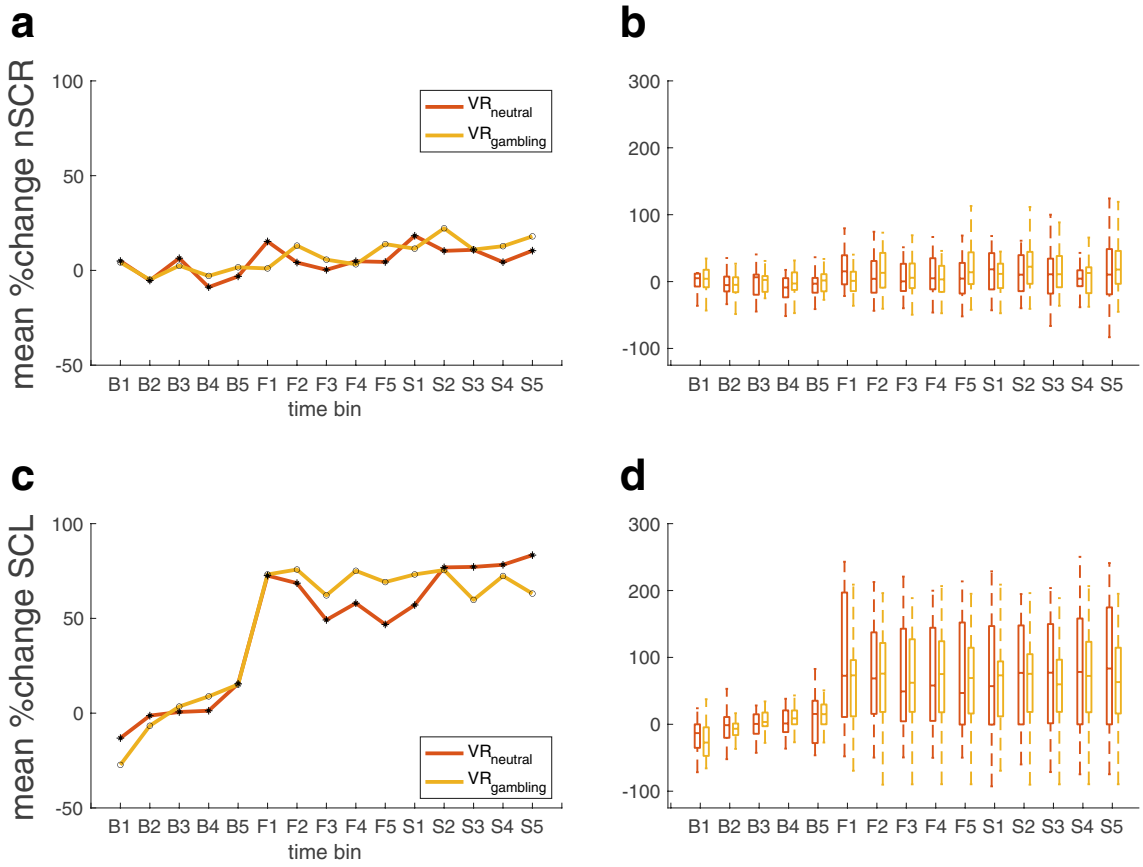


Figure 6. Results of the EDA measurements divided into 15 time points over the course of the baseline phase, measured before participants entered the VR-environments, and the first and second exploration phases. Each of the three phases is divided into five one-minute bins (B1-5: pre-VR baseline, F1-5: first exploration phase in VR, S1-5: second exploration phase VR). (**a**) Median percent change from baseline mean for no. of spontaneous SCRs over all participants. (**b**) Boxplot of percentage change from baseline mean for no. spontaneous SCRs over all participants. (**c**) Median percent change from baseline mean of SCL over all participants. (**d**) Boxplots of percentage change from base line mean of SCL over all participants.

tion process might be more sensitive to state-dependent changes in mood, fatigue or motivation. Full results for the split-half reliability analyses can be found in the supplementary materials (Supplementary Tables S3–S5).

Electrodermal activity (EDA). The data of 8 of the 34 participants had to be excluded from the EDA analysis, due to technical problems or missing data during one of the testing sessions. Physiological reactivity in the remaining 26 (18 female) participants was analyzed by converting the SCL signal as well as the nSCRs into percent change from the mean level during the base line phase. Both signals were then binned into five one-minute intervals for each of the three phases (baseline, first exploration and second exploration phase). All comparisons were tested with the Wilcoxon Signed Rank Test. Entering the VR-environments (comparing bin B5 to bin F1 for both environments individually) resulted in a significant increase in the SCL values for both VR-environments ($VR_{neutral}$: $Z = -3.67$, $p < 0.001$, $r = 0.72$; $VR_{gambling}$: $Z = -3.543$, $p = 0.002$, $r = 0.695$) (Fig. 6c, d). The effect was large in both sessions ($r > 0.5$). However, for the number of spontaneous SCRs (nSCRs), this effect was only significant in the neutral VR-environment (neutral: $Z = -2.623$, $p = 0.009$, $r = 0.515$; gambling: $Z = -0.013$, $p = 0.99$, $r = 0.002$). There was no significant difference between the two sessions, but the effect was of medium size ($Z = -1.7652$, $p = 0.078$, $r = 0.346$) (Fig. 6a, b). To test whether entering the specific experimental areas of the two VR-environments (virtual café vs. virtual casino) had differential effects on physiological responses, the increase in sympathetic arousal from the end of the first exploration phase to the start of the second exploration phase was examined (comparing bin F5 to bin S1, see Fig. 6b, d). The SCL (neutral: $Z = -0.7238$, $p = -0.469$, $r = 0.142$; gambling: $Z = -0.089$, $p = 0.929$, $r = 0.017$) as well as the nSCRs (neutral: $Z = -1.943$, $p = 0.052$, $r = 0.381$; gambling: $Z = 0.982$, $p = 0.326$, $r = 0.193$) assessed for each session individually showed no significant effect. The effect size was medium ($r = 0.381$) for the nSCRs of the $VR_{neutral}$-session and small for all other comparisons ($r < 0.3$). Furthermore, the Wilcoxon Signed-Ranks test indicated no significant differences between the two experimental areas on both sympathetic arousal measures (SCL: $Z = -0.572$, $p = 0.381$, $r = 0.11$; nSCRs.: $Z = -1.7652$, $p = 0.078$, $r = 0.346$) (Fig. 6b, d). For the nSCRs however, the effect was of a medium size ($r = 0.346$).

Discussion

Here we carried out an extensive investigation into the reliability of temporal discounting measures obtained in different virtual reality environments as well as standard lab-based testing. This design allowed us the joint assessment of physiological arousal and decision-making, an approach with potential applications to cue-reactivity studies in substance use disorders or behavioral addictions such as gambling disorder. Participants performed a temporal discounting task within two different VR-environments (a café environment and a casino/sports betting environment: $VR_{neutral}$ vs. $VR_{gambling}$) as well as in a standard computer-based lab testing session. Exposure to VR generally increased sympathetic arousal as assessed via electrodermal activity (EDA), but these effects were not differentially modulated by the different VR environments. Results revealed good to excellent test–retest reliability of model-based ($\log(k)$) and model-free (AUC) measures of temporal discounting across all testing environments. However, the DDM_S parameters modelling latent decision processes showed substantially lower test–retest reliabilities between the three sessions. The split-half reliability within each session was mostly good to excellent indicating that the lower test–retest reliability was likely caused by the participants current state and not by factors within the modelling process itself.

To test how well temporal discounting, as a measure of choice impulsivity, performs in virtual environments we implemented a VR-design that is built for possible future application in a cue-reactivity context. Healthy controls displayed little evidence for systematic differences in choice preferences between the Lab-session and the VR-sessions. This was observed for model-free measures (AUC), as well as the $\log(k)$ parameter of the hyperbolic discounting model with the softmax choice rule and the drift diffusion model with non-linear drift rate scaling (DDM_S). Model comparison revealed that the DDM_S accounted for the data best, confirming previous findings[43,45,46,48]. Although generally, discount rates assessed in the three sessions were of similar magnitude, in the DDM_S there was moderate evidence for reduced discounting (i.e., smaller values of $\log(k)$) in the $VR_{neutral}$ session. The reasons for this could be manifold. One possibility is that environmental novelty plays a role, such that perceived novelty of the $VR_{neutral}$ session might have been lower than for the $VR_{gambling}$ and Lab-sessions. Exposure to novelty can stimulated dopamine release[77], which is known to impact temporal discounting[78]. Nonetheless, effect sizes were medium (0.37 and 0.29) and the dBFs revealed only moderate evidence. Numerically, the mean $\log(k)$'s of the softmax model showed the same tendency, but here effects were less pronounced. One possibility is that the inclusion of additional latent variables in the DDM_S might have increased sensitivity to detect this effect. There was also evidence for a session effect on the scaling parameter (ν_{coeff}). Here, the impact of trial-wise value differences on the drift rate was attenuated in the Lab-session, with dBFs revealing strong ($VR_{neutral}$) or moderate evidence ($VR_{gambling}$) for a reduction in ν_{coeff} in the Lab-session. Again, effect sizes were medium. Nevertheless, the data suggest increased sensitivity to value differences in VR. This effect might be due to the option presentation in the Lab-session compared to the VR-sessions. The presentation of options within VR might have been somewhat more salient, which might have increased attention allocated to the value differences within the VR-sessions. However, this remains speculative until further research reproduces and further assesses these specific effects on the DDM parameters. Boundary separation (α), drift rate asymptote (ν_{max}) and starting point (z) showed little evidence for systematic differences between sessions. The only DDM_S parameter showing extreme evidence for a systematic difference between the lab- and VR-sessions was the non-decision time (τ). This effect is unsurprising, as it describes RT components attributable to perception and/or motor execution. Given that indicating a response with a controller in three-dimensional space takes longer than a simple button press, this leads to substantial increases in τ during VR testing. Finally, the good test–retest reliability of $\log(k)$ from the DDM_S furthermore indicates that RTs obtained in VR can meaningfully be modeled using the DDM. The potential utility of this modeling approach in the context of gambling disorder is illustrated by a recent study that reported reduced boundary separation (α) in participants suffering from gambling disorder compared to healthy controls in a reinforcement learning task[48]. Given that there are mixed results when it comes to the effect of addiction related cues on RTs[79–81], the effects of these cues on the latent decision variables included in the DDM could provide additional insights. Taken together, these results show that VR immersion in general does not influence participants inter-temporal preferences in a systematic fashion and might open up a road to more ecologically valid lab experiments, e.g., focusing on behavioral cue-reactivity in addiction. This is in line with other results showing the superiority of VR compared to classical laboratory experiments[6].

The present data add to the discussion concerning the reliability of behavioral tasks[9,50–53,55] in particular in the context of computational psychiatry[15,82]. To examine test–retest reliability, the three sessions were performed on different days and with a mean interval of 3.85 days between sessions. The test–retest reliability for the AUC and the $\log(k)$ parameter of the hyperbolic discounting model with softmax choice rule were both excellent. For the $\log(k)$ of the DDM_S the ICC was good, but slightly lower than for AUC and softmax. Nevertheless, the discount rate $\log(k)$ was overall stable regardless of the analytical approach. The ICC of 0.7 observed for the DDM_S was comparable to earlier studies on temporal discounting reliability[52,53]. Kirby and colleagues[52] for instance demonstrated a reliability of 0.77 for a 5-week interval and 0.71 for 1 year. This shows that at least over shorter periods from days to weeks, temporal discounting performed in VR has a reliability comparable to standard lab-based testing. Enkavi and colleagues[49] stress that in particular difference scores between conditions (e.g. Stroop, Go-NoGo etc.), show unsatisfactory reliability due to the low between participants variation created by commonly used behavioral tasks. Assessment of difference scores was not applicable in the present study. Nevertheless, there was no positive evidence for systematic effects on $\log(k)$ (with the exception of the potential novelty effects discussed above), and the test–retest reliability between all conditions was at least good across analysis schemes, indicating short-term stability of temporal discounting measured in VR. It is worth noting, however, that temporal discounting shares some similarities with questionnaire-based measures. As in questionnaires, in temporal discounting tasks participants are explicitly instructed to indicate their preferences. This might be one reason why the reliability of temporal discounting is often substantially higher than that of other behavioral

tasks[49,52,53,55]. Other parameters of the DDM_S showed lower levels of test–retest reliability. Especially the ν_{coeff} parameters were less reliable, at least when estimated jointly with ν_{max}. In the DDM_L, which does not suffer from potential trade-offs between these different drift rate components, the ICC of ν_{coeff} was good (Supplementary Table S2). Similarly, here log(k) also showed an excellent ICC.

The substantially lower test–retest reliability exhibited by the parameters of the DDM_S that represent latent decisions processes, compared to log(k) or AUC warrants further discussion. Prior publications from our lab[24,41] have extensively reported parameter recovery of the DDMs model and revealed a good recovery performance. The low test–retest reliability is therefore unlikely to be due to poor identifiability of model parameters. One possible reason for this discrepancy between log(k)/AUC and the other parameters is that the tendency to discount value over time might be a stable trait-like factor, while the latent decision processes reflected in the other DDM_S parameters might be more substantially influenced by state effects. While this could explain the low test–retest reliability, it would predict that these parameters should nonetheless be stable within sessions. We addressed this issue in a further analysis of within-session split-half reliability (see Supplementary Tables S3–S5). The results showed a good-to-excellent within-session stability for most parameters, with the drift rate coefficient ν_{coeff} being a notable exception. This is compatible with the idea that latent decision processes reflected in the DDM_S parameters might be affected by factors that differ across testing days, but are largely stable within sessions, such as mood, fatigue or motivation.

VR has previously been used to study cue-reactivity in participants suffering from gambling disorder[2,3,83], but also in participants experiencing nicotine[84] and alcohol[1] use disorders. Our experimental set-up extends these previous approaches in several ways. First, we included both a neutral and a gambling-related environment. This allows us to disentangle general VR effects from specific contextual effects. Second, our reliability checks for temporal discounting show that model-based constructs with clinical relevance for addiction[18,23] can be reliably assessed when behavioral testing is implemented directly in the VR environment. Together, these advances might yield additional insights into the mechanisms underlying cue-reactivity in addiction, and contextual effects in psychiatric disorders more generally.

Understanding how addictions manifest on a computational and physiological level is important to further the understanding the mechanisms underlying maladaptive decision-making. Although alterations in neural reward circuits, in particular in ventral striatum and ventromedial prefrontal cortex, are frequently observed in gambling disorder, there is considerable heterogeneity in the directionality of these effects[85]. Gambling-related visual cues interfere with striatal valuation signals in participants suffering from gambling disorder, and might thereby increase temporal discounting[12]. In the present work, assessment of physiological reactivity to VR was limited to electrodermal activity (EDA). EDA is an index of autonomic sympathetic arousal, which is in turn related to the emotional response to addiction related cues[39,86–88]. The skin conductance level (SCL) is increased in participants with substance use disorders in response to drug related cues[86]. Additionally, it has been shown that addiction related cues in VR can elicit SCR responses in teen[87] and adult[88] participants suffering from a nicotine addiction. In our study, we mainly used this physiological marker to assess how healthy participants react to VR exposure. For the number of spontaneous responses in the EDA signal (nSCRs), the increase upon exposure to VR (B5 vs. F1) was only significant in the $VR_{neutral}$ environment. The effect size for the difference between both environments was medium. Given that the two starting areas of the VR-environments were identical, this difference might have been caused by random fluctuations. However, an increase in the number of spontaneous SCRs during VR immersion has been reported previously[5] and thus warrants further investigation. The SCL, on the other hand, increased substantially upon exposure to VR, as indicated by a significant increase between the last minute of baseline recording (B5) and the first minute of the first exploration phase (F1). The effect sizes indicated a large effect. SCL then remained elevated throughout both exploration phases (F1 to S5) but did not increase further when the virtual café/casino area was entered. These results suggest that exposure to VR increases sympathetic arousal as measured with SCL in healthy control participants independent of the presented VR environment.

There are several limitations that need to be acknowledged. First, there was considerable variability in test–retest intervals across participants. While most of the sessions were conducted within a week, in some participants this interval was up to 3 weeks, reducing the precision of conclusions regarding temporal stability of discounting in VR. Other studies, however, have used intervals ranging from 5 to 57 weeks[52] or three months[53], and have reported comparable reliabilities. Moreover, there is evidence for a heritability of temporal discounting of around 30 and 50 percent at the ages of 12 and 14 years respectively[89]. This increases the confidence in the results obtained here. Nevertheless, a more systematic assessment of how long these trait indicators remain stable in VR would be desirable and could be addressed by future research. Second, the sample size was lower compared to larger studies conducted online[49], and the majority of participants was female. Both factors limit the generalizability of our results. However, large-scale online studies have shortcomings of their own, including test batteries that take multiple hours and/or multiple sessions to complete[49,50], potentially increasing participants' fatigue, and which might have detrimental effects on data quality. We also note that the present sample size was sufficiently large to reveal stable parameter estimates, showing that in our design participants performed the task adequately. Thirdly, the immersion in VR might have been reduced by the available physical lab space. To ensure safety, the experimenter had to at times instruct participants to stay within the designated VR-zone. This distraction might have reduced the effects caused by the VR-environments, because participants were not able to fully ignore the actual physical surroundings. Additionally, it might have influenced the EDA measurements in an unpredictable way. Future research would benefit from the implementation of markers within the VR-environments in order to ensure safety without breaking immersion. Moreover, participants had to spend about thirty minutes in the full VR-setup. The behavioral tasks were presented after the exploration phase, such that participants might have been fatigued or experienced discomfort during task completion. Finally, the study at hand did not include participants that gamble frequently or are suffering from gambling disorder and

is therefore not a cue-reactivity study itself, but rather a methodological validation for future studies using this and similar designs. Due to the fact that participants here were supposed to be fairly unfamiliar with gambling environments this study could not determine how ecologically valid the gambling environment actually is. This needs to be addressed in future research. In relation to that, cue-reactivity in gambling disorder is determined by many individual factors[37]. The VR-design presented here is designed for slot machine and sports betting players, and thus not applicable for other forms of gambling.

Overall, our results demonstrate the methodological feasibility of a VR-based approach to behavioral and physiological testing in VR with potential applications to cue-reactivity in addiction. Healthy non-gambling control participants showed little systematic behavioral and physiological effects of the two VR environments. Moreover, our data show that temporal discounting is reliable behavioral marker, even if tested in very different experimental settings (e.g. standard lab testing vs. VR). It remains to be seen if such gambling-related environments produce cue-reactivity in participants suffering from gambling disorder. However, results from similar applications have been encouraging[2,3]. These results show the promise of VR applications jointly assessing of behavioral and physiological cue-reactivity in addiction science.

Received: 21 August 2020; Accepted: 15 March 2021
Published online: 29 March 2021

References

1. Ghiță, A. *et al.* Cue-elicited anxiety and alcohol craving as indicators of the validity of ALCO-VR software: a virtual reality study. *J. Clin. Med.* **8**, 1153 (2019).
2. Bouchard, S. *et al.* Using virtual reality in the treatment of gambling disorder: the development of a new tool for cognitive behavior therapy. *Front. Psychiatry* **8**, 27 (2017).
3. Giroux, I. *et al.* Gambling exposure in virtual reality and modification of urge to gamble. *Cyberpsychol. Behav. Soc. Netw.* **16**, 224–231 (2013).
4. Wang, Y. G., Liu, M. H. & Shen, Z. H. A virtual reality counterconditioning procedure to reduce methamphetamine cue-induced craving. *J. Psychiatr. Res.* **116**, 88–94 (2019).
5. Peterson, S. M., Furuichi, E. & Ferris, D. P. Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading. *PLoS ONE* **13**, 1–17 (2018).
6. Rubo, M. & Gamer, M. Stronger reactivity to social gaze in virtual reality compared to a classical laboratory environment. *Br. J. Psychol.* https://doi.org/10.1111/bjop.12453 (2020).
7. Detez, L. *et al.* A psychophysiological and behavioural study of slot machine near-misses using immersive virtual reality. *J. Gambl. Stud.* **35**, 929–944 (2019).
8. Dickinson, P., Gerling, K., Wilson, L. & Parke, A. Virtual reality as a platform for research in gambling behaviour. *Comput. Hum. Behav.* **107**, 106293 (2020).
9. Amlung, M. *et al.* Delay discounting as a transdiagnostic process in psychiatric disorders: a meta-analysis. *JAMA Psychiatr.* https://doi.org/10.1001/jamapsychiatry.2019.2102 (2019).
10. Kirby, K. N. & Petry, N. M. Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls. *Addiction* **99**, 461–471 (2004).
11. Peters, J. *et al.* Lower ventral striatal activation during reward anticipation in adolescent smokers. *Am. J. Psychiatry* **168**, 540–549 (2011).
12. Miedl, S. F., Büchel, C. & Peters, J. Cue-induced craving increases impulsivity via changes in striatal value signals in problem gamblers. *J. Neurosci.* **34**, 4750–4755 (2014).
13. Potenza, M. N. Review. The neurobiology of pathological gambling and drug addiction: an overview and new findings. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 3181–3189 (2008).
14. Wiehler, A. & Peters, J. Reward-based decision making in pathological gambling: the roles of risk and delay. *Neurosci. Res.* **90**, 3–14 (2015).
15. Huys, Q. J. M., Maia, T. V. & Frank, M. J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **19**, 404–413 (2016).
16. Farrell, S. & Lewandowsky, S. *Computational Modeling of Cognition and Behaviour* (Cambridge University Press, 2018). https://doi.org/10.1017/CBO9781316272503
17. Miedl, S. F., Peters, J. & Büchel, C. Altered neural reward representations in pathological gamblers revealed by delay and probability discounting. *Arch. Gen. Psychiatry* **69**, 177–186 (2012).
18. Bickel, W. K., Koffarnus, M. N., Moody, L. & Wilson, A. G. The behavioral- and neuro-economic process of temporal discounting: a candidate behavioral marker of addiction. *Neuropharmacology* **76**(Pt B), 518–527 (2014).
19. Lempert, K. M., Steinglass, J. E., Pinto, A., Kable, J. W. & Simpson, H. B. Can delay discounting deliver on the promise of RDoC?. *Psychol. Med.* **49**, 190–199 (2019).
20. Dixon, M. R., Jacobs, E. A., Sanders, S. & Carr, J. E. Contextual control of delay discounting by pathological gamblers. *J. Appl. Behav. Anal.* **39**, 413–422 (2006).
21. Genauck, A. *et al.* Cue-induced effects on decision-making distinguish subjects with gambling disorder from healthy controls. *Addict. Biol.* **25**, 1–10 (2020).
22. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Rev.* **18**, 247–291 (1993).
23. Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F. & Baxter, C. Remember the future: Working memory training decreases delay discounting among stimulant addicts. *Biol. Psychiatry* **69**, 260–265 (2011).
24. Bickel, W. K., Moody, L. & Quisenberry, A. Computerized working-memory training as a candidate adjunctive treatment for addiction. *Alcohol Res. Curr.* **36**, 123 (2014).
25. Carter, B. L. & Tiffany, S. T. Meta-analysis of cue–reactivity in addiction research. *Addiction* **94**, 327–340 (1999).
26. Starcke, K., Antons, S., Trotzke, P. & Brand, M. Cue-reactivity in behavioral addictions: a meta-analysis and methodological considerations. *J. Behav. Addict.* **7**, 227–238 (2018).
27. Berridge, K. C. & Robinson, T. E. Liking, wanting, and the incentive-sensitization theory of addiction. *Am. Psychol.* **71**, 670–679 (2016).
28. Anselme, P. Motivational control of sign-tracking behaviour: a theoretical framework. *Neurosci. Biobehav. Rev.* **65**, 1–20 (2016).
29. Berridge, K. C. From prediction error to incentive salience: mesolimbic computation of reward motivation. *Eur. J. Neurosci.* **35**, 1124–1143 (2012).

30. Volkow, N. D. *et al.* Cocaine cues and dopamine in dorsal striatum: mechanism of craving in cocaine addiction. *J. Neurosci.* **26**, 6583–6588 (2006).
31. van Holst, R. J., van Holstein, M., van den Brink, W., Veltman, D. J. & Goudriaan, A. E. Response inhibition during cue reactivity in problem gamblers: an fmri study. *PLoS ONE* **7**, 1–10 (2012).
32. Brevers, D., He, Q., Keller, B., Noël, X. & Bechara, A. Neural correlates of proactive and reactive motor response inhibition of gambling stimuli in frequent gamblers. *Sci. Rep.* **7**, 1–11 (2017).
33. Brevers, D., Sescousse, G., Maurage, P. & Billieux, J. Examining neural reactivity to gambling cues in the age of online betting. *Curr. Behav. Neurosci. Rep.* **6**, 59–71 (2019).
34. Crockford, D. N., Goodyear, B., Edwards, J., Quickfall, J. & El-Guebaly, N. Cue-induced brain activity in pathological gamblers. *Biol. Psychiatry* **58**, 787–795 (2005).
35. Goudriaan, A. E., De Ruiter, M. B., Van Den Brink, W., Oosterlaan, J. & Veltman, D. J. Brain activation patterns associated with cue reactivity and craving in abstinent problem gamblers, heavy smokers and healthy controls: an fMRI study. *Addict. Biol.* **15**, 491–503 (2010).
36. Kober, H. *et al.* Brain activity during cocaine craving and gambling urges: an fMRI study. *Neuropsychopharmacology* **41**, 628–637 (2016).
37. Limbrick-Oldfield, E. H. *et al.* Neural substrates of cue reactivity and craving in gambling disorder. *Transl. Psychiatry* **7**, e992 (2017).
38. Potenza, M. N. *et al.* Gambling urges in pathological gambling. *Arch. Gen. Psychiatry* **60**, 828 (2003).
39. Braithwaite, J. J., Watson, D. G., Jones, R. & Rowe, M. A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* **49**, 1017–1034 (2013).
40. Forstmann, B. U., Ratcliff, R. & Wagenmakers, E.-J. Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annu. Rev. Psychol.* **67**, 641–666 (2016).
41. Ballard, I. C. & McClure, S. M. Joint modeling of reaction times and choice improves parameter identifiability in reinforcement learning models. *J. Neurosci. Methods* **317**, 37–44 (2019).
42. Shahar, N. *et al.* Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Comput. Biol.* **15**, e1006803 (2019).
43. Fontanesi, L., Gluth, S., Spektor, M. S. & Rieskamp, J. A reinforcement learning diffusion decision model for value-based decisions. *Psychon. Bull. Rev.* **26**, 1099–1121 (2019).
44. Pedersen, M. L., Frank, M. J. & Biele, G. The drift diffusion model as the choice rule in reinforcement learning. *Psychon. Bull. Rev.* **24**, 1234–1251 (2017).
45. Peters, J. & D'Esposito, M. The drift diffusion model as the choice rule in inter-temporal and risky choice: a case study in medial orbitofrontal cortex lesion patients and controls. *PLoS Comput. Biol.* **16**, 1–26 (2020).
46. Wagner, B., Clos, M., Sommer, T. & Peters, J. Dopaminergic modulation of human inter-temporal choice: a diffusion model analysis using the D2-receptor-antagonist haloperidol. *bioRxiv* (2020).
47. Miletić, S., Boag, R. J. & Forstmann, B. U. Mutual benefits: combining reinforcement learning with sequential sampling models. *Neuropsychologia* **136**, 107261 (2020).
48. Wiehler, A. & Peters, J. Diffusion modeling reveals reinforcement learning impairments in gambling disorder that are linked to attenuated ventromedial prefrontal cortex value representations. *bioRxiv* 2020.06.03.131359 (2020). https://doi.org/10.1101/2020.06.03.131359.
49. Enkavi, A. Z. *et al.* Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 5472–5477 (2019).
50. Eisenberg, I. W. *et al.* Applying novel technologies and methods to inform the ontology of self-regulation. *Behav. Res. Ther.* **101**, 46–57 (2018).
51. Hedge, C., Powell, G. & Sumner, P. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* **50**, 1166–1186 (2018).
52. Kirby, K. N. One-year temporal stability of delay-discount rates. *Psychon. Bull. Rev.* **16**, 457–462 (2009).
53. Ohmura, Y., Takahashi, T., Kitamura, N. & Wehr, P. Three-month stability of delay and probability discounting measures. *Exp. Clin. Psychopharmacol.* **14**, 318–328 (2006).
54. Peters, J. & Büchel, C. Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. *J. Neurosci.* **29**, 15727–15734 (2009).
55. Odum, A. L. Delay discounting: trait variable?. *Behav. Process.* **87**, 1–9 (2011).
56. Petry, J. *Psychotherapie der Glücksspielsucht* (Psychologie Verlags Union, 1996).
57. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
58. Kool, W., Cushman, F. A. & Gershman, S. J. When does model-based control pay off?. *PLoS Comput. Biol.* **12**, e1005090 (2016).
59. Johnson, M. W. & Bickel, W. K. Within-subject comparison of real and hypothetical money rewards in delay discounting. *J. Exp. Anal. Behav.* **77**, 129–146 (2002).
60. Bickel, W. K., Pitcock, J. A., Yi, R. & Angtuaco, E. J. C. Congruence of BOLD response across intertemporal choice conditions: fictive and real money gains and losses. *J. Neurosci.* **29**, 8839–8846 (2009).
61. Myerson, J., Green, L. & Warusawitharana, M. Area under the curve as a measure of discounting. *J. Exp. Anal. Behav.* **76**, 235–243 (2001).
62. Green, L., Myerson, J. & Macaux, E. W. Temporal discounting when the choice is between two delayed rewards. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1121–1133 (2005).
63. Mazur, J. E. An adjusting procedure for studying delayed reinforcement. *Sci. Res.* **1987**, 55–73 (1987).
64. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
65. R Core Team. *R: A Language and Environment for Statistical Computing* (R Core Team, 2013).
66. Plummer, M. A program for analysis of Bayesian graphical models. *Work* **124**, 1–10 (2003).
67. Wabersich, D. & Vandekerckhove, J. Extending JAGS: a tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behav. Res. Methods* **46**, 15–28 (2014).
68. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**(4), 457–472 (1992).
69. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. J. R. *Stat. Soc. Ser. B Stat. Methodol.* **64**, 583–639 (2002).
70. Marsman, M. & Wagenmakers, E. J. Three insights from a Bayesian interpretation of the one-sided p value. *Educ. Psychol. Meas.* **77**, 529–539 (2017).
71. Beard, E., Dienes, Z., Muirhead, C. & West, R. Using Bayes factors for testing hypotheses about intervention effectiveness in addictions research. *Addiction* **111**, 2230–2247 (2016).
72. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
73. Bach, D. R., Friston, K. J. & Dolan, R. J. Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *Int. J. Psychophysiol.* **76**, 52–55 (2010).
74. Benedek, M. & Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80–91 (2010).

75. Kerby, D. The simple difference formula: an approach to teaching nonparametric correlation. *Compr. Psychol.* **3**, 11 (2014).
76. Rosenthal, R., Cooper, H. & Hedges, L. Parametric measures of effect size. *Handb. Res. Synth.* **621**(2), 231–244 (1994).
77. Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T. & Genzel, L. Novelty and dopaminergic modulation of memory persistence: a tale of two systems. *Trends Neurosci.* **42**, 102–114 (2019).
78. D'Amour-Horvat, V. & Leyton, M. Impulsive actions and choices in laboratory animals and humans: effects of high vs. low dopamine states produced by systemic treatments given to neurologically intact subjects. *Front. Behav. Neurosci.* **8**, 1–20 (2014).
79. Juliano, L. M. & Brandon, T. H. Reactivity to instructed smoking availability and environmental cues: evidence with urge and reaction time. *Exp. Clin. Psychopharmacol.* **6**(1), 45 (1998).
80. Sayette, M. A. *et al.* The effects of cue exposure on reaction time in male alcoholics. *J. Stud. Alcohol* **55**, 629–633 (1994).
81. Vollstädt-Klein, S. *et al.* Validating incentive salience with functional magnetic resonance imaging: association between mesolimbic cue reactivity and attentional bias in alcohol-dependent patients. *Addict. Biol.* **17**, 807–816 (2012).
82. Hedge, C., Bompas, A. & Sumner, P. Task reliability considerations in computational psychiatry. *Biol. Psychiatry Cogn. Neurosci.* **5**, P837–P839 (2020).
83. Bouchard, S., Loranger, C., Giroux, I., Jacques, C. & Robillard, G. Using virtual reality to provide a naturalistic setting for the treatment of pathological gambling. In *The Thousand Faces of Virtual Reality* (ed. Sik-Lanyi, C.) (InTech, 2014). https://doi.org/10.5772/59240.
84. Gamito, P. *et al.* Eliciting nicotine craving with virtual smoking cues. *Cyberpsychol. Behav. Soc. Netw.* **17**, 556–561 (2014).
85. Clark, L., Boileau, I. & Zack, M. Neuroimaging of reward mechanisms in Gambling disorder: an integrative review. *Mol. Psychiatry* **24**, 674–693 (2019).
86. Havermans, R. C., Mulkens, S., Nederkoorn, C. & Jansen, A. The efficacy of cue exposure with response prevention in extinguishing drug and alcohol cue reactivity. *Behav. Interv. Pract. Resid. Commun. Based Clin. Progr.* **22**(2), 121–135 (2007).
87. Bordnick, P. S., Traylor, A. C., Graap, K. M., Copp, H. L. & Brooks, J. Virtual reality cue reactivity assessment: a case study in a teen smoker. *Appl. Psychophysiol. Biofeedback* **30**, 187–193 (2005).
88. Choi, J. S. *et al.* The effect of repeated virtual nicotine cue exposure therapy on the psychophysiological responses: a preliminary study. *Psychiatry Investig.* **8**, 155–160 (2011).
89. Anokhin, A. P., Golosheykin, S., Grant, J. D. & Heath, A. C. Heritability of delay discounting in adolescence: a longitudinal twin study. *Behav. Genet.* **41**, 175–183 (2011).

Acknowledgements

We thank Mohsen Shaverdy and Diego Saldivar for the implementation of the VR environments and task programming and all members of the Peters Lab at the University of Cologne for helpful discussions. This work was supported by Deutsche Forschungsgemeinschaft (DFG, Grant PE1627/5-1 to J.P.).

Author contributions

Conceptualization: J.P., L.B.; Data curation: L.B.; Formal analysis: L.B.; Funding acquisition: J.P.; Investigation: L.B., L.S.; Methodology: J.P., L.B.; Project administration: L.B.; Writing – original draft: L.B.; Writing – review & editing: J.P., L.B., L.S.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-86388-8.

Correspondence and requests for materials should be addressed to L.R.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Study 2

Study 2 summary. The second study we implemented for the dissertation project presented here built upon the feasibility findings of study 1 and depicts the major addiction-related research focus therein. After reliability and general validity of our design had been established in the first study, we invited participants reporting frequent gambling behavior as well as non-gambling participants to our VR lab to test if and how our VR environments can provoke cue-reactivity effects in the participants suffering from problematic gambling behavior. We tested subjective, physiological, and behavioral cue-reactivity with the measures and tasks described above. Replicating earlier studies on group differences between non-gambling control participants and participants suffering from GD in temporal discounting tasks and the TST, we found increased temporal discounting and decreased MB control in the gambling group. However, these differences were not systematically modulated by the VR environment presented to the participants, suggesting the absence of strong behavioral cue-reactivity in our design. Importantly, participants in the GD group indicated a strong and selective increase of subjective craving in response to exposure to the virtual casino. Taken together these results warrant caution when applying VR in research and therapy in the context of addiction-related disorders, because VR might not generally replicate the effects of real-life gambling environments. Nevertheless, VR shows promise for research and therapeutic intervention strategies in the context of GD, given that we were able to induce the desire to gamble in participants reporting frequent gambling behavior.

Increased temporal discounting and reduced model-based control in problem gambling are not substantially modulated by exposure to virtual gambling environments.

Luca R. Bruder*, Ben Wagner, David Mathar, Jan Peters

Department of Psychology, Biological Psychology, University of Cologne, Germany

*** Corresponding author**

Contact: Luca Bruder (lbruder@uni-koeln.de), Jan Peters (jan.peters@uni-koeln.de)

Author Contributions

Conceptualization: J. P., L. B.

Data curation: L. B.

Formal analysis: L. B., B.W., D. M.

Funding acquisition: J. P.

Investigation: L. B.

Methodology: J. P., L. B.

Project administration: L. B.

Writing – original draft: L. B.

Writing – review & editing: J. P., L. B., B. W., D. M.

Additional Information

The authors declare no competing interests.

Acknowledgments

We thank Mohsen Shaverdy and Diego Saldivar for the implementation of the VR environments and task programming and all members of the Peters Lab at the University of Cologne for helpful discussions. This work was supported by Deutsche Forschungsgemeinschaft (DFG, grant PE1627/5-1 to J.P.).

Abstract

High-performance virtual reality (VR) technology has opened new possibilities for the examination of the reactivity towards addiction-related cues (cue-reactivity) in addiction. In this preregistered study (https://osf.io/4mrta), we investigated the subjective, physiological, and behavioral effects of gambling-related VR environment exposure in participants reporting frequent or pathological gambling (n=31) as well as non-gambling controls (n=29). On two separate days, participants explored two rich and navigable VR-environments (neutral: café vs. gambling-related: casino/sports-betting facility), while electrodermal activity and heart rate were continuously measured using remote sensors. Within VR, participants performed a temporal discounting task and a sequential decision-making task designed to assess model-based and model-free contributions to behavior. Replicating previous findings, we found strong evidence for increased temporal discounting and reduced model-based control in participants reporting frequent or pathological gambling. Although VR gambling environment exposure increased subjective craving, there was if anything inconclusive evidence for further behavioral or physiological effects. Instead, VR exposure substantially increased physiological arousal (electrodermal activity), across groups and conditions. VR is a promising tool for the investigation of context effects in addiction, but some caution is warranted since effects of real gambling environments might not generally replicate in VR**.** Future studies should delineate how factors such as cognitive load and ecological validity could be balanced to create a more naturalistic VR experience.

Introduction

Goal-directed decision-decision making is a key aspect of optimal behavior in a complex environment. Alterations therein in the form of prepotent, impulsive response patterns and habitual tendencies may result in adverse consequences in the long run. A prominent example for this are addiction related disorders such as substance-use-disorders [1–3] or behavioral addictions like gambling disorder (GD)[4–6]. Assessing the processes underlying decision making in general and impaired decision making specifically is difficult, because often not even the agent itself knows why a decision was made. A tool to study the processes that underly decision-making impairments is computational psychiatry[7]. The young field of computational psychiatry employs theoretically grounded mathematical models to quantify these processes and assess how they are perturbed in psychiatric disorders[8], with the aim of establishing common (transdiagnostic) computational markers [9]. This might in turn inform the development of effective interventions and/or treatment targets and might support the identification of vulnerable individuals.

Studies investigating maladaptive decision-making have identified several related but distinct processes that play a role across psychiatric disorders. One process is the discounting of reward value over time (temporal discounting), as both steep and shallow discounting are associated with different psychiatric conditions[1]. In temporal discounting tasks, participants repeatedly choose between a fixed immediate reward and larger rewards that are temporally delayed[10]. The degree of temporal discounting is then estimated from choices and/or response time (RT) distributions via computational models. Altered temporal discounting is suspected to be a transdiagnostic marker for several psychiatric disorders[1,9], with addictions and related disorders forming a prominent example[9,11].

Another cognitive process that has received considerable attention for quantifying goal-directed control of decision-making is reinforcement learning (RL)[12]. RL is thought to depend on two systems, a habitual "model-free" system that learns stimulus-response associations, and a goal-directed "model-based" system that computes action consequences via a model of the environment[13]. The degree to which participants use goal-directed or habitual RL is often assessed with the 2-step task[14,15], a two-stage decision-making task where first stage decisions probabilistically determine the presented second stage, and thereby the rewards that can be obtained. Reduced model-based RL is associated with a range of subclinical symptoms[16] and addiction related disorders including GD[17] and substance use disorder[18].

Prominent characteristics of addiction are compulsive drug seeking and insensitivity to negative consequences[19]. Incentive-sensitization theory[20–22] postulates that neural circuits mediating the incentive motivation to obtain rewards become sensitized to reward-predictive cues, giving rise to craving and drug-seeking behavior. These effects are thought to be mediated by the mesocorticolimbic dopamine system[22]. For example, exposure to addiction-related cues is correlated with the modulation of striatal value signals during temporal discounting[6], and increases striatal dopamine release in humans[23]. Across substance-use disorders and behavioral addictions, such effects are referred to as *cue-reactivity*[24–26]. Cue-reactivity manifests on a physiological and subjective level[24,25]. Additionally, exposure to addiction-related cues might increase temporal discounting[6,27,28], modulate risk-taking[29] and impair cognitive performance[30].

Cue-reactivity in GD has been examined using visual cues[4,6,29,31–38] or real-life gambling environment exposure[27,28]. Both methods arguably represent extremes on the spectrum between highly controlled laboratory environments and field studies. Field studies have high ecological validity but lack control over confounding factors and complicate the assessment of physiological variables. Conversely, lab studies yield control over confounding variables but lack ecological validity. We recently proposed a virtual reality (VR) approach that combines ecological validity with the advantages of a highly controlled lab environment[39–41]. VR allows the concurrent measurement of physiological, subjective, and behavioral cue-reactivity in an ecologically valid virtual environment. Participants are equipped with head-mounted displays and immersed in two rich and navigable VR environments: a (neutral) café or a (gambling-related) casino. Participants explore the environments and subsequently perform behavioral tasks within them.

We have shown that behavioral data obtained in VR yield reliable estimates of temporal discounting[39], and allow for a comprehensive RT-based modeling via the drift-diffusion model (DDM)[39,42,43]. Here, the decision process is modelled as a dynamic diffusion process between two boundaries, providing both a more detailed account of the underlying latent decision processes[44–48] and more stable parameter estimates[49,50]. Recent studies have successfully applied this approach to disentangle dopamine effects on temporal discounting[48], quantify reinforcement learning impairments in GD[51] and clarify effects of medial orbitofrontal cortex lesions on decision-making[46].

This pre-registered VR study had three aims. First, we extended previous VR work on cue-reactivity in addiction [40,52,53] by comprehensively modeling the effects of gambling-related environments in GD on cognitive processes during decision-making. We hypothesized that participants reporting frequent and/or pathological gambling behavior would show overall increased temporal discounting[9,11] and reduced model-based RL[17] compared to non-gambling controls. Additionally, we predicted that exposure to a VR gambling context would increase subjective craving (urge-to-gamble), further increase temporal discounting[5,10] and further reduce model-based RL[17] in GD compared to controls. We hypothesized that physiological cue-reactivity in GD would manifest as increased heart rate and skin-conductance responses during exposure to a VR gambling context.

Methods

Participants. Thirty-one participants (three female) reporting regular gambling and at least one DSM-V[19] criterion for gambling disorder aged between 20 and 41 (mean = 26.06, std = 5.43) as well as thirty non-gambling control participants (three female) in a matched age range (mean = 26.83, std = 4.65) were invited to the lab on three different testing days. Groups were additionally matched on years of education and smoking (Table 1). Participants were recruited via flyers posted at local gambling venues and via postings in local internet forums. No participant reported a history of traumatic brain injury, psychiatric or neurological disorders or severe motion sickness.

Participants provided informed written consent prior to their participation, and the study procedure was approved by the Ethics Board of the Germany Psychological Society. The procedure was in accordance with the 1964 Helsinki declaration and its later amendments.

Table 1. Summary of group characteristics. p-values printed in bold font are significant at .05.

Screening	Gambling Group	Control Group	Group Difference
Gender (male/ female)	28/3	26/3	$X^2(1) < 0.001$, p = 1
Age	26.06 (5.34)	26.83 (4.65)	U = 426, p = 0.73
Years of education	12.1 (1.68)	12.07 (1.25)	U = 454.5, p = 0.94
DSM-V criteria for GD[19]	5.32 (2.41)	0 (0)	**U = 899, p < 0.001**
Beck Depression Inventory (BDI)[54]	13.45 (10.08)	6.45 (5.53)	**U = 639, p < 0.01**
Symptom-Checklist 90 (SCL90) [55]	0.93 (0.7)	0.34 (0.29)	**U = 684, p < 0.001**
South Oaks Gambling Screen (SOGS)[56]	8.23 (4.55)	0.24 (0.83)	**U = 873.5, p < 0.001**
Kurzfragebogen zum Glücksspielverhalten (KFG)[57]	24.55 (13.61)	3.21 (11.3)	**U = 851, p < 0.001**
Alcohol Use Disorder Identification Test (AUDIT)[58]	9.84 (8.87)	4.79 (3.08)	U = 580.5, p = 0.052
Fagerström Test for Nicotine Dependence (FTND)[59]	1.32 (2.5)	0.69 (1.37)	U = 483, p = 0.52
Gambling-Related Cognitions Scale (GRCS)[60]	70.35 (26.25)	26.55 (8.14)	**U = 878, p < 0.001**
Gamblers' Beliefs Questionnaire (GBQ)[61]	30.29 (19.32)	26.21 (15.89)	U = 483, p = 0.62

VR-Setup. The VR-environments were presented using a wireless HTC VIVE head-mounted display (HMD). The setup provided a 110° field of view, a 90 Hz refresh rate, and a resolution of 1440 x 1600 Pixel per eye. Participants had an area of about 6m^2 open space to navigate the virtual environment. For the execution of the behavioral tasks and additional movement control participants held one VR-controller in their dominant hand. The VR-software was run on a PC with the following specifications: CPU: Intel Core i7-3600, Memory: 32.0 GB RAM, Windows 10, GPU: NVIDIA GeForce GTX 1080 (Ti). The VR-environments themselves were designed in Unity. Auditory stimuli were presented using on-ear headphones.

VR-Environments. The VR set-up consisted of two environments, one neutral environment ($VR_{neutral}$) and one gambling-related environment ($VR_{gambling}$). Environments contained an identical starting area and different experimental areas. Participants were placed in the middle of a small rural shopping street with a small park adjacent to it. Participants heard low street noises. The starting area was intended to familiarize participants with VR and control for possible confounding effects of entering VR. From this starting area participants could move to the experimental area by entering one of the houses on the shopping street. The entrances were in the same location for both environments. The actual experimental area differed for the two VR environments (see Figure 1). In the $VR_{neutral}$ environment the experimental area comprised a small café including customers and buffet (Figure 1b, c). Participants were surrounded by low conversation and music. In the $VR_{gambling}$ environment the experimental participants were presented a small casino containing slot machines and a sports betting room (Figure 1e, f). Participants could hear slot machine sounds and sports. The floorplan of the $VR_{gambling}$ experimental area was a mirrored version of the $VR_{neutral}$ experimental areas floorplan (Figure 1a, d). Both environments contained eight animated human avatars that performed non-repetitive movements like gambling or ordering food.

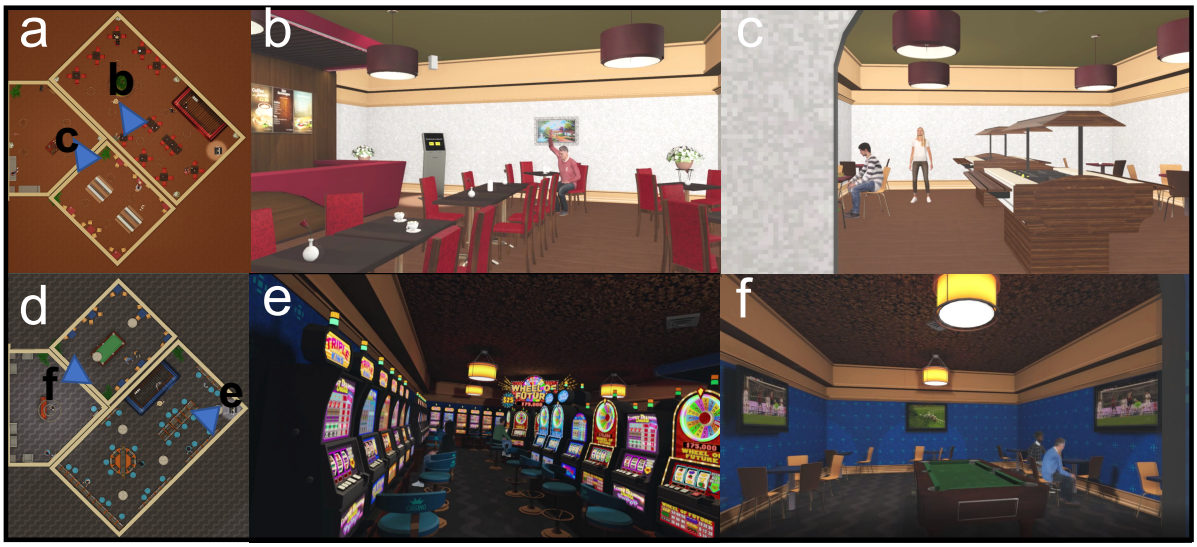


Figure 1. Experimental areas of the VR-environments a) Floorplan of the café within the VR-neutral environment b) View of the main room of the café c) View of the buffet area of the café d) Floorplan of the casino within the VR-gambling environment e) View of the main room of the casino f) View of the sports bar within the casino

Experimental procedure. Participants were invited for three testing sessions on three different days. The mean interval between sessions was 2.06 days with a range from 1 to 7 days. During the first session, participants complete a range of questionnaires and working memory tasks (see baseline screening). This session took approximately two and a half hours. In the second and third session participants entered one of the two VR environments. The order of the two VR environments was counter-balanced across participants. Upon arrival at the lab for the VR-sessions participants were first introduced to the VR equipment and handling. Subsequently they received detailed instructions for the behavioral tasks to be performed in VR. Participants were then seated, and a five-minute baseline measurement of physiological measures was obtained (baseline phase). The experimenter then helped the participants to apply the VR headset. Upon VR immersion, participants found themselves in the starting (outdoor) area of the VR environment and were instructed to explore it for five minutes (first exploration phase). Participants were then instructed to enter the interior experimental area of the VR environment and explore it for an additional five-minute period (second exploration phase). Experimental phases were each divided into five one-minute bins (B1 to B5 for the baseline phase, F1 to F5 for the first exploration phase and S1 to S5 for the second exploration phase). Following S5, participants were asked to proceed to the behavioral tasks, which were presented on a terminal within the experimental area.

Physiological measurements. Electrodermal activity (EDA)[62] was measured using a BioNomadix-PPGED wireless remote sensor together with a Biopac MP160 data acquisition

system (Biopac Systems, Santa Barbara, CA, USA). A GSR100C amplifier module with a gain of 5V, low pass filter of 10 Hz and a high pass filter DC were included in the recording system. The system was connected to the acquisition computer running under Biopac's AcqKnowledge software. Triggers for the events within the VR-environments were send to the acquisition PC via digital channels from the VR-PC. Disposable Ag/AgCl electrodes were attached to the thenar and hypothenar eminences of the non-dominant palm. Isotonic paste (Biopac Gel 101) was used to ensure optimal signal transmission. The signal was measured in micro-Siemens units (mS). The same Biopac MP160 system with a BioNomadix-PPGED wireless remote sensor was used to record the heart rate at the fingertip.

Temporal discounting task. Participants completed two behavioral tasks in each VR environment: a temporal discounting task[10] and the Two-Step task[14,15] (a sequential RL task). The temporal discounting task consisted out of 96 choices between an immediate (smaller-but-sooner, SS) reward fixed at 20 Euros, and larger but delayed (larger-later, LL) rewards. LL options were created by multiplying the SS option with a range of factors (range 1.025 to 3.85) combined with different temporal delays (range 1 to 122 days). Two sets of LL options were created by taking all possible combinations of six delays and 16 factors for each set. The two sets were matched for mean LL amount and mean delay. The order of presentation of the two sets was counterbalanced across participants. Participants were informed that one trial per session would be randomly selected and paid out in form of voucher for a widely known online store.

Options per trial were presented in two yellow squares on a black background on a display positioned within the experimental area of the VR environments (Figure 2). Offers were randomly assigned to the right or left side of the virtual display. Participants could take as long as they wanted to make a response, but they were instructed to decide intuitively. After a decision was made by aiming at the preferred option with the VR controller and pulling the trigger, a short inter-trial-interval (ISI) of .5 to 1 seconds followed. During this ISI the two yellow squares were filled with questions marks. Subsequently the next trial started.

Sequential RL task. Next, participants completed 200 trials of a modified version of the 2-step task proposed by Daw and colleagues[14]. The task consists of two stages. On each stage, participants choose between two abstract visual cues. The first stage (S1) consisted of two stimuli, one of which had to be chosen. Depending on the choice in S1, participants were then taken to one of two second stages (S2) that were characterized by different colors and different available stimuli. Here, one of the two S2 stimuli was selected, and rewarded with points

ranging from 0 to 99. One S1 cue led to one second stage with a probability of 70% ("common" transition) and to the other with a probability of 30% ("rare" transition). Transitions were reversed for the other S1 option and fixed across trials. Following selection of an S2 option, the obtained points were shown below the chosen picture for one second. S2 rewards were determined by four independent Gaussian random walks with reflecting boundaries (min: 0, max: 99). These walks were pre-computed and counterbalanced across sessions. Participants were informed about the task structure and their comprehension of the task was repeatedly questioned by the experimenter. Additionally, participants completed five example trials to get used to the task timing. In both stages, participants had three seconds to log their response. If they failed to do so they received zero points and the next trial started. Motivation to perform the task correctly was ensured by linking the final score to an additional financial reward that participants could receive. This was done by multiplying the final score of the participant with a factor of .0005 resulting in an additional reward of between five and nine €.

Baseline screening. During the first session, participants answered a range of questionnaires, and their working memory capacity was probed. A list of the questionnaires can be found in Table 1. To probe the working memory capacity of the participants we applied four established tests. First, participants completed the Rotation Span Task, which tests the ability to memorize a sequence of arrow orientations while being distracted by a letter rotation task[63]. Second, participants had to memorize sequences of letters, while doing simple calculations. This Operation Span Task was adapted from the Complex Span Task[64]. Third, we tested the listening span of participants by having them remember the last words of sentences presented in variably sized blocks. This task was adapted from the German version of Reading Span Task[65]. Fourth, participants performed a forward and a backward version of the Digit Span Task. Participants heard sequences of digits and had to remember them in the same or reversed order[66].

Temporal discounting: model-agnostic analysis. Temporal discounting data were first analyzed with a model-free approach, without a-priori model assumptions. Points of subjective equivalence between SS and LL rewards (indifference points) were estimated by fitting logistic functions to the choices for each delay. Indifference points were then plotted per participant and the area under the resulting curve (AUC) was calculated following standard procedures[67]. AUC values were compared across groups and sessions using a mixed ANOVA.

Temporal discounting: computational modeling. The effect a delay has on the subjective valuation of a reward can be accurately modelled by a hyperbolic function[68,69]. We therefore employed hierarchical Bayesian modeling to determine each participant's rate of discounting for the different sessions[8]. A hierarchical hyperbolic discounting model was fit to data of each group individually (Eq. 1):

$$SV(LL_t) = \frac{A_t}{(1 + (\exp(k + s_k * I_t)) * D_t)} \quad (1)$$

Here, $SV(LL_t)$ is the subjective (discounted) value of the LL on trial *t*. A and D represent the objective reward amount and the delay to the LL, respectively. The steepness of the hyperbolic discounting function is governed by the parameter *k* (discount rate). To avoid numerical instability of *k*-values close to, we estimated the parameter in log-space (log[k]). The parameter s_k represents the change in *log(k)* from the $VR_{neutral}$ to the $VR_{gambling}$ session. Finally, I_t is a dummy variable coding the experimental session for trial *t*.

The hyperbolic model was subsequently combined with two different choice rules, a softmax action selection rule[12] and the drift diffusion model (DDM)[45]. Softmax determines the probability of choosing the LL option on a given trial *t* (Eq. 2):

$$P(LL_t) = \frac{\exp(SV(LL_t) * (\beta + s_\beta * I_t))}{\exp(SV(SS_t) * (\beta + s_\beta * I_t)) + \exp(SV(LL_t) * (\beta + s_\beta * I_t))} \quad (2)$$

Here, the β -parameter determines choice stochasticity with respect to model-based subjective values. For $\beta = 0$ choices are random, whereas higher β values reflect stronger dependency of choices on modeled values. Again, we included a term (s_β) modeling condition-dependent changes in decision noise.

The second choice rule we applied was the DDM. In addition to the choices, the DDM includes response times (RTs) in the analysis to decompose RT distributions into latent decision processes. Recently, the DDM has increasingly been applied in the context value-based decision-making, including temporal discounting[39,46,48] and reinforcement learning[44,45] and models two-alternative forced-choice decisions as a noisy evidence accumulation process between two boundaries. As soon as the evidence in favor of one option exceeds the corresponding boundary, the process terminates, and the choice is executed. The upper boundary coded LL choices, whereas the lower boundary coded SS choices. RTs of SS choices were multiplied with -1 prior to the model estimation. We excluded, the 2.5% slowest and fastest trials of each participant from the analysis to prevent outlier trials from negatively impacting model fit[45,46].

Within the DMM, the RT on trial t is then distributed according to the Wiener first passage time (*wfpt*) (Eq. 3):

$$RT_t \sim \text{wfpt}\,(\alpha + s_\alpha * I_t, \tau + s_\tau * I_t, z + s_z * I_t, v + s_v * I_t) \quad (3)$$

Here α represents the boundary separation reflecting the speed-accuracy tradeoff. The non-decision time τ models RT components unrelated to the evidence accumulation process (e.g. perception and motor preparation). The starting point of the diffusion process is modeled via the bias z. Finally, the rate of evidence accumulation is described by the drift-rate v. The s-parameters (s_α, s_τ, s_z, s_v) again model condition effects on the corresponding parameters, and I_t is the dummy-coded condition predictor for trial t.

We compared three different versions of the DDM. First, we fit a null model (DDM_0) with a constant drift rate[46,48]. We then compared two DDMs that incorporated hyperbolic discounting, either via a linear (DDM_L) or a non-linear (DDM_S) mapping of trial-wise value differences on drift rates. For the DDM_L, v_{coeff} maps the value differences onto v:

$$v_t = (v_{coeff} + s_{v_{coeff}} * I_t) * (SV(LL_t) - SV(SS_t)) \quad (4)$$

One problem with such a linear mapping is that v might increase infinitely with high value differences, leading to a substantial under-prediction of RTs for high value differences[46]. As in previous work[46,48], we therefore also examined a non-linear model (DDM_S) (Eq. 5 and 6)[44]:

$$v_t = S\,[(v_{coeff} + s_{v_{coeff}} * I_t) * (SV(LL_t) - SV(SS_t))] \quad (5)$$

$$S(m) = \frac{2 * (v_{max} + s_{v_{max}} * I_t)}{1 + \exp(-m)} - (v_{max} + s_{v_{max}} * I_t) \quad (6)$$

This formulation caps the drift rate at v_{max}.

We and others have shown that both group- and subject-level DDM_S parameters recover well when combined with value-based decision models including temporal discounting[44,46,48].

2-step task: model-agnostic analysis. Here we modelled second stage RTs using a hierarchical generalized linear model (HGLM) with the factors transition (common or rare), session ($VR_{gambling}$ or $VR_{neutral}$) and group (gambling or non-gambling control)[70]. The 2.5% slowest and fastest trials per each participant were excluded from the analysis to reduce the negative impact of outlier trials on model fit[45,46]. We also performed a standard model-agnostic analysis of the 2-step task, modelling the probability to repeat previous trial first stage choices (pStay) as a function of transition, group and previous reward[14]. Since the present 2-step task version employed continuous rewards for S2, this analysis was conducted using a moving average of recent rewards to categorize trials into rewarded and unrewarded trials. In addition,

we assessed the final score participants achieved in both conditions with a mixed model ANOVA including group and session as factors.

2-step task: computational modeling. We again combined an RL model that learns state-action values (Q-values) for options in the different stages with both a softmax and a DDM choice rule. The softmax model was based on the hybrid model proposed by Otto and colleagues[71]. Here, both model-free (MF) and model-based (MB) Q-values are tracked for S1 options, whereas S2 options only have MF Q-values. Q_{MF}-values for both stages are updated on trial *t* via the prediction error δ (Eq. 7 to 10). The subscript j (j ϵ [1,2]) denotes the two possible actions in each stage, while *i* denotes the stage presented.

$$Q_{MF}(s_{2j,t}, a_{j,t}) = Q_{MF}(s_{2j,t-1}, a_{ij,t-1}) + (\eta_2 + s_{\eta_2} * I_t) * \delta_{2,t} \quad (7)$$

$$Q_{MF}(s_{1i,t}, a_{j,t}) = Q_{MF}(s_{1,t-1}, a_{j,t-1}) + (\eta_1 + s_{\eta_1} * I_t)\delta_{1,t-1} + (\eta_2 + s_{\eta_2} * I_t)\delta_{s2,} \quad (8)$$

$$\delta_{1,t} = Q_{MF}(s_{2,t-1}, a_{j,t-1}) - Q_{MF}(s_{1,t-1}, a_{j,t-1}) \quad (9)$$

$$\delta_{2,t} = r_{2,t-1} - Q_{MF}(s_{2i,t-1}, a_{j,t-1}) \quad (10)$$

Here, *r* corresponds to the S2 reward obtained on the previous trial. As there are only S2 rewards, S1 Q_{MF}-values are based on the Q_{MF}-values of the S2 options (equation 11) and on the S2 prediction error. η_1 and η_2 represent the S1 and S2 learning rates (i.e., the impact of prediction errors on future reward expectation). Learning rates were modeled in standard normal space [-4, 4], and back-transformed to the interval [0, 1] via the inverse cumulative normal distribution function.

In contrast, model-based Q-values (Q_{MB}, see Eq. 11) take the S_1-S_2 transition probabilities as well as S_2 Q_{MF}-values into account:

$$Q_{MB}(s_1, a_j) = P(s_{21}|s_1, a_j)\, max\, Q_{MF}(s_{21}, a) + P(s_{22}|s_1, a_j)\, max\, Q_{MF}(s_{22}, a) \quad (11)$$

Since there are no further stages to the task, for S2, $Q_{MF} = Q_{MB}$. The S2 softmax choice rule is analogous to the one described for temporal discounting. Here, β_2 governs the choice stochasticity (Eq. 12):

$$(14)P(a_{i,t} = a|s_{2,t}) = \frac{exp((\beta_2 + s_{\beta_2} * I_t)Q_{MF_{S2}}(a))}{\sum_{a'} exp((\beta_2 + s_{\beta_2} * I_t)Q_{MF_{S2}}(a'))} \quad (12)$$

The S1 softmax choice rule then includes separate weighting parameters for Q_{MF} and Q_{MB} (β_{MF} and β_{MB}) that model the impact of these Q-values on S1 choices (Eq. 13):

$$P(a_{i,t} = a|s_{1,t}) = \frac{exp\left((\beta_{MB} + I_t * s_{\beta MB})Q_{MB_{S_1}}(a) + (\beta_{MF} + I_t * s_{\beta MF})Q_{MF_{S_1}}(a) + (\rho + I_t * s_\rho) * rep(a)\right)}{\sum_{a'} exp\left((\beta_{MB} + I_t * s_{\beta MB})Q_{MB_{S_1}}(a') + (\beta_{MF} + I_t * s_{\beta MF})Q_{MF_{S_1}}(a') + (\rho + I_t * s_\rho) * rep(a')\right)}, \quad (13)$$

Action selection in S1 is thus modelled by weighing the influence of MB and MF Q-values in a softmax choice rule. Additionally, the parameter ρ models perseveration, i.e., the tendency to repeat the previous action. *Rep* takes a value of 1 if the corresponding action was taken on trial t-1, and 0 otherwise. I_t again represents the dummy-coded session.

Finally, we included a term that decayed the Q-values of unchosen stimuli in both stages[72,73] with a decay-rate η_{decay} towards the center of the reward walks (.5) (Eq. 14). For the modelling process the reward values were multiplied by 100 as these values were presented to the participants.

$$Q_{unchosen}(s_{i,t}, a_{i,t}) = Q_{unchosen}(s_{i,t-1}, a_{i,t-1}) * \left(\eta_{decay} + s_{\eta decay} * I_t\right) + \left(1 - \left(\eta_{decay} + s_{\eta decay} * I_t\right)\right) * 0.5 \quad (14)$$

We again extended this modeling approach using the same three DDM choices rules previously described for the temporal discounting task, with the key difference that now trial-wise drift-rates depended on the differences in Q-values between options ($\Delta_{Q_{MB}}$ and $\Delta_{Q_{MF}}$ respectively). For S2, this depended only on Q_{MF} value differences (Eq. 15):

$$v_{S2,t} = \left(v_{coeffS2} + s_{vcoeff_{S2}} * I_t\right) * \Delta_{Q_{MF_{S2}}} \quad (15)$$

For S1, to take both Q_{MF} and Q_{MB} values into account, we included separate drift-rate coefficients $v_{coeffMB}$ and $v_{coeffMF}$ (Eq. 16):

$$v_{S1,t} = \left(v_{coeff_{MB}} + I_t * s_{vcoeff_{MB}}\right) * \Delta_{Q_{MB}} + \left(v_{coeff_{MF}} + I_t * s_{vcoeff_{MF}}\right) * \Delta_{Q_{MF_{S1}}} + (\rho + I_t * s_\rho) \quad (16)$$

As for the temporal discounting models, the DDM_L simply used the drift rates from equations 15 and 16. In contrast, for the DDM_S these drift-rates were additionally passed through a sigmoid function S (see Eq. 17):

$$S(m) = \frac{2 * \left(v_{max_{Si}} + s_{v_{max_{Si}}} * \ I_t\right)}{1 + exp\ (-m)} - \left(v_{max_{Si}} + s_{v_{max_{Si}}} * \ I_t\right) \quad (17)$$

For all models, separate DDM parameters were estimated for the first and second stage of the task.

Hierarchical Bayesian Models. All models were fit to the data of all participants in a hierarchical Bayesian estimation scheme. Models not including session effects were fit separately per session, yielding separate estimates per participant and session. For group-level means we employed weakly informative uniform or normal priors over numerically plausible ranges (see Supplementary Table 1). For parameters modelling condition effects we used Gaussian priors with means of 0. Participant-level parameters were drawn from group-level Gaussian distributions, the means, and precisions of which were again estimated from the data. We used Markov Chain Monte Carlo to estimate posterior distributions for all model parameters. Temporal discounting modeling was done in R [74] using the JAGS software package (version 4.3.0)[75] and the Wiener module for JAGS [76]. 2-step task modeling used Python in conjunction with the *pystan* toolbox and STAN (version 2.27) [77]. We ran four chains for each model. For JAGS, chains consisted out of one million samples. Only the last 15000 were kept and the rest was discarded as a burn-in. For STAN, chains consisted of 7000 samples of which the first 3000 were discarded as a burn-in. Chain convergence was assessed using the $\hat{R}$ statistic[78] where values ≤ 1.01 were considered acceptable. To determine the best fitting DDM in both tasks we calculated the Watanabe-Akaike Information Criterion (WAIC)[79] using the loo-package (version 2.4.1) in R. Lower WAIC values represent better fits.

Session and group effects on model parameters. To quantify potential session, group and interaction effects, we compared the group-level posterior distributions across groups and conditions. We then examined the 95% highest density intervals (HDI) and calculated directional Bayes Factors (dBF) quantifying the degree of evidence for a reduction vs. an increase in a parameter. Because priors for group-level parameters were symmetric, dBFs can simply calculated as the ratio of the posterior mass of the difference distribution above zero to the posterior mass below zero[80]. Here, dBFs above 3 are interpreted as moderate evidence in favor of a positive effect, while Bayes Factors above 12 are interpreted as strong evidence for a positive effect[81]. Specifically, a dBF of 3 would imply that a positive effect is three times more likely, given the data, than a negative effect. Bayes Factors below 0.33 are likewise interpreted as moderate evidence in favor of the alternative model with reverse directionality.

A dBF above 100 is considered extreme evidence[81]. The cutoffs used here can be considered liberal, because they are usually used if the test is against a H_0 implying an effect of exactly 0.

Analysis of physiological data. Electrodermal activity (EDA) is a frequently employed index of sympathetic activity[82]. VR-related physiological arousal was first assessed using the tonic skin conductance level (SCL)[62] via continuous deconvolution analysis (CDA), using default settings in the Ledalab toolbox[83] for Matlab (The MathWorks). SCL was subsequently transformed into percentage change from the mean signal of the five-minute baseline phase and split into fifteen one-minute bins, five per experimental phase (pre-VR baseline [B], first exploration [F] and second exploration [S]). We additionally examined the number of spontaneous phasic skin conductance responses (SCR)[83]. The phasic component of the EDA signal was again divided into fifteen one-minute bins. For each bin we calculated the number of spontaneous SCRs. The resulting values were transformed into percentage change from the mean number of spontaneous SCRs during the five baseline bins.

Heart rate (HR) was analyzed in a similar fashion. The signal was converted into the percent change from baseline mean and divided into fifteen one-minute bins described above. Again, we compared bin B5 to F1 to assess the effect of immersion into VR and bins F5 and S1 to test for physiological cue-reactivity using Wilcoxon Signed-Ranks Tests[84].

To test for general VR effects, the last baseline bin (B5) was compared to the first bin of the first exploration phase (F1) via non-parametric Wilcoxon Signed-Ranks Test[84]. To test for specific cue-reactivity effects, the last bin of the first exploration phase (F5) was compared to the first bin of the second exploration phase (S1), i.e. focusing on the time point were the specific gambling vs. neutral environment was entered, using non-parametric Wilcoxon Signed-Ranks Tests[84]. Effect sizes were calculated as the Z statistic divided by the square-root of N and reported as r[85]. Absolute r values above .5 are considered large effects, while values between .3 and .5 are considered medium. Values below .3 are considered small. The significance level was Bonferroni corrected for the number of statistical tests for each signal type.

Subjective urge to gamble. Participants rated their subjective urge to gamble on a 10-point likert scale at five time points: at the beginning of the experiment (B1), at the end of the baseline phase (B5), at the end of the first exploration phase (F5), at the end of the second exploration phase (S5) and after completion of behavioral testing (end). Participants were prompted verbally and verbally reported their urge to gamble. Such verbal reports have been

466 used successfully in previous cue-reactivity research [86] and for comparable VR designs[41].
467 Ratings were analyzed using a mixed model ANOVA.
468
469

Results

Subjective urge to gamble. Participants reported their subjective desire to gamble on a 10-point likert scale at five time points (see methods section). A mixed model ANOVA revealed a main effect of group ($F(1) = 78.94$, $p < .001$, $\eta_p^2 = .434$, i.e. an overall greater urge to gamble in the GD group) and a significant three-way interaction between group, session, and time point ($F(3.059) = 2.81$, $p = .04$, $\eta_p^2 = .002$). Simple effects analysis showed that this interaction was mostly due to an increased urge to gamble upon entering the gambling area in $VR_{gambling}$ in the GD group but not the control group ($t = -11.197$, $p < .001$) (Figure 2). This increase was also significantly greater than the increase caused by the experimental area of the $VR_{neutral}$ environment ($t = -11.973$, $p < .001$). $VR_{gambling}$ exposure thus increased subjective craving specifically in the gambling group.

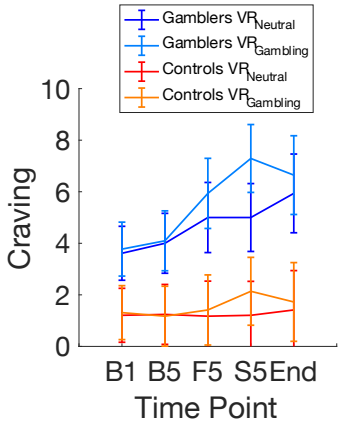


Figure 2. Subjective urge-to-gamble (craving) was verbally reported via a 10-point-likert scale at five time points, at the beginning and end of the pre-VR baseline phase (B1, B5), at the end of the first VR exploration phase (F5, VR outdoor area) and at the end of the second VR exploration phase (S5, $VR_{neutral}$ vs. $VR_{gambling}$) and upon completion of the experimental tasks (End).

Temporal discounting: model-agnostic analysis. A mixed model ANOVA on area-under-the curve values (AUC, a model-agnostic measure of temporal discounting, see methods section) revealed significantly steeper discounting (i.e., smaller AUC values) in gamblers vs. controls ($F = 16.871$, $p < .001$, $\eta_p^2 = .216$) (Figure 3 a and b). There was, however, no significant main effect of the VR session ($F = .546$, $p = .463$, $\eta_p^2 < .001$) and no significant group x session interaction ($F = 1.686$, $p = .199$, $\eta_p^2 = .002$).

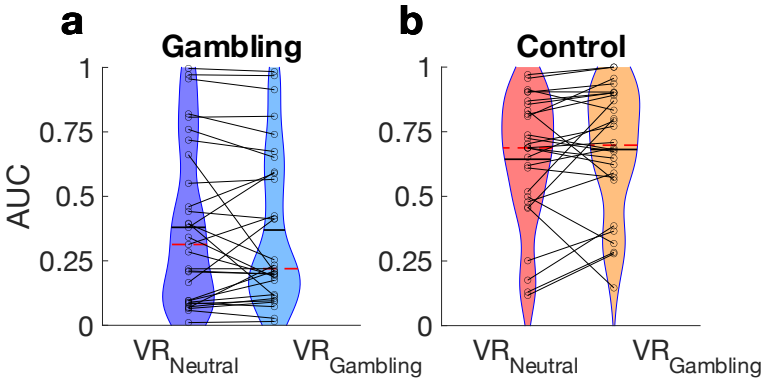


Figure 3. Violin plot of AUC values for the GD group (a) and non-gambling controls (b). Lines connect single participant data points. Dashed red lines represent the median and solid black lines the mean.

Temporal discounting softmax choice rule. Group level posteriors revealed extreme evidence for an increased discount rate log(k) in the gambling group (dBF>100, 95% HDI: min = .52, max = 2.501) (Figure 4 a and Table 2). However, there was only moderate evidence a further increase in discounting in the $VR_{gambling}$ environment gamblers (s_k dBF = 3.338; 95% HDI: min = -.156, max = .325) (Figure 4 c and Table 3). In contrast, controls showed strong evidence for reduced discounting in the $VR_{gambling}$ session (dBF = 0.072, 95% HDI: min = -.733, max = .081). Decision noise (β) was overall lower in controls vs. gamblers (dBF <0.01, 95% HDI: min = -.492, max = -.08) (Figure 4 b and Table 2). There was no conclusive evidence for a VR effect on β in either group (Controls: dBF = 1.410, 95% HDI: min = -.055, max = .065; Gambling: dBF = 2.476, 95% HDI: min = -.087, max = .17) (Figure 4 d and Table 3).

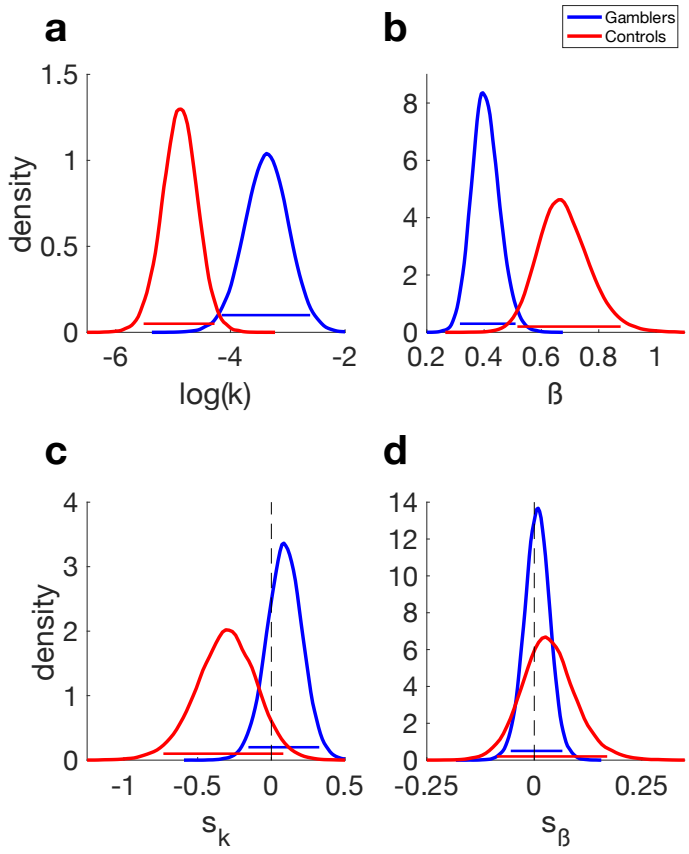


Figure 4. Hyperbolic temporal discounting model with a softmax choice rule. Group level posterior distributions for the gambling (blue) and control group (red). a) discount-rate log(k). b) softmax β. c) shift parameter s_k describing the shift in log(k) from the $VR_{neutral}$ to the $VR_{gambling}$ session. d) shift parameter s_β describing the shift in softmax β from the $VR_{neutral}$ to the $VR_{gambling}$ session. Horizontal lines denote 95% highest posterior density intervals.

Table 2. Posterior means and directional Bayes Factors (dBF) for the parameters of the hyperbolic temporal discounting model with softmax (left) and DDM_S choice rule (right). dBF values around 1 indicate that values are evenly distributed around 0. dBFs are calculated as BF = i/(1-i), with i being the probability mass of the posterior distributions above zero. dBFs for group difference are based on the difference distributions between groups. Values are reported as Gambling > Control.

	Softmax Model			**DDMs**				
Model parameter	**GD: Mean**	**CON: Mean**	**dBF group**	**GD: Mean**	**CON: Mean**	**dBF group**	**GD: dBF vs. 0**	**CON: dBF vs. 0**
log(k)	-3.371	-4.884	448.116	-3.428	-4.791	353.78	-	-
$\boldsymbol{\beta}$	.406	.679	.003	-	-	-	-	-
$\boldsymbol{v_{coeff}}$	-	-	-	.21	.267	.106	>100	>100
$\boldsymbol{v_{max}}$	-	-	-	1.772	1.86	.387	>100	>100
$\boldsymbol{\alpha}$	-	-	-	2.85	3.013	.224	-	-
$\boldsymbol{z}$	-	-	-	.499	.544	.01	-	-
$\boldsymbol{\tau}$	-	-	-	.911	1.09	.023	-	-

Table 3. Mean and Directional Bayes Factors (dBF) values for the shift parameters of the hyperbolic temporal discounting model with a softmax and a DDMs choice rule. dBFs values around 1 indicate that values are evenly distributed around 0. dBFs are calculated as BF = i/(1-i), with i being the probability mass of the posterior distributions above zero. dBFs for group difference are based on the difference distributions between groups. Values are reported as Gambling > Control. dBFs for group comparisons are based on the difference distributions of the posteriors of both groups.

	Softmax Model					DDMs				
Model parameter (shift)	**GD: Mean**	**CON: Mean**	**dBF group**	**GD: dBF vs. 0**	**Con: dBF vs. 0**	**GD: Mean**	**CON: Mean**	**dBF group**	**GD: dBF vs. 0**	**Con: dBF vs. 0**
s_k	.088	-.309	19.457	3.338	.072	-.01	-.442	21.888	1.119	.031
s_β	.006	2.476	.564	1.41	2.476	-		-	-	-
s_{vcoeff}	-	-	-	-	-	.03	.096	.137	6.643	48.758
s_{vmax}	-	-	-	-	-	.047	.104	.516	3.499	4.675
s_α	-	-	-	-	-	-.196	-.029	.263	.089	.707
s_z	-	-	-	-	-	.001	.007	.615	1.195	2.166
s_τ	-	-	-	-	-	.02	.065	.369	1.577	6.964

Table 4. Summary of the WAICs of all drift diffusion models (DDM) in all sessions. Ranks are based on the lowest WAIC in all sessions.

	Gambling		Controls		
Model	**$VR_{neutral}$**	**$VR_{gambling}$**	**$VR_{neutral}$**	**$VR_{gambling}$**	**Rank**
DDM_0	8197.4	7563.7	7558.3	6925.3	3
DDM_L	6596.5	6021.7	6296.5	5748	2
DDM_S	6243.5	5611	5619	5142.4	1

Temporal discounting drift diffusion model choice rule. Model comparison revealed that the DDM_S (DDM including a sigmoidal drift rate modulation) accounted for the data best (Table 4), replicating previous findings[39,46,48,51]. Further analyses focused on the DDM_S. The group-level posterior distributions for log(k) again indicated extreme evidence for stronger discounting in the gambling group (dBF > 100, 95% HDI: min = .43, max = 2.293) (Figure 5 a and Table 2). There was only anecdotal evidence for an increase in temporal discounting in the $VR_{gambling}$ session (modelled by the s_k parameter) in the gambling group (dBF = 1.119, 95% HDI: min = -.234, max = .249) (Figure 6 a and Table 3). In the control group however, there was strong evidence for decreased discounting in the $VR_{gambling}$ session (dBF = 0.031, 95% HDI: min = -.915, max = .004) (Figure 6 a and Table 3). For the v_{coeff} parameter, mapping trial wise value differences to the drift rate, we found moderate evidence in favor of a higher influences of trial wise value differences in the control group compared to the gambling group (dBF = .106, 95% HDI: min = -.144, max = .029) (Figure 5 d and Table 2), mirroring the effects observed for the decision noise parameter β in the softmax model. The gambling and the control groups displayed moderate and very strong evidence in favor of an increase in v_{coeff} in the $VR_{gambling}$ session (s_{vcoeff} Gambling: dBF = 6.643, 95% HDI: min = -.019, max = .087;

Control: dBF = 48.758, 95% HDI: min = .007, max = .2) (Figure 6 d and Table 3). For v_{max}, the upper boundary for the trial-wise value difference's influence on the drift rate, the dBF indicated only anecdotal evidence for a difference between groups (dBF = 0.387, 95% HDI: min = -.37, max = .193) (Figure 5 b and Table 2). Furthermore, both groups showed moderate evidence for an increase in v_{max} in the VRgambling session (sv_{max} Gambling: dBF = 3.499, 95% HDI: min = -.073, max = .167; Control: dBF = 4.675, 95% HDI: min = -.113, max = .325) (Figure 6 b and Table 3). The group level posterior distributions for the α parameter, governing the trade-off between speed and accuracy, showed only moderate evidence in favor of an increased accuracy focus in controls (dBF = 0.224, 95% HDI: min = -.524, max =.194) (Figure 5 e and Table 2). s_α parameter group level posterior distributions revealed strong evidence for a decrease in boundary separation in $VR_{gambling}$ in gamblers (dBF = 0.089, 95% HDI: min = -.478, max = .089) (Figure 6 e and Table 3). For controls, there was no conclusive evidence for directional effects (dBF = 0.707, 95% HDI: min = -.313, max = .259) (Figure 6 e and Table 3). The bias parameter z revealed extreme evidence for a more pronounced bias towards SS vs. LL choices in gamblers vs. controls (dBF = 0.009, 95% HDI: min = -.083, max = -.008) (Figure 5 c and Table 2). Session effects on z were of inconclusive directionality (s_z Gambling: dBF = 1.195, 95% HDI: min = -.024, max = .026; Control: dBF = 2.166, 95% HDI: min = -.02, max = .034) (Figure 6 c and Table 3). For the non-decision time τ there was strong evidence for a reduced value in gamblers (dBF = 0.023, 95% HDI: min = -.381, max = -.005), reflecting faster motor and/or perceptual RT components (Figure 5 f and Table 2). There was moderate evidence for an increased non-decision time in the control group in the VRgambling session ($s\tau$ dBF = 6.924, 95% HDI: min = -.094, max = .135). This was not the case for the gambling group ($s\tau$ dBF = 1.576, 95% HDI: min = -.045, max = .175) (Figure 6 f and Table 3).

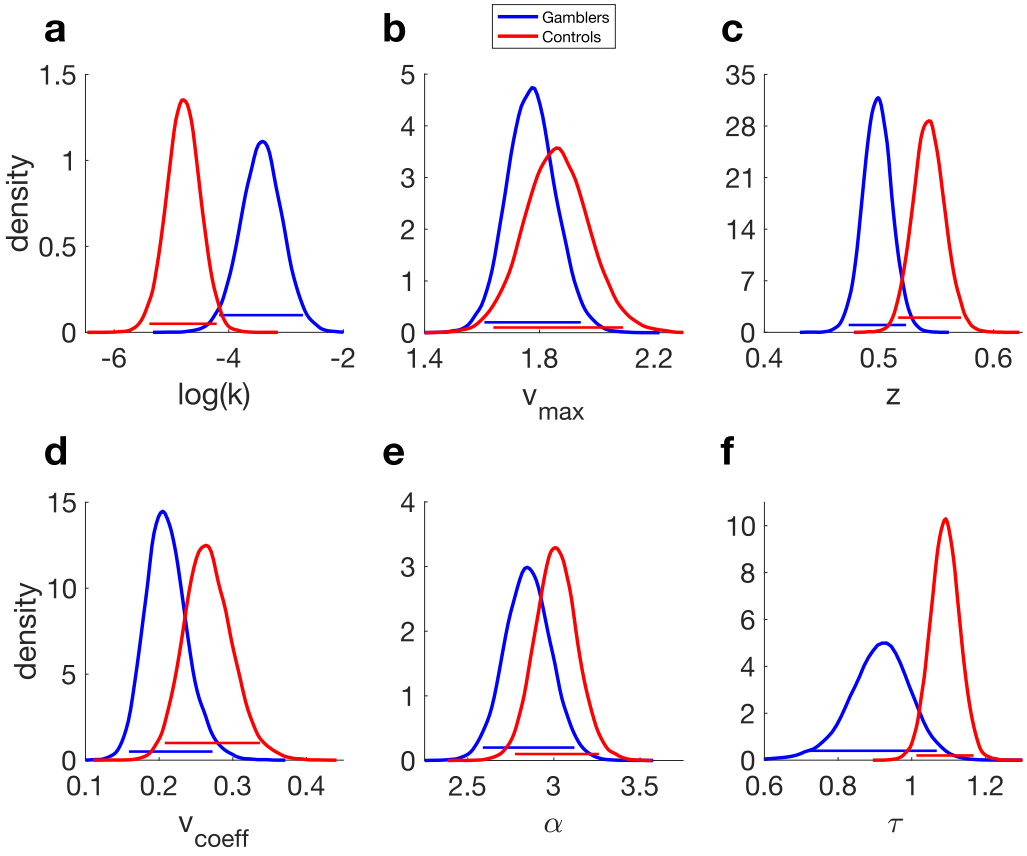


Figure 5. Hyperbolic temporal discounting model with drift diffusion model choice rule: parameter posterior distributions from the $VR_{neutral}$ condition. a) discount-rate log(k). b) maximum drift-rate v_{max}. c) starting-point z. d) drift-rate coefficient v_{coeff}. e) boundary separation α. f) non-decision time τ. Horizontal lines denote 95% highest posterior density intervals.

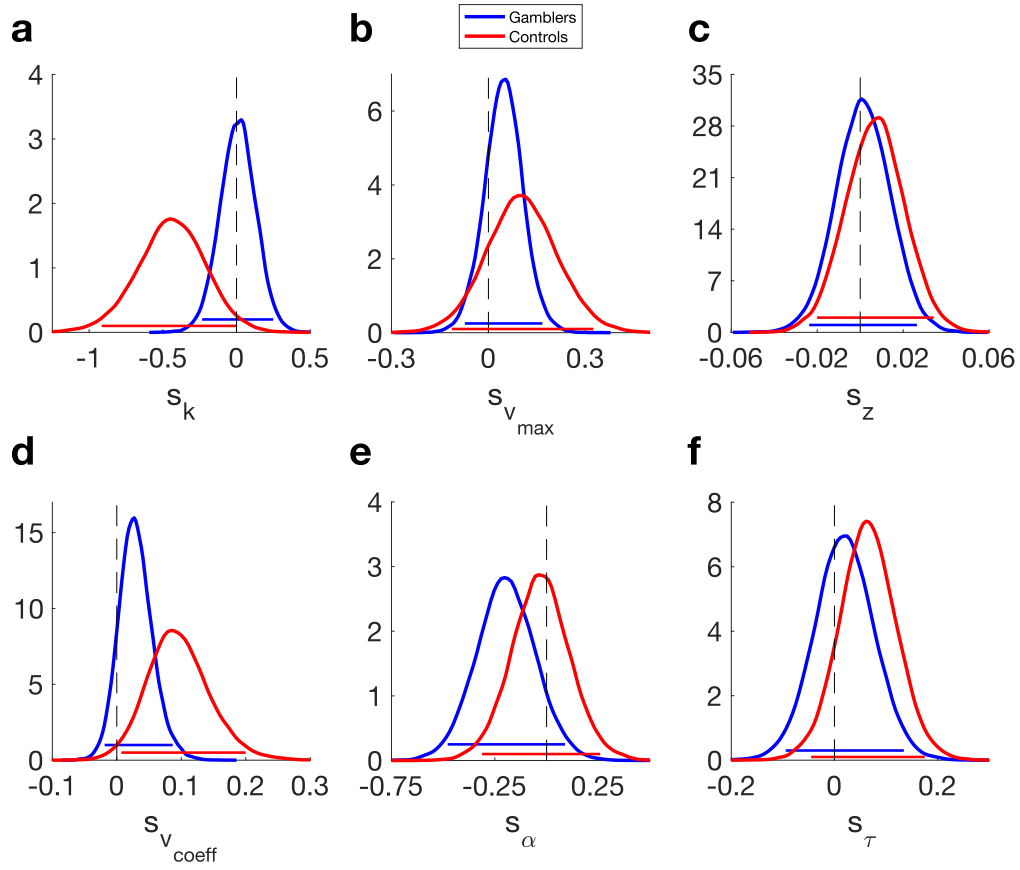


Figure 6. Hyperbolic temporal discounting model with drift diffusion model choice rule: s-parameters modeling the change in each parameter from the $VR_{neutral}$ to the $VR_{gambling}$ session. a) discount-rate log(k). b) maximum drift-rate v_{max}. c) starting-point z. d) drift-rate coefficient v_{coeff}. e) boundary separation α. f) non-decision time τ. Horizontal lines denote 95% highest posterior density intervals.

2-step task: model-agnostic analyses. Numerically, non-gambling controls scored more points overall (Mean [SD]: Gambling group, $VR_{gambling}$: 118.835 [9.931], $VR_{neutral}$: 116.751 [10.768], Control group, $VR_{gambling}$: 120.096 [11.066], $VR_{neutral}$: 121.535 [10.524]), although a mixed model ANOVA revealed that the main effect of group did not reach significance ($F(1) = 3.788$, $p = .056$, $\eta_p^2 = .021$). Likewise, there was no significant effect of session ($F(1) = .021$, $p = .887$, $\eta_p^2 = <.001$), nor a group x session interaction ($F(1) = .615$, $p = .436$, $\eta_p^2 = .007$).

Second stage RTs were then analyzed as a function of transition type (common vs. rare), session ($VR_{gambling}$ vs. $VR_{neutral}$) and group (gambling vs. non-gambling) using a hierarchical linear mixed model (see Figure 7 a to d). Here, increases in S2 RTs following rare transitions are taken as a measure of MB control[70]. This revealed significant effects of transition ($p < .001$, see Table 5), session ($p = .002$, see Table 5) and group ($p < .001$, see Table 5) as well as a significant interaction between the transition and group ($p < .001$, see Table 5), reflecting a greater increase in S2 RTs following rare transitions in controls. A further model-agnostic analysis of the 2-step task typically entails an analysis of S1 stay probabilities (i.e. the

probability of repeating the S1 choice made on the previous trial) as a function of previous reward and transition[14]. Since in the 2-step task version employed here, S2 rewards were not probabilistic but continuous we ran an analogous analysis using a moving average of recent rewards to categorize trials according to previous rewards. Results are presented in the supplementary materials (Supplementary Table 3). Resonating with the results from the RT analysis, this analysis revealed MF and MB influences on S1 decisions in both groups (the factor Reward and the interaction of the factors Reward*Transition both were significant at $p < .001$, Supplementary Table 3). Additionally, we observed a greater MB effect in the non-gambling group (significant three-way interaction of Reward*Transition*Group at $p < .001$, Supplementary Table 3).

Table 5. Results of the Hierarchical General Linear Model analysis of second stage RTs from the 2-step task, with Transition (common vs. rare), Session (gambling vs. neutral) and Group (Gambling vs. Non-gambling) as fixed effects and subject as random effect.

Fixed effects				
	Estimate	**Std. Error**	***t***	***p*-value**
Intercept	-.067	.026	-2.61	**.013**
Transition	-.183	.01	-17.761	**<.001**
Session	.037	.012	3.043	**.002**
Group	-.157	.012	-13.038	**<.001**
Trans*Session	-.022	.015	-1.494	.135
Trans*Group	1.121	.014	8.46	**<.001**
Session*Group	-.015	.017	-.888	.375
Trans*Session*Group	-.004	.02	-.179	.858

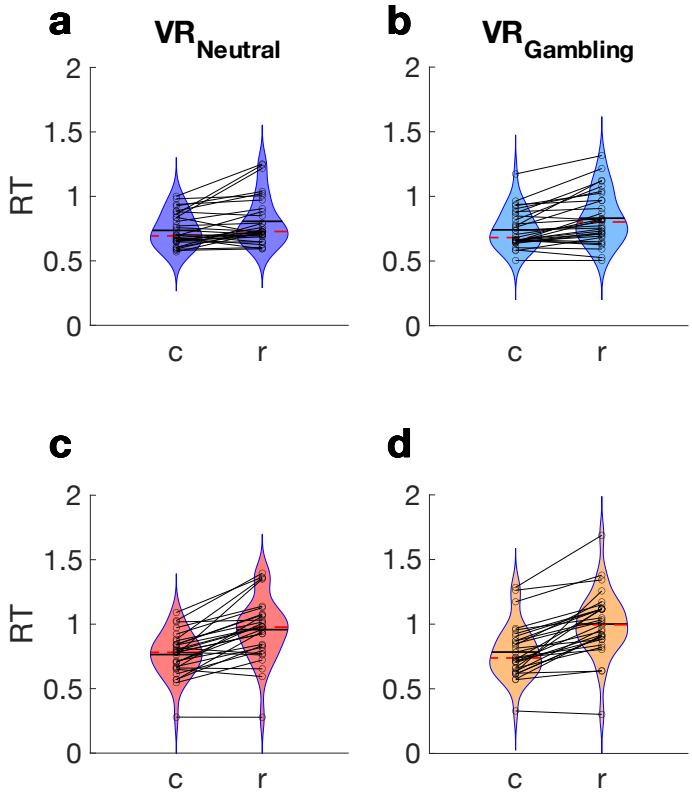


Figure 7. Violin plots of median S2 response times (RTs) from the 2-step task per group (top row: gambling group, bottom row: non-gambling group) and condition. Lines connect individual participant data points. a) gambling group, $VR_{neutral}$, b) gambling group, $VR_{gambling}$. c) control group, $VR_{neutral}$. d:) control group, $VR_{gambling}$. "c" denotes common transitions, "r" denotes rare transitions. Dashed red lines represent the median and solid black lines the mean.

2-step task: softmax choice rule. As pre-registered, we initially applied the hybrid model proposed by Daw and colleagues[14] and extended by Otto et al.[71]. The model includes independent learning rates for S1 and S2, MF and MB β weights for S1, and a single softmax slope parameter β for S2 (see methods). In both groups, S1 choices were modulated by MF Q-values (β_{MF} Gambling: dBF >100, 95% HDI: min = 1.504, max = 3.247, Controls: dBF >100, 95% HDI: min = 1.325, max = 2.682) (Figure 8 c) and MB Q-values (β_{MB} Gambling: dBF = 64.204, 95% HDI: min = .245, max = 5.379, Controls: dBF >100, 95% HDI: min = 4.505, max = 9.597) (Figure 8 b). β_{MB} was greater in controls (dBF = .011, 95% HDI: min = -7.791, max = -.64), but this was not the case for β_{MF} (BFs: 2.988, 95% HDI: min = -.757, max = 1.473). There was only moderate evidence for a gambling-session related increase in β_{MB} in both groups (Gambling: dBF: 4.692, 95% HDI: min = -.922, max = 2.811, Control: dBF: 8.442, 95% HDI: min = -.657, max = 3.455) (Figure 8 g). Other parameters modeling $VR_{gambling}$ effects only revealed if anything anecdotal evidence (see Figure 8 a to j and Tables 6 and 7). For a full list of all directional Bayes Factors please refer to Tables 6 and 7.

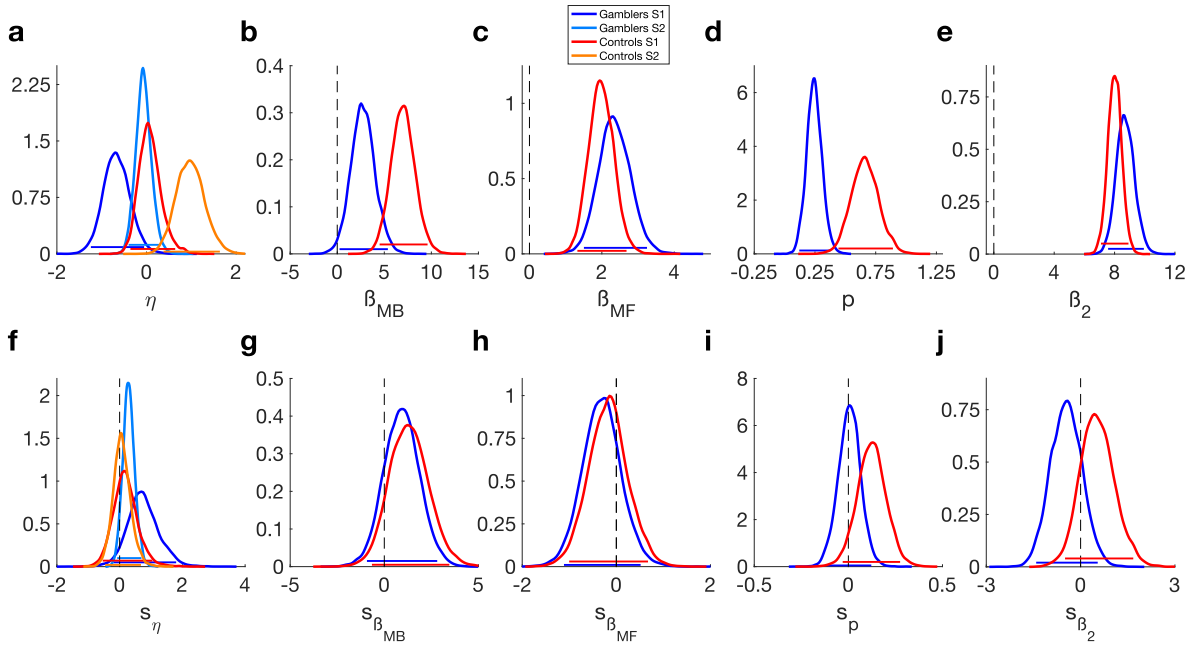


Figure 8. Hybrid model with softmax choice rule posterior distributions. Top row: Parameters values for the $VR_{neutral}$ session, Bottom row: shifts in parameter values in $VR_{gambling}$ session. a) First and second stage learning rates. b) MB weights β_{MB}. c) MF weights β_{MF}. d) perseverance parameter p. e) Second stage softmax β_2. f) shift in first and second stage learning rates. g) shift in β_{MB}. h) shift in β_{MF}. i) shift in p. j) shift in β_2. Horizontal lines denote 95% highest posterior density intervals.

Table 6. Mean and Directional Bayes Factors (dBF) values for the parameters of the Hybrid Rl model with a softmax and a DDMs choice rule. dBFs values around 1 indicate that values are evenly distributed around 0. dBFs are calculated as BF = i/(1-i), with i being the probability mass of the posterior distributions above zero. dBFs for group difference are based on the difference distributions between groups. Values are reported as Gambling > Control. dBFs for group comparisons are based on the difference distributions of the posteriors of both groups.

	Softmax Model					**DDMs**				
Model Parameter	**GD: Mean**	**Con: Mean**	**dBF group**	**GD: dBF vs. 0**	**Con: dBF vs. 0**	**GD: Mean**	**Con: Mean**	**dBF group**	**GD: dBF vs. 0**	**Con: dBF vs. 0**
η_1	.749	.188	.034	-	-	-.355	-.282	.818	-	-
η_2	.293	.075	.002	-	-	.012	.62	.057	-	-
β_{MB}	2.744	6.981	.011	64.204	>100	-	-	-	-	-
β_{MF}	2.348	1.985	2.988	>100	>100	-	-	-	-	-
β_2	8.718	8.014	4.503	>100	>100	-	-	-	-	-
p	.245	.662	<.001	-	-	.311	.708	.011	-	-
$\nu_{coeff\ MB}$	-	-	-	-	-	3.17	8.895	.009	69.665	>100
$\nu_{coeff\ MF}$	-	-	-	-	-	2.959	2.271	3.802	>100	>100
ν_{coeff2}	-	-	-	-	-	6.87	6.482	1.869	>100	>100
ν_{max1}	-	-	-	-	-	.567	1.26	.185	30.118	>100
ν_{max2}	-	-	-	-	-	2.491	2.707	.252	>100	>100
α_1	-	-	-	-	-	1.358	1.427	.22	-	-
α_2	-	-	-	-	-	1.527	1.587	.162	-	-
τ_1	-	-	-	-	-	1.738	1.896	.002	-	-
τ_2	-	-	-	-	-	.391	.421	.196	-	-

Table 7. Posterior means and Directional Bayes Factors (dBF) for parameters modeling condition effects for 2-step task models (left: hybrid RL model softmax choice rule, right: hybrid RL model with DDM_S choice rule). dBF values around 1 indicate that values are evenly distributed around 0. dBFs are calculated as BF = i/(1-i), with i being the probability mass of the posterior distributions above zero. dBFs for group difference are based on the difference distributions between groups. Values are reported as Gambling > Control.

	Softmax Model					DDMs				
Model Parameter (shift)	**GD: Mean**	**Con: Mean**	**dBF groups**	**GD: dBF vs. 0**	**Con: dBF vs. 0**	**GD: Mean**	**Con: Mean**	**dBF groups**	**GD: dBF vs. 0**	**Con: dBF vs. 0**
$s\eta_1$	.749	.188	.034	14.656	2.324	.353	.1	1.823	3.388	1.872
$s\eta_2$	.293	.075	.002	14.949	1.563	.093	.251	.531	1.385	3.345
$s_{\beta_{MB}}$	.924	1.312	.617	4.692	8.442	-	-	-	-	-
$s_{\beta_{MF}}$	-.304	-.18	.629	.304	.513	-	-	-	-	-
s_{β_2}	-.453	.534	.109	.22	5	-	-	-	-	-
Sp	.006	.123	.121	1.293	18.58	-1.236	-.364	.472	-	-
$s_{\nu\text{coeff MB}}$	-	-	-	-	-	1.963	-.443	2.192	3.582	.772
$s_{\nu\text{coeff MF}}$	-	-	-	-	-	-1.236	-.364	.472	.282	.58
$s_{\nu\text{coeff2}}$	-	-	-	-	-	.936	.83	1.027	4.322	4.864
$s_{\nu\text{max1}}$	-	-	-	-	-	-.942	-1.035	1.234	<.001	.01
$s_{\nu\text{max2}}$	-	-	-	-	-	-.006	-.266	3.629	.897	.17
$s_{\alpha 1}$	-	-	-	-	-	.015	.069	.32	1.447	8.881
$s_{\alpha 2}$	-	-	-	-	-	.015	.08	.118	1.991	57.06
$s_{\tau 1}$	-	-	-	-	-	.006	-.004	2.19	1.921	.623
$s_{\tau 2}$	-	-	-	-	-	-.002	-.001	.887	.718	.907

2-step task: drift diffusion model choice rule. Next, we replaced the softmax choice rule with the DDM. In line with the results from the temporal discounting task, a non-linear drift-rate scaling accounted for the data best, and this was the case in both groups (Table 8). In both groups, S1 choices were affected by both MF Q-value differences ($\nu_{\text{coeffMF}} > 0$, Gambling: dBF >100, 95% HDI: min = 1.828, max = 4.3; Control: dBF >100, 95% HDI: min = 1.34, max = 3.408) (Figure 9 f and Table 6) and MB Q-value differences ($\nu_{\text{coeffMB}} > 0$, Gambling: dBF = 69.665, 95% HDI: min = .414, max = 6.363; Control: dBF >100, 95% HDI: min = 5.638, max = 12.665) (Figure 9 e and Table 6). As in the softmax model, we observed extreme evidence for a greater MB effect (ν_{coeffMB}) in the non-gambling control group (dBF:.009, 95% HDI: min = -10.483, max = -1.159) (Figure 9 e and Table 6). For MF Q-values we observed moderate evidence for a higher ν_{coeffMF} ion the gambling group compared to the non-gambling control group (BFs: 3.802, 95% HDI: min = -.915, max = 2.346) (Figure 9 f and Table 6). In both groups, we there was only anecdotal or inconclusive evidence for parameter changes in the VR_{gambling} session (Figure 10 a to h and Table 7).

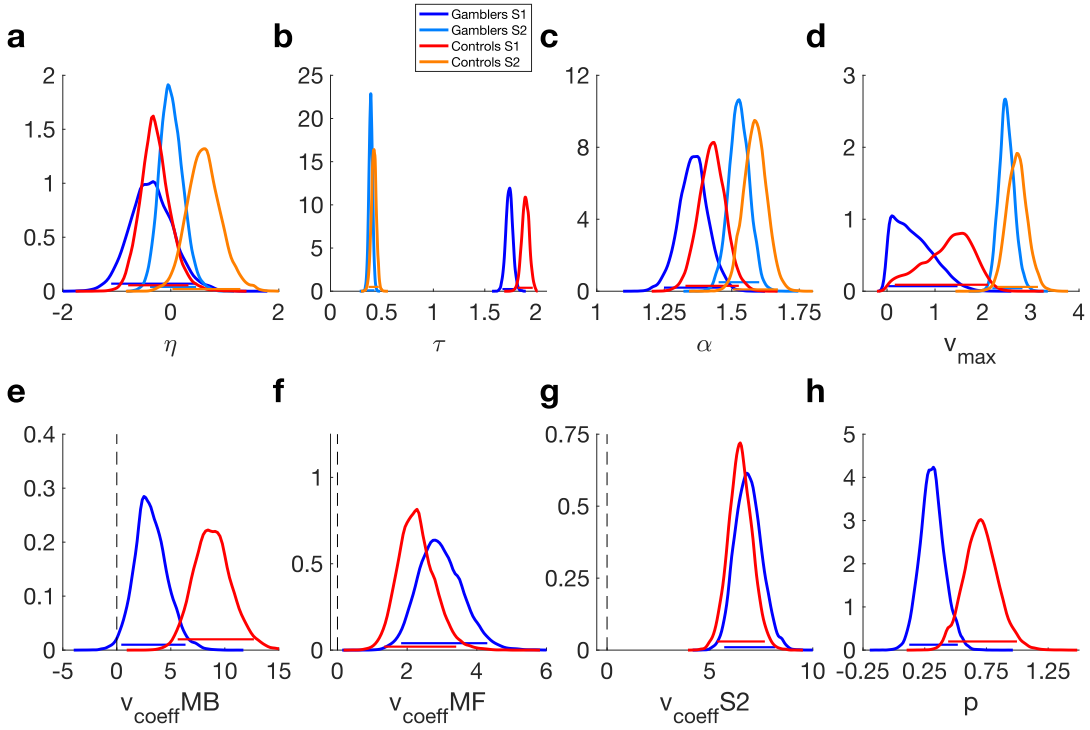


Figure 9. Hybrid RL model with a non-linear drift diffusion model choice rule. a) learning rates for S1 and S2 η. b) non-decision time τ. c) boundary separation α. d) maximum drift-rate v_{max}. e) MB drift-rate coefficient $v_{coeffMB}$. f) MF drift-rate coefficient $v_{coeffMF}$. g) drift-rate coefficient v_{coeff} for S2. h) perseverance parameter p. Horizontal lines denote 95% highest posterior density intervals.

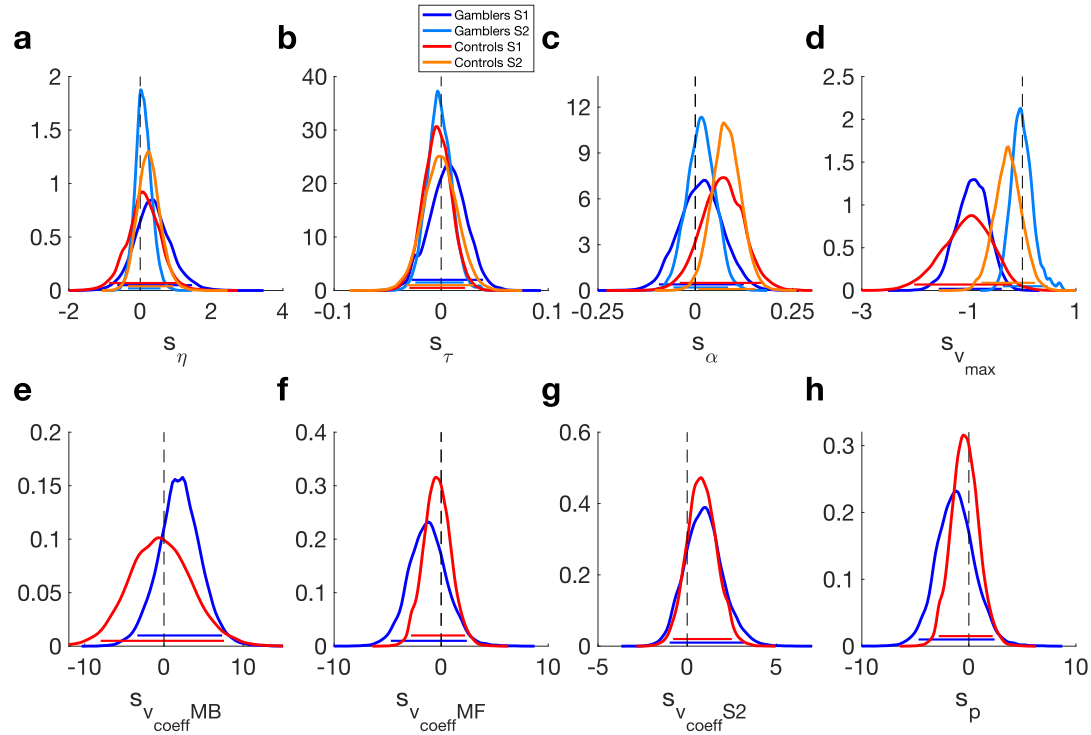


Figure 10. Posterior distributions of parameters modeling condition effects for the hybrid RL model with drift diffusion model choice rule (DDM_S). a) shift in the learning rate for S1 and S2 s_η. b) shift in non-decision time s_τ. c) shift in boundary separation s_α. d) shift in maximum drift-rate s_{vmax21}. e) shift in MB drift-rate coefficient $s_{vcoeffMB}$. f) shift in MF drift-rate coefficient $s_{vcoeffMF}$. g) shift in drift-rate

coefficient s_{vcoeff} for S2. h) shift in perseverance parameter s_p. Horizontal lines denote 95% highest posterior density intervals.

Table 8. Summary of the WAICs of all DDM models in all sessions. Ranks are based on the lowest WAIC in all sessions.

	Gambling		Control		
Model	**$VR_{neutral}$**	**$VR_{gambling}$**	**$VR_{neutral}$**	**$VR_{gambling}$**	**Rank**
DDM_0	11840	13186.2	11774	13122.7	3
DDM_L	7783.5	9244.9	6733.5	7871.9	2
DDM_S	7579.3	9122.2	6535.6	7593.5	1

Electrodermal activity (EDA). As preregistered, psychophysiological cue-reactivity was analyzed by converting the number of spontaneous skin conductance responses (nSCR) and the skin conductance level (SCL) into percentage change from baseline. Per phase, values were binned into fifteen one-minute intervals (five each for the baseline phase [B], first exploration phase [F] and second exploration phase [S]). All comparisons were tested for significance with the Wilcoxon Signed Rank Test. The significance level was Bonferroni corrected. Entering VR (i.e., B5 vs. F1) led to a significant increase in SCL in both groups and there was no significant difference between conditions (Figures 11 and 12 c and d and Supplementary Tables 4 and 5). The effect size was large throughout ($r > .5$). There was no corresponding significant change in nSCRs in either group (Figures 11 and 12 a and b Supplementary Tables 6 and 7). Psychophysiological cue-reactivity was examined by comparing the difference between F5 and S1 (i.e. the effect of entering the experimental areas: virtual café vs. virtual casino) between $VR_{neutral}$ and $VR_{gambling}$ per group. There were no significant effects on either nSCR or SCL and therefore no evidence for psychophysiological cue-reactivity in nSCRs (Figure 11 and 12 a and b and Supplementary Tables 6 and 7) and SCL (Figures 11 and 12 c and d and Supplementary Tables 4 and 5).

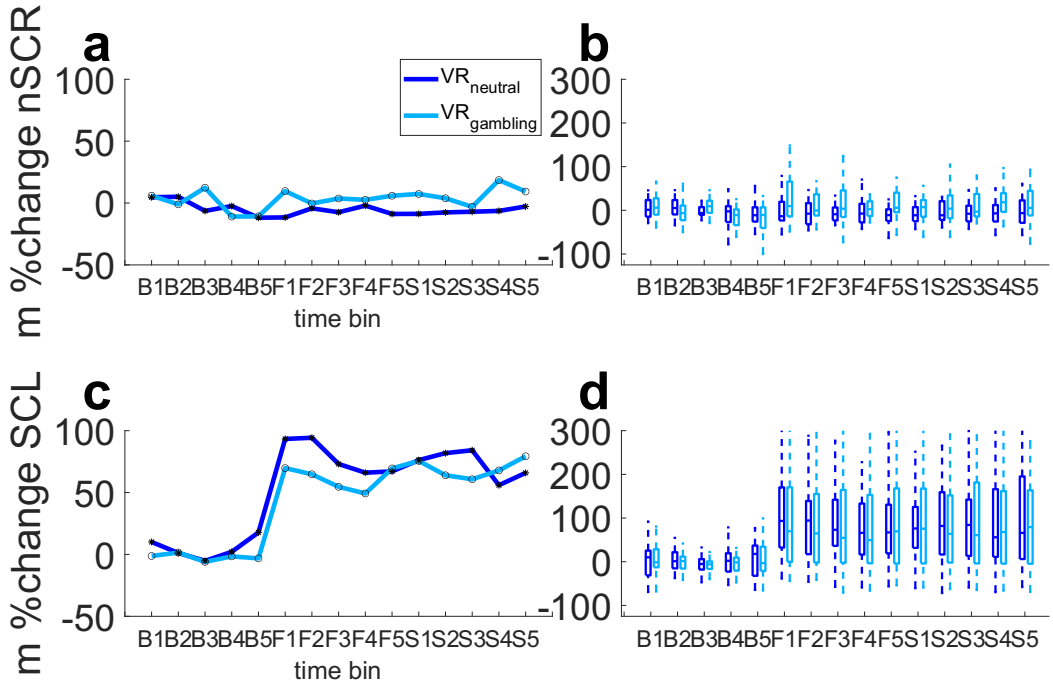


Figure 11. Results of the EDA measurements in the gambling group divided into 15 time points over the course of the baseline phase, measured before participants entered the VR-environments, and the first and second exploration phases. Each of the three phases is divided into five one-minute bins (B1-5: pre-VR baseline, F1-5: first exploration phase in VR, S1-5: second exploration phase VR). a: Median percent change from baseline mean for no. of spontaneous SCRs over gambling participants. b: Boxplot of percentage change from baseline mean for no. spontaneous SCRs over gambling participants. c: Median percent change from baseline mean of SCL over gambling participants. d: Boxplots of percentage change from base line mean of SCL over gambling participants.

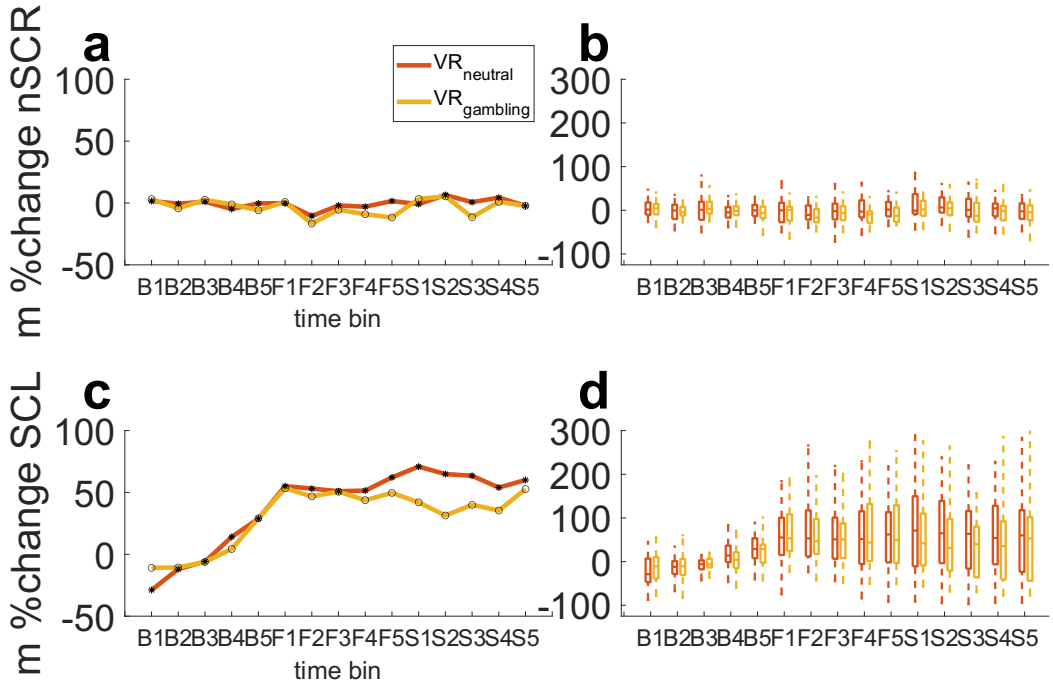


Figure 12. Results of the EDA measurements in the control group divided into 15 time points over the course of the baseline phase, measured before participants entered the VR-environments, and the first and second exploration phases. Each of the three phases is divided into five one-minute bins (B1-5: pre-VR baseline, F1-5: first exploration phase in VR, S1-5: second exploration phase VR). a: Median percent change from baseline mean for no. of spontaneous SCRs over control participants. b: Boxplot of percentage change from baseline mean for no. spontaneous SCRs over control participants. c: Median percent change from baseline mean of SCL over control participants. d: Boxplots of percentage change from base line mean of SCL over control participants.

Heart Rate (HR). Analysis of HR proceeded along similar lines as the analysis of the EDA data described above. HR was first converted into percent signal change from baseline, and then divided into fifteen one-minute bins (five each for the baseline phase [B], first exploration phase [F] and second exploration phase [S]). We observed no overall significant increase in HR in response to VR immersion (B5 vs. F1) in either group and environment, with the $VR_{gambling}$ environment in the gambling group forming the single exception (Supplementary Tables 8 and 9). However, this effect was not significantly greater than in $VR_{neutral}$. Likewise, entering the experimental areas of the VR environments did not significantly increase HR (Supplementary Tables 8 and 9) in either group or environment. Accordingly, there was no evidence for significant cue-reactivity effects on HR.

Discussion

Here we investigated the subjective, behavioral, and physiological effects of virtual reality (VR) gambling environment exposure in regular gamblers (GD group) and matched non-gambling controls. The joint assessment of these three levels of cue-reactivity in VR enabled us to thoroughly delineate several possible effects, thereby informing potential future application of VR in addiction research. Participants explored two rich and navigable virtual environments (a café environment and a casino/sports betting environment: $VR_{neutral}$ vs. $VR_{gambling}$) and within both environments performed two behavioral tasks with high relevance for gambling disorder and addiction, temporal discounting, and a 2-step sequential RL task. In both groups, exposure to VR substantially increased sympathetic arousal as reflected in the tonic skin conductance level (SCL). However, despite the fact that the $VR_{gambling}$ environment selectively increased subjective craving in the GD group, no physiological measure showed a pattern consistent with physiological cue-reactivity in the gambling group. Analysis of the behavioral data revealed that previously observed group differences between gamblers and controls were replicated in VR. First, gamblers discounted delayed rewards more steeply than controls[1,5,9,10], but this effect was not differentially modulated by the $VR_{gambling}$ environment. Second, gamblers relied less on a model-based (MB) decision strategy during the 2-step sequential RL task[17], but this effect was again not differentially modulated by the $VR_{gambling}$ environment.

Self-reported cue-reactivity. Participants verbally reported subjective urge-to-gamble (craving) at four time points during exposure to the virtual environments. Craving was overall higher in the gambling group, and exposure to the $VR_{gambling}$ environment selectively increased the subjective craving in gamblers but not controls. This is in line with earlier results[40] showing that virtual gambling environments induce subjective craving in frequent gamblers on a level comparable to gambling on real video slot machines. Thus, our VR environment exhibited ecological validity with respect to self-reported urge-to-gamble in gambling participants.

Physiological cue-reactivity. How addiction manifests on a computational, physiological and neural level has important implications for treatment and relapse prevention. Alterations in neural reward processing in ventral striatum and ventromedial prefrontal cortex have been frequently observed in gambling disorder, albeit with considerable heterogeneity in the directionality of these effects[87]. Gambling-related visual cues might interfere with striatal valuation signals in GD, and might in turn increase temporal discounting[6]. Here, analysis of

778 physiological cue-reactivity was limited to heart rate (HR) and electrodermal activity (EDA).
779 EDA indexes sympathetic arousal, which is tightly linked to processing of addiction-related
780 cues[62,88–90]. For example, addiction related cues presented in VR increased SCR amplitudes in
781 participants suffering from nicotine addiction [89,90]. In contrast, in the present study there was
782 no significant effect of entering the specific gambling-related or neutral sections of the VR
783 environments (i.e. comparing F5 vs. S1) in either group for any physiological measure. Instead,
784 in both groups, the tonic skin conductance level (SCL) increased substantially upon VR
785 immersion in both groups (effect size *r* between .66 and .81 in both groups and both sessions)
786 and remained elevated until the end of the experiment, an effect we have observed previously
787 in non-gambling controls[39]. No corresponding effects were observed for heart rate or the
788 number of spontaneous skin conductance responses. These results suggest a general increase of
789 sympathetic activity during VR immersion, but do not show evidence for physiological cue-
790 reactivity in GD in the present VR setting.
791
792 *Behavioral performance.* The behavioral tasks replicated earlier results in gamblers and
793 controls[5,17]. Regular slot machine gamblers discounted rewards substantially more steeply
794 than controls, and this was the case across model-agnostic (AUC) and model-based measures
795 (i.e. *log(k)* in softmax and diffusion models), resonating with a range of earlier results in
796 behavioral addictions and substance-use disorders [1,5,9,10]. We likewise replicated the recent
797 finding of reduced MB decision-making in participants suffering from GD[17]. The model-
798 agnostic analysis of RTs revealed that both groups showed longer S2 RTs after rare transitions,
799 reflecting MB decision-making[14,17]. However, this effect was significantly increased in non-
800 gambling controls. These results were also supported by comprehensive computational
801 modelling. A hybrid reinforcement learning (RL) model showed very strong evidence for
802 reduced MB decision making in the gambling group, and this was the case for both softmax
803 and DDM choice rules.
804
805 *Behavioral cue-reactivity.* In contrast to these robust group differences, and contrasting
806 with our preregistered hypothesis, neither temporal discounting nor model-based RL were
807 substantially modulated by the virtual gambling environment in the GD group. A seminal study
808 by Dixon and colleagues[27] showed increased discounting in gamblers when tested in a
809 gambling environment, an effect that we recently replicated[28]. Other studies showed increased
810 temporal discounting[6] or increased risk-taking[29] in participants suffering from GD in the
811 presence of gambling-related visual cues[6,29,91]. The group level posterior distribution of the

shift parameter s_k of the hyperbolic discounting model indicated only inconclusive evidence for increased discounting in the gambling group, and this was the case for both softmax and DDM choice rules. There was, however, strong evidence for *decreased* discounting in the $VR_{gambling}$ session in the non-gambling control group. There was also strong evidence that this decrease was significantly more pronounced than in gambling group (comparison of the group level posteriors of the s_k parameter). The fact that non-gambling control participants tended to discount rewards less steeply in the $VR_{gambling}$ session is of potential interest, but it is important to note that in a previous study employing the same VR design[39] we observed the reverse effect in a group of non-gambling controls. In that study, controls if anything showed reduced temporal discounting in the $VR_{neutral}$ session. Effects of virtual gambling environment exposure in non-gambling controls are therefore overall inconclusive.

An analysis of the DDM parameters revealed strong evidence for a decrease in boundary separation in the $VR_{gambling}$ session in the GD group. Participants reporting regular slot machine gambling appeared to increasingly trade-off accuracy in favour of speed in the $VR_{gambling}$ session. However, there was only moderate evidence for this decrease being stronger in the gambling group than in the non-gambling control group (comparison of the group level posteriors of the s_α parameter). It is thus difficult to draw strong conclusion from this observation. Nevertheless, a tendency towards a decrease in boundary separation would indicate that regular slot machine gamblers might attend less to actual value differences between the SS and the LL, instead preferring a more rapid response rate. In addition, the group level posterior distribution of $s_{v_{coeff}}$ in controls suggests that the value difference between the LL and the SS had a stronger influence in the trial-wise RT in the $VR_{gambling}$ session, indicating that non-gambling control participants placed a stronger weight onto the value differences between SS and LL in the $VR_{gambling}$ session. Again, there was only moderate evidence for this effect being stronger in the non-gambling control group than in the gambling group (comparison of the group level posteriors of the $s_{v_{coeff}}$ parameter).

Taken together, earlier findings of steeper discounting in gamblers when tested in real gambling venues [27] which we recently replicated[28], did not translate to VR. Whether this is due to the specific design of the present VR environments, or due to more general effects of VR on cognitive load and/or arousal (see below) is an open question.

We hypothesized that exposure the $VR_{gambling}$ environment would reduce MB and increase of MF decision-making in regular slot machine gamblers. This hypothesis was based on the idea that addiction-related cues might trigger pathological habits[21,92], which in turn could be related to reduced MB decision-making[16,93]. An alternative view is that addiction

might instead be associated with excessive goal-directed behavior, in particular in the presence of addiction related cues[94]. We observed the latter effect in a recent study in regular slot-machine gamblers[28]. Regular gamblers showed a substantial increase in model-based control when tested in a gambling venue. However, in the present VR setting, we observed little evidence for either of these effects. Gamblers showed no evidence for longer RTs after rare transitions in the $VR_{gambling}$ session, an index for MB decision-making. The hybrid model parameter s_β showed only moderate evidence for an increase in MB decision-making in the $VR_{gambling}$ environment in both groups. 95% HDIs were overlapping with zero and the change was of similar magnitude in both groups, suggesting the absence of specific cue-reactivity effects on MB decision-making in the gambling group. A similar picture emerged for the corresponding parameters from the hybrid model with DDM choice rule. Taken together, there was no conclusive evidence for VR effects on either MB or MF decision-making in either group. As reductions in MB control constitute potential transdiagnostic markers for compulsivity-related disorders[16,93] it is interesting that the $VR_{gambling}$ environment did not cause effects in the gambling group. Possible explanations for the lack of cue-reactivity effects are discussed below.

Conclusion. In contrast to the absence of behavioral and physiological cue-reactivity effects, the present VR set-up increased the subjective urge to gamble in participants reporting frequent slot machine gambling. This does not render VR generally unsuitable, as it has been shown here and by other groups that gambling related VR environments can induce craving[40]. Rather, behavioral, and physiological cue-reactivity effects might depend on specific VR design features. It is of course possible that different VR designs might have yielded the predicted effects. In addition to the specifics of the VR environments, more general effects of VR immersion might have precluded us from detecting physiological and/or behavioral cue-reactivity effects. In particular, SCL exhibited a substantial overall effect of VR immersion across groups and conditions, replicating our previous observation[39]. This might reflect increased cognitive load of VR immersion [95,96], which could interfere with the expression of behavioral effects of gambling-related environments. Likewise, physiological correlates of VR-related cognitive load could have precluded us from detecting more subtle modulation of SCL due to cue-reactivity effects. The lack of behavioral and/or physiological correlates for the reported increase in subjective craving may warrant caution for future applications of VR in exposure therapy and addiction science. Exposure therapy aims to confront patients with key stimuli and train strategies to overcome craving or fear responses. It is therefore important to

note that VR might be limited in ecological validity to reproduce real-life behavioral effects[27,28].

Limitations. There are several limitations that need to be acknowledged here. First, although groups were matched on key variables, group differences on depressive symptoms and overall psychopathology remained, as in previous studies [6,17]. Second, participants spend between thirty and forty minutes in VR. Behavioral tasks were performed after the initial exploration phase. Participants might therefore have been fatigued or experienced some discomfort during task performance. Future studies might thus benefit from shorter designs. Third, the immersion in VR was constrained by the available physical lab space. The experimenter had to ensure the safety of participant by giving external instructions when needed. Distractions caused by these instructions might have reduced the immersion experienced by the participants. Additionally, such instructions might have affected physiological measures. Future research would benefit from implementing clear markers within the VR environments to ensure safety without breaking immersion. Finally, the virtual slot machines used in our design did not exactly match most recent machines used in local gambling facilities. This might have reduced the level of realism that the VR environment conveyed. However, since an increase in the subjective urge to gamble was observed in the GD group, this indicates sufficient ecological validity to produce subjective craving. Future research could benefit from improved quality of graphical assets, e.g., by creating objects that more closely resemble current video slot machines.

Overall, we reproduced established group differences in decision-making between participants suffering from GD and non-gambling control participants in a VR setting: Participants reporting frequent gambling showed higher levels of temporal discounting and reduced MB decision-making, compared to non-gambling controls. However, we found little evidence for behavioral or physiological effects of virtual gambling environments in the GD group, despite these environments eliciting increased subjective craving in gamblers. Some caution is therefore warranted when applying VR in experimental or therapeutical contexts, as established behavioral effects of gambling environments[27,28] might not generally replicate in VR. Future studies should delineate how cognitive load and ecological validity could be balanced in VR to create a more naturalistic VR experience.

References

1. Amlung, M. *et al.* Delay Discounting as a Transdiagnostic Process in Psychiatric Disorders: A Meta-analysis. *JAMA psychiatry* (2019) doi:10.1001/jamapsychiatry.2019.2102.
2. Kirby, K. N. & Petry, N. M. Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls. *Addiction* **99**, 461–471 (2004).
3. Peters, J. *et al.* Lower ventral striatal activation during reward anticipation in adolescent smokers. *Am. J. Psychiatry* **168**, 540–549 (2011).
4. Potenza, M. N. Review. The neurobiology of pathological gambling and drug addiction: An overview and new findings. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3181–3189 (2008).
5. Wiehler, A. & Peters, J. Reward-based decision making in pathological gambling: The roles of risk and delay. *Neurosci. Res.* **90**, 3–14 (2015).
6. Miedl, S. F., Büchel, C. & Peters, J. Cue-induced craving increases impulsivity via changes in striatal value signals in problem gamblers. *J. Neurosci.* **34**, 4750–4755 (2014).
7. Huys, Q. J. M., Maia, T. V & Frank, M. J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **19**, 404–413 (2016).
8. Farrell, S. & Lewandowsky, S. *Computational modeling of cognition and behavior*. *Computational Modeling of Cognition and Behavior* (Cambridge University Press, 2018). doi:10.1017/CBO9781316272503.
9. Lempert, K. M., Steinglass, J. E., Pinto, A., Kable, J. W. & Simpson, H. B. Can delay discounting deliver on the promise of RDoC? *Psychol. Med.* **49**, 190–199 (2019).
10. Miedl, S. F., Peters, J. & Büchel, C. Altered neural reward representations in pathological gamblers revealed by delay and probability discounting. *Arch. Gen. Psychiatry* **69**, 177–186 (2012).
11. Bickel, W. K., Koffarnus, M. N., Moody, L. & Wilson, A. G. The behavioral- and neuro-economic process of temporal discounting: A candidate behavioral marker of addiction. *Neuropharmacology* **76 Pt B**, 518–527 (2014).
12. Sutton & Barto. *Reinforcement Learning: An Introduction*. (MIT Press, 1998).
13. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).
14. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215

(2011).

15. Kool, W., Cushman, F. A. & Gershman, S. J. When Does Model-Based Control Pay Off? *PLoS Comput. Biol.* **12**, e1005090 (2016).

16. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goaldirected control. *Elife* **5**, 1–24 (2016).

17. Wyckmans, F. *et al.* Reduced model-based decision-making in gambling disorder. *Sci. Rep.* **9**, 19625 (2019).

18. Sebold, M. *et al.* Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology* **70**, 122–131 (2014).

19. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (Arlington, VA: Author., 2013).

20. Robinson, T. E. & Berridge, K. C. The incentive sensitization theory of addiction: Some current issues. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3137–3146 (2008).

21. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Res. Rev.* **18**, 247–291 (1993).

22. Berridge, K. C. & Robinson, T. E. Liking, wanting, and the incentive-sensitization theory of addiction. *Am. Psychol.* **71**, 670–679 (2016).

23. Volkow, N. D. *et al.* Cocaine cues and dopamine in dorsal striatum: Mechanism of craving in cocaine addiction. *J. Neurosci.* **26**, 6583–6588 (2006).

24. Carter, B. L. & Tiffany, S. T. Meta--analysis of cue--reactivity in addiction research. *Addiction* (1999).

25. Starcke, K., Antons, S., Trotzke, P. & Brand, M. Cue-reactivity in behavioral addictions: A meta-analysis and methodological considerations. *J. Behav. Addict.* **7**, 227–238 (2018).

26. Courtney, K. E., Schacht, J. P., Hutchison, K., Roche, D. J. O. & Ray, L. A. Neural substrates of cue reactivity: Association with treatment outcomes and relapse. *Addict. Biol.* **21**, 3–22 (2016).

27. Dixon, M. R., Jacobs, E. A., Sanders, S. & Carr, J. E. Contextual Control of Delay Discounting by Pathological Gamblers. *J. Appl. Behav. Anal.* **39**, 413–422 (2006).

28. Wagner, B., Mathar, D. & Peters, J. Gambling environment exposure increases temporal discounting but improves model-based control in regular slot-machine gamblers . *bioRxiv* (2021) doi:https://doi.org/10.1101/2021.07.15.452520.

29. Genauck, A. *et al.* Cue-induced effects on decision-making distinguish subjects with

gambling disorder from healthy controls. *Addict. Biol.* **25**, 1–10 (2020).

30. Tolliver, B. K. *et al.* Impaired cognitive performance in subjects with methamphetamine dependence during exposure to neutral versus methamphetamine-related cues. *Am. J. Drug Alcohol Abuse* **38**, 251–259 (2012).

31. Potenza, M. N. *et al.* Gambling Urges in Pathological Gambling. *Arch. Gen. Psychiatry* **60**, 828 (2003).

32. van Holst, R. J., van Holstein, M., van den Brink, W., Veltman, D. J. & Goudriaan, A. E. Response inhibition during cue reactivity in problem gamblers: An fmri study. *PLoS One* **7**, 1–10 (2012).

33. Brevers, D., Sescousse, G., Maurage, P. & Billieux, J. Examining Neural Reactivity to Gambling Cues in the Age of Online Betting. *Curr. Behav. Neurosci. Reports* **6**, 59–71 (2019).

34. Brevers, D., He, Q., Keller, B., Noël, X. & Bechara, A. Neural correlates of proactive and reactive motor response inhibition of gambling stimuli in frequent gamblers. *Sci. Rep.* **7**, 1–11 (2017).

35. Crockford, D. N., Goodyear, B., Edwards, J., Quickfall, J. & El-Guebaly, N. Cue-induced brain activity in pathological gamblers. *Biol. Psychiatry* **58**, 787–795 (2005).

36. Goudriaan, A. E., De Ruiter, M. B., Van Den Brink, W., Oosterlaan, J. & Veltman, D. J. Brain activation patterns associated with cue reactivity and craving in abstinent problem gamblers, heavy smokers and healthy controls: An fMRI study. *Addict. Biol.* **15**, 491–503 (2010).

37. Kober, H. *et al.* Brain Activity during Cocaine Craving and Gambling Urges: An fMRI Study. *Neuropsychopharmacology* **41**, 628–637 (2016).

38. Limbrick-Oldfield, E. H. *et al.* Neural substrates of cue reactivity and craving in gambling disorder. *Transl. Psychiatry* **7**, e992 (2017).

39. Bruder, L. R., Scharer, L. & Peters, J. Reliability assessment of temporal discounting measures in virtual reality environments. *Sci. Rep.* 1–16 (2021) doi:10.1038/s41598-021-86388-8.

40. Bouchard, S. *et al.* Using Virtual Reality in the Treatment of Gambling Disorder: The Development of a New Tool for Cognitive Behavior Therapy. *Front. psychiatry* **8**, 27 (2017).

41. Giroux, I. *et al.* Gambling exposure in virtual reality and modification of urge to gamble. *Cyberpsychol. Behav. Soc. Netw.* **16**, 224–231 (2013).

42. Ratcliff, R. & McKoon, G. The diffusion decision model: Theory and data for two-

choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).

43. Forstmann, B. U., Ratcliff, R. & Wagenmakers, E.-J. Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annu. Rev. Psychol.* **67**, 641–666 (2016).

44. Fontanesi, L., Gluth, S., Spektor, M. S. & Rieskamp, J. A reinforcement learning diffusion decision model for value-based decisions. *Psychon. Bull. Rev.* **26**, 1099–1121 (2019).

45. Pedersen, M. L., Frank, M. J. & Biele, G. The drift diffusion model as the choice rule in reinforcement learning. *Psychon. Bull. Rev.* **24**, 1234–1251 (2017).

46. Peters, J. & D'Esposito, M. The drift diffusion model as the choice rule in inter-temporal and risky choice: A case study in medial orbitofrontal cortex lesion patients and controls. *PLoS Comput. Biol.* **16**, 1–26 (2020).

47. Miletić, S., Boag, R. J. & Forstmann, B. U. Mutual benefits: Combining reinforcement learning with sequential sampling models. *Neuropsychologia* **136**, (2020).

48. Wagner, B., Clos, M., Sommer, T. & Peters, J. Dopaminergic modulation of human inter-temporal choice: a diffusion model analysis using the D2-receptor-antagonist haloperidol. *J. Neurosci.* **40**, 7936–7948 (2020).

49. Ballard, I. C. & McClure, S. M. Joint modeling of reaction times and choice improves parameter identifiability in reinforcement learning models. *J. Neurosci. Methods* **317**, 37–44 (2019).

50. Shahar, N. *et al.* Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Comput. Biol.* **15**, e1006803 (2019).

51. Wiehler, A. & Peters, J. Diffusion modeling reveals reinforcement learning impairments in gambling disorder that are linked to attenuated ventromedial prefrontal cortex value representations. *bioRxiv* 2020.06.03.131359 (2020) doi:10.1101/2020.06.03.131359.

52. Detez, L. *et al.* A Psychophysiological and Behavioural Study of Slot Machine Near-Misses Using Immersive Virtual Reality. *J. Gambl. Stud.* **35**, 929–944 (2019).

53. Dickinson, P., Gerling, K., Wilson, L. & Parke, A. Virtual reality as a platform for research in gambling behaviour. *Comput. Hum. Behav.* **107**, (2020).

54. Beck, A. T., Steer, R. A. & Brown, G. K. Beck depression inventory-II. *San Antonio* 490–498 (1996).

55. Derogatis, L. R., Lipman, R. S. & Covi, L. SCL-90: an outpatient psychiatric rating

scale--preliminary report. *Psychopharmacol Bull* 13–28 (1973).

56. Lesieur, H. & Blume, S. The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *Am. J. Psychiatry* (1987).

57. Petry, J. *Psychotherapie der Glücksspielsucht*. (Psychologie Verlags Union, 1996).

58. Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D. & Bradley, K. A. The AUDIT Alcohol Consumption Questions (AUDIT-C): An Effective Brief Screening Test for Problem Drinking. *Arch. Intern. Med.* 1789–1795 (1998).

59. Fagerström, K. Determinants of tobacco use and renaming the FTND to the Fagerstrom Test for Cigarette Dependence. *Nicotine Tob. Res.* **14**, 75–78 (2012).

60. Raylu, N. & Oei, T. P. S. The Gambling Related Cognitions Scale (GRCS): Development, confirmatory factor validation and psychometric properties. *Addiction* **99**, 757–769 (2004).

61. Steenbergh, T. A., Meyers, A. W., May, R. K. & Whelan, J. P. Development and validation of the Gamblers' Beliefs Questionnaire. *Psychol. Addict. Behav.* **16**, 143–149 (2002).

62. Braithwaite, J. J., Watson, D. G., Jones, R. & Rowe, M. A Guide for Analysing Electrodermal Activity (EDA)& Skin Conductance Responses (SCRs)for Psychological Experiments. *Psychophysiology* **2013**, (2013).

63. Unsworth, N., Heitz, R. P., Schrock, J. C. & Engle, R. W. An automated version of the operation span task. *Behav. Res. Methods* 498–505 (2005).

64. Redick, T. S. *et al.* Measuring Working Memory Capacity With Automated Complex Span Tasks. *Eur. J. Psychol. Assess.* **28**, 164–171 (2012).

65. van den Noort, M., Bosch, P., Haverkort, M. & Hugdahl, K. A Standard Computerized Version of the Reading Span Test in Different Languages. *Eur. J. Psychol. Assess.* **24**, 35–42 (2008).

66. Wechsler, D. *Wechsler adult intelligence scale--Fourth Edition (WAIS--IV)*. (NCS Pearson, 2008).

67. Myerson, J., Green, L. & Warusawitharana, M. Area Under the Curve As a Measure of Discounting. *J. Exp. Anal. Behav.* **76**, 235–243 (2001).

68. Mazur, J. E. An adjusting procedure for studying delayed reinforcement. **1987**, 55–73 (1987).

69. Green, L., Myerson, J. & Macaux, E. W. Temporal discounting when the choice is between two delayed rewards. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1121–1133 (2005).

70. Otto, A. R., Skatova, A., Madlon-Kay, S. & Daw, N. D. Cognitive Control Predicts Use of Model-based Reinforcement Learning. *J. Cogn. Neurosci.* **23**, (2015).
71. Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20941–20946 (2013).
72. Toyama, A., Katahira, K. & Ohira, H. Reinforcement Learning With Parsimonious Computation and a Forgetting Process. *Front Hum Neurosci.* **13**, 1–16 (2019).
73. Toyama, A., Katahira, K. & Ohira, H. A simple computational algorithm of model-based choice preference. *Cogn. Affect. Behav. Neurosci.* **17**, 764–783 (2017).
74. R Core Team. R: A Language and Environment for Statistical Computing. (2013).
75. Plummer, M. A Program for analysis of Bayesian graphical models. *Work. Pap.* (2003).
76. Wabersich, D. & Vandekerckhove, J. Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behav. Res. Methods* **46**, 15–28 (2014).
77. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, (2017).
78. Gelman, A. & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci. 7* **4**, 457–472 (1992).
79. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
80. Marsman, M. & Wagenmakers, E. J. Three Insights from a Bayesian Interpretation of the One-Sided P Value. *Educ. Psychol. Meas.* **77**, 529–539 (2017).
81. Beard, E., Dienes, Z., Muirhead, C. & West, R. Using Bayes factors for testing hypotheses about intervention effectiveness in addictions research. *Addiction* **111**, 2230–2247 (2016).
82. Bach, D. R., Friston, K. J. & Dolan, R. J. Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *Int. J. Psychophysiol.* **76**, 52–55 (2010).
83. Benedek, M. & Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80–91 (2010).
84. Kerby, D. The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Compr. Psychol.* **3**, (2014).
85. Rosenthal, R., Cooper, H. & Hedges, L. Parametric measures of effect size. *Handb.*

Res. Synth. **621(2)**, 231–244 (1994).

86. Tiffany, S. T. A critique of contemporary urge and craving research: Methodological, psychometric, and theoretical issues. *Adv. Behav. Res. Ther.* **14**, 123–139 (1992).

87. Clark, L., Boileau, I. & Zack, M. Neuroimaging of reward mechanisms in Gambling disorder: an integrative review. *Mol. Psychiatry* **24**, 674–693 (2019).

88. Havermans, R. C., Mulkens, S., Nederkoorn, C. & Jansen, A. The efficacy of cue exposure with response prevention in extinguishing drug and alcohol cue reactivity. ,. *Behav. Interv. Pract. Resid. Community-Based Clin. Programs* **22(2)**, 121–135 (2007).

89. Bordnick, P. S., Traylor, A. C., Graap, K. M., Copp, H. L. & Brooks, J. Virtual reality cue reactivity assessment: A case study in a teen smoker. *Appl. Psychophysiol. Biofeedback* **30**, 187–193 (2005).

90. Choi, J. S. *et al.* The effect of repeated virtual nicotine cue exposure therapy on the psychophysiological responses: A preliminary study. *Psychiatry Investig.* **8**, 155–160 (2011).

91. Van Holst, R. J., Veltman, D. J., Bchel, C., Van Den Brink, W. & Goudriaan, A. E. Distorted expectancy coding in problem gambling: Is the addictive in the anticipation? *Biol. Psychiatry* **71**, 741–748 (2012).

92. Antons, S., Brand, M. & Potenza, M. N. Neurobiology of cue-reactivity, craving, and inhibitory control in non-substance addictive behaviors. *J. Neurol. Sci.* **415**, (2020).

93. Gillan, C. M. *et al.* Comparison of the Association between Goal-Directed Planning and Self-reported Compulsivity vs Obsessive-Compulsive Disorder Diagnosis. *JAMA Psychiatry* **77**, 77–85 (2020).

94. Hogarth, L. Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory. *Neuropsychopharmacology* **45**, 720–735 (2020).

95. Frederiksen, J. G. *et al.* Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surg. Endosc.* **34**, 1244–1252 (2020).

96. Paas, F. Training Strategies for attaining transfer of Problemsolving skill in statistics: A cognitive load task. *J. Educ. Psychol.* **84**, 429–434 (1992).

Supplementary Materials

Supplementary Table 1. Priors for all parameters used in the computational modelling analysis of the temporal discounting task.

Model Parameter	Prior	Model Parameter (shift)	Prior
log(k)	uniform(-20,3)	s_k	normal(0,.2)
β	uniform(0,10)	s_β	normal(0,.2)
ν_{coeff}	normal(0,10)	$s_{\nu coeff}$	normal(0,10)
ν_{max}	normal(0,10)	$s_{\nu max}$	normal(0,10)
α	uniform(.01,3)	s_α	normal(0, 2)
z	uniform(.1, .9)	s_z	normal(0, 2)
τ	uniform(0.01, 2)	s_τ	normal(0, 2)

Supplementary Table 2. Priors for all parameters used in the computational modelling analysis of the 2-step task.

Model Parameter	Prior	Model Parameter (shift)	Prior
η_1	uniform(-4,4)	$s\eta_1$	normal(0, 1)
η_2	uniform(-4,4)	$s\eta_2$	normal(0, 1)
β_{MB}	normal(0,15)	$s_{\beta_{MB}}$	normal(0,3)
β_{MF}	normal(0,15)	$s_{\beta_{MF}}$	normal(0,3)
β_2	normal(0,15)	s_{β_2}	normal(0,3)
p	normal(0,10)	s_p	normal(0, 2)
$\nu_{coeff\ MB}$	normal(0,10)	$s_{\nu coeff\ MB}$	normal(0,10)
$\nu_{coeff\ MF}$	normal(0,10)	$s_{\nu coeff\ MF}$	normal(0,10)
ν_{coeff2}	normal(0,10)	$s_{\nu coeff2}$	normal(0,10)
ν_{max1}	normal(0,10)	$s_{\nu max1}$	normal(0,10)
ν_{max2}	normal(0,10)	$s_{\nu max2}$	normal(0,10)
α_1	uniform(.01,3)	$s_{\alpha 1}$	normal(0, 2)
α_2	uniform(.01,3)	$s_{\alpha 2}$	normal(0, 2)
τ_1	uniform(.01, 2)	$s_{\tau 1}$	normal(0, 2)
τ_2	uniform(.01, 2)	$s_{\tau 2}$	normal(0, 2)

Supplemental model-agnostic analysis 2-step task. As described in our preregistration we performed an additional model-agnostic HGLM analysis of the 2-step task. We modelled the probability to repeat the choice from the previous trial (1 if the choice was repeated and 0 if it was not) as a function of the transition in the previous trial (rare or common), the reward received in the previous trial (0 if the reward was lower than the mean of the rewards received in the past 20 trials and 1 otherwise), the group (gambling or non-gambling control), and finally the session ($VR_{neutral}$ vs $VR_{gambling}$).

The results were mostly in line with the results obtained from the other analysis. The effect of Reward ($z = -4.024$, $p < .001$) and the interaction Reward*Transition ($z = 9.308$, $p < .001$) were significant indicating that overall participants in both groups showed MB and MF

decision-making. Additionally, the three-way interaction Trans*Rew*Group was significant ($z = -6.389$, $p < .001$) indicating stronger MB decision-making in the non-gambling control group. Finally, none of the terms including the session were significant suggesting the absence of effects specifically caused by the VR environments.

Supplementary Table 3. Results of the supplementary HGLM analysis of the 2-step task. The probability to repeat S1 choices (pStay) is modelled as a function of previous transition and reward, the group affiliation and session. p-values printed in bold font are significant at a threshold of .05.

Fixed effects				
	Estimate	**Std. Error**	**z**	**p. val**
Intercept	1.274	.119	10.664	**<.001**
Transition	-.859	.109	-7.914	**<.001**
Reward	-.567	.141	-4.024	**<.001**
Session	.17	.106	1.607	.108
Group	-.794	.1	-7.924	**<.001**
Trans*Rew	1.758	.189	9.308	**<.001**
Trans*Sess	-.087	.131	-.661	.509
Rew*Sess	.011	.168	.065	.948
Trans*Group	.728	.123	5.893	**<.001**
Rew*Group	.554	.153	3.63	**<.001**
Sess*Group	-.208	.141	-1.476	.14
Trans*Rew*Sess	.036	.2	.178	.859
Trans*Rew*Group	-1.17	.183	-6.389	**<.001**
Trans*Sess*Group	-.037	.174	-.21	.834
Rew*Sess*Group	-.028	.218	-.129	.897
Trans*Rew*Sess*Group	.214	.261	.82	.412

Supplementary Table 4. Results of Wilcoxon signed rank tests of the SCL data for the gambling group. p-values printed in bold font are significant at a Bonferroni corrected threshold of .004.

Comparison				
	Z	**df**	**p. val**	**r**
B5 vs F1 $VR_{neutral}$	-4.099	30	**<.001**	.804
B5 vs F1 $VR_{gambling}$	-3.552	30	**<.001**	.661
$VR_{neutral}$ – $VR_{gambling}$	-.683	30	.495	-.134
F5 vs S1 $VR_{neutral}$	-2.141	30	.032	-.42
F5 vs S1 $VR_{gambling}$	.729	30	.466	.143
$VR_{neutral}$ – $VR_{gambling}$	-1.753	30	.008	-.344

Supplementary Table 5. Results of Wilcoxon signed rank tests of the SCL data for the control group. p-values printed in bold font are significant at a Bonferroni corrected threshold of .004.

Comparison	Z	df	p. val	r
B5 vs F1 $VR_{neutral}$	-3.552	28	**<.001**	.697
B5 vs F1 $VR_{gambling}$	-3.917	28	**<.001**	.768
$VR_{neutral}$ – $VR_{gambling}$	1.139	28	.255	.223
F5 vs S1 $VR_{neutral}$	-1.389	28	.165	-.272
F5 vs S1 $VR_{gambling}$	1.389	28	.165	.272
$VR_{neutral}$ – $VR_{gambling}$	-1.731	28	.084	-.339

Supplementary Table 6. Results of Wilcoxon signed rank tests of the nSCRs data for the gambling group. p-values printed in bold font are significant at a Bonferroni corrected threshold of .004.

Comparison	Z	df	p. val	r
B5 vs F1 $VR_{neutral}$	-.25	30	.802	-.049
B5 vs F1 $VR_{gambling}$	-2.811	30	.005	-.551
$VR_{neutral}$ – $VR_{gambling}$	1.684	30	.092	-.33
F5 vs S1 $VR_{neutral}$	-.713	30	.476	-.139
F5 vs S1 $VR_{gambling}$	1.116	30	.265	.219
$VR_{neutral}$ – $VR_{gambling}$	-1.048	30	.295	-.205

Supplementary Table 7. Results of Wilcoxon signed rank tests of the SCL data for the control group. p-values printed in bold font are significant at a Bonferroni corrected threshold of .004.

Comparison	Z	df	p. val	r
B5 vs F1 $VR_{neutral}$	.888	28	.375	.174
B5 vs F1 $VR_{gambling}$	.114	28	.891	.027
$VR_{neutral}$ – $VR_{gambling}$	.137	28	.89	.025
F5 vs S1 $VR_{neutral}$	-1.776	28	.076	-.348
F5 vs S1 $VR_{gambling}$	-2.852	28	.004	-.559
$VR_{neutral}$ – $VR_{gambling}$	.433	28	.665	.085

Supplementary Table 8. Results of Wilcoxon signed rank tests of the HR data for the gambling group. p-values printed in bold font are significant at a Bonferroni corrected threshold of .004.

Comparison	Z	Df	p. val	r
B5 vs F1 $VR_{neutral}$	-2.664	30	.008	-.523
B5 vs F1 $VR_{gambling}$	-3.53	30	**< .001**	-.692
$VR_{neutral}$ – $VR_{gambling}$	-.387	30	.699	-.076
F5 vs S1 $VR_{neutral}$	-1.001	30	.316	-.197
F5 vs S1 $VR_{gambling}$	.182	30	.855	-.036
$VR_{neutral}$ – $VR_{gambling}$	-.729	30	.4662	-.143

Supplementary Table 9. Results of Wilcoxon signed rank tests of the HR data for the control group. p-values printed in bold font are significant at a Bonferroni corrected threshold of .004.

Comparison	Z	df	p. val	r
B5 vs F1 $VR_{neutral}$	.615	28	.539	.121
B5 vs F1 $VR_{gambling}$	-1.389	28	.165	-.272
$VR_{neutral}$ – $VR_{gambling}$	1.23	28	.219	-.241
F5 vs S1 $VR_{neutral}$	-.046	28	.964	-.009
F5 vs S1 $VR_{gambling}$	1.526	28	.127	.299
$VR_{neutral}$ – $VR_{gambling}$	-.66	28	.509	-.13

General Discussion

To summarize, we conducted two studies employing a virtual reality (VR) cue-reactivity design for GD that combined the assessment of subjective, physiological, and behavioral cue-reactivity. The first study aimed to examine the reliability of temporal discounting measures in VR and standard lab environments in a group of non-gambling control participants. Additionally, we aimed to explore the feasibility of applying sequential sampling models to temporal discounting data obtained in VR. The study had three main results. First, temporal discounting measures (AUC and *log(k)*) showed a good to excellent test-retest reliability across the VR sessions and the lab session. This was true for the model-free AUC values as well as for the model-based *log(k)* from the hyperbolic models with the softmax choice rule and the drift-diffusion model (DDM) choice rules. Second, the test-retest reliability of the other parameters of the non-linear DDM_S modeling latent decision processes across sessions was substantially lower. However, the split-half reliability within each session was mostly good to excellent suggesting that lower test-retest reliability was likely a result of the current state of the participants and not caused by factors within the modelling process itself. Third, there was only little and inconclusive evidence that the different VR environments and the lab environment did modulate the discounting behavior of participants differently, indicating a lack of general VR effects on temporal discounting. Finally, the exposure to VR generally increased sympathetic physiological arousal as assessed by the tonic skin conductance level (SCL). This was not modulated differently by the two VR environments. Taken together the results of the first study demonstrate the methodological feasibility of a VR-based approach to behavioral and physiological testing in VR and lay the groundwork for the second study presented here.

The second study employed the VR design validated in the first study to investigate the subjective, behavioral, and physiological effects of VR gambling environment exposure in a group of regular gamblers (GD group) and matched non-gambling controls. Participants explored the two VR environments (neutral and gambling) and subsequently performed the temporal discounting tasks and the Two-Step task (TST). The study produced several important findings. Firstly, exposure to VR in general increased sympathetic physiological arousal as reflected in the tonic SCL. Secondly, we replicated earlier studies showing steeper temporal discounting[8,12,16,17] and decreased reliance on model-based (MB) control[24] in the GD group compared to non-gambling controls. Thirdly, the $VR_{gambling}$ environment selectively increased subjective craving in the GD group, as expected based on previous findings[5]. Interestingly however, none of the three physiological measures and neither of the two behavioral measures

showed conclusive evidence consistent with cue-reactivity in the GD group. Possible explanations for this observation are discussed below.

The behavioral results of both studies taken together warrant important conclusions for the application of VR in the context of cue-reactivity in addiction related disorders. In the first study we replicated earlier results showing a good to excellent test-retest reliability for temporal discounting measures[131,132] based on computational modeling and more model agnostic measures. By doing so we also further informed the discussion about the reliability of temporal discounting as a stable trait indicator[8,131,132,188–190]. While the test-retest reliability of the parameters in the different DDM models describing latent decision processes was substantially lower, we could still establish a strong within-session reliability. Therefore, we conclude that temporal discounting tendencies might be more of a stable trait-like characteristic, whereas latent decisions processes might be more dependent on states like mood and fatigue. Additionally, the non-decision time parameter (τ) showed a consistent pattern of increased perception and/or motor execution in VR, which should be the case as task execution within VR needs more movement due to VR controllers. These results are important because cue-reactivity research in VR and in general could profit from the usage of sequential sampling models, such as the DDM, as there are mixed results when it comes to the effects of addiction-related cues on reaction times (RT)[191–193]. Based on these encouraging results we included a group of frequent slot machine gamblers in the second study to investigate the cue-reactivity effects of our VR design. In addition, we added comprehensive analyses of the TST, which was impossible in the first study as the data we obtained was corrupted. In this second study we replicated results showing steeper temporal discounting in the GD group[8,12,16,17] across the model-agnostic AUC and the *log(k)* parameter of all models. In line with that, we also demonstrated reduced MB control across model agnostic analyses as well as computational modelling approaches in the GD group using the TST. This also replicated previous results[24]. Taken together, these results further increase our confidence in the application of established behavioral tasks in VR and the validity of using DDMs to analyze the resulting behavior. The most striking finding, however, was that we failed to observe conclusive evidence in favor of the existence of any form of behavioral cue-reactivity provoked by the $VR_{gambling}$ environment in both groups. We thus could not replicate the behavioral results observed in earlier studies conducted in real life gambling facilities, that showed behavioral cue-reactivity in participants suffering from GD[32,33]. The failure to replicate earlier findings is even more remarkable given that the GD group reported a selectively increased urge to gamble in the $VR_{gambling}$ environment, suggesting cue-reactivity on a subjective level. Our VR design thus induced the conscious wish

to gamble but failed to provoke any of the putative behavioral correlates associated with this induced craving.

Resonating with the behavioral results, both studies showed that exposure to VR increased sympathetic physiological arousal as measured with the SCL. This is again in line with prior research showing increased sympathetic arousal during VR exposure[3]. Nevertheless, we observed no evidence for any form of physiological cue-reactivity in both groups in the second study. Therefore, our design again did not provoke physiological cue-reactivity effects observed in previous studies assessing these responses using picture stimuli[194–196].

The lack of behavioral and/or physiological correlates of the reported increased subjective urge to gamble has important implications for future applications of VR exposure therapy and addiction science. Exposure therapy uses relevant stimuli to confront patients and train these patients to employ cognitive strategies to overcome craving and fear responses. This has been applied successfully in patients suffering from post-traumatic stress disorder and phobias using VR[197]. As the application of exposure therapy in addictions has not been as successful without VR in the past[198], our hope is that VR might supplement exposure therapy and thereby help to overcome these short comings. It is therefore important to note that, contrary to our expectations, VR might not provide an ecological validity high enough to produce real-life behavioral effects of gambling related contexts[32,33]. Importantly, this does not render VR generally unsuitable, as it has been successfully applied in exposure therapy elsewhere[197] and could induce craving in the studies presented here and conducted by other groups[5]. It is possible that cue-reactivity in VR depends on specific design features. Therefore, it might be that different VR designs would have produced the predicted effects. In addition to the specific design of the VR environment, more general effects of VR immersion might have prevented us from detecting the expected physiological and/or behavioral cue-reactivity effects. In particular, we observed a substantial increase in the SCL upon VR immersion in both studies and across all groups. The SCL then remained elevated throughout the experiment. An elevated sympathetic physiological arousal might reflect increased cognitive load during VR immersion[199,200]. This in turn could interfere with the behavioral tasks performed by the participants and preclude us from observing the expected cue-reactivity effects. In similar fashion, the elevated SCL correlated to VR-related cognitive load could have prevented the detection of smaller modulations of the SCL caused by reactivity to the different VR-environments. Future research could include measures of cognitive load in VR[201] to assess and possibly address this problem in future VR designs. Additionally, future research could train

participants in the usage of VR, reducing the cognitive load of participants in the actual study[202].

There is a set of limitations of this dissertation project that needs to be acknowledged here. First, in both studies participants spend about thirty to forty minutes immersed in VR. The behavioral tasks were performed after the two initial exploration phases. It is thus possible that participants were already fatigued or felt discomforted when the behavioral tasks started. These effects might not have been strong enough to produce observable effects when compared to the lab session in the first study but might nevertheless have blurred possible cue-reactivity effects in the second study. Future research should therefore contemplate to reduce the length of VR designs or split the exploration of the VR environments from the completion of behavioral tasks. Second, the physical space in which participants could move during VR immersion in both studies was constrained. The experimenter had to ensure the safety of participants by giving instructions if needed. These instructions might have reduced the immersion produced by our VR design. Furthermore, they might have caused unpredictable effects in the physiological measures, reducing our ability to detect important but subtle effects. Future studies would greatly benefit from increased automation of the VR environments. This could be realized by adding clearer boundaries, more space and recorded verbal instructions to our VR design. Third, the virtual slot machines used to build our virtual gambling environment did not exactly resemble the machines most frequently found in local gambling facilities. This discrepancy might have reduced the ecological validity of our VR design. However, the GD group in the second study reported an increased urge to gamble, suggesting an ecological validity sufficient to provoke craving. Future research could improve upon this by increasing the quality of graphical assets and design these to match current video slot machines more closely. Fourth, the first study only investigated the test-retest reliability and feasibility of employing the temporal discounting task in VR and did not conduct a thorough investigation of the same for the TST. It is therefore not clear how well the TST can be applied in VR. Even without VR the TST has been shown to produce rather poor test-retest reliabilities[46,189]. However, these results could be substantially improved by including RT data in the modelling process[46], as we did in the second study. Despite the unclear test-retest reliability of the TST, we have confidence in the results obtained with the TST in VR. We reproduced earlier group differences between non-gambling controls and the GD group[24], demonstrating that the TST produces similar results in VR. Furthermore, we demonstrated in the first study that sequential sampling models including RTs seems to produce valid results in VR. We employed a similar

choice rule for the TST in the second study as we did in the first and thus believe that the results can be interpreted with confidence.

Conclusion

Overall, the results obtained by both studies presented in the dissertation project at hand revealed further evidence for the validity of temporal discounting and the TST as possible diagnostic markers of GD. We demonstrated high reliability of the temporal discounting task and reproduced established group differences in decision-making between participants suffering from GD and non-gambling controls in both behavioral tasks[8,12,24]. Additionally, we demonstrated that behavioral data obtained by both tasks in VR can be meaningfully interpreted with comprehensive computational modelling, especially with models including RTs such as the DDM. In the context of cue-reactivity we found mixed results. While our design was effective in eliciting subjective craving in participants suffering from GD, we observed little evidence for behavioral or sympathetic physiological cue-reactivity. Therefore, caution is warranted when applying VR in experimental or therapeutical contexts, as established behavioral effects of real-life gambling environments[32,33] might not generally carry over to VR designs. Nevertheless, VR holds great promise for the research and therapy of addiction related disorders as it has been used successfully in other disorders[197]. This justifies the investment into VR environments for cue-reactivity studies and therapy, at least for now. Future research should delineate how cognitive load and ecological validity could be balanced in VR to create a more realistic VR experience.

References

1. Wang, Y. guang, Liu, M. hui & Shen, Z. hua. A virtual reality counterconditioning procedure to reduce methamphetamine cue-induced craving. *J. Psychiatr. Res.* **116**, 88–94 (2019).
2. Rubo, M. & Gamer, M. Stronger reactivity to social gaze in virtual reality compared to a classical laboratory environment. *Br. J. Psychol.* 1–14 (2020) doi:10.1111/bjop.12453.
3. Peterson, S. M., Furuichi, E. & Ferris, D. P. Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading. *PLoS One* **13**, 1–17 (2018).
4. Ghiţă, A. *et al.* Cue-Elicited Anxiety and Alcohol Craving as Indicators of the Validity of ALCO-VR Software: A Virtual Reality Study. *J. Clin. Med.* **8**, 1153 (2019).
5. Bouchard, S. *et al.* Using Virtual Reality in the Treatment of Gambling Disorder: The Development of a New Tool for Cognitive Behavior Therapy. *Front. psychiatry* **8**, 27 (2017).
6. Detez, L. *et al.* A Psychophysiological and Behavioural Study of Slot Machine Near-Misses Using Immersive Virtual Reality. *J. Gambl. Stud.* **35**, 929–944 (2019).
7. Dickinson, P., Gerling, K., Wilson, L. & Parke, A. Virtual reality as a platform for research in gambling behaviour. *Comput. Hum. Behav.* **107**, (2020).
8. Amlung, M. *et al.* Delay Discounting as a Transdiagnostic Process in Psychiatric Disorders: A Meta-analysis. *JAMA psychiatry* (2019) doi:10.1001/jamapsychiatry.2019.2102.
9. Kirby, K. N. & Petry, N. M. Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls. *Addiction* **99**, 461–471 (2004).
10. Peters, J. *et al.* Lower ventral striatal activation during reward anticipation in adolescent smokers. *Am. J. Psychiatry* **168**, 540–549 (2011).
11. Potenza, M. N. Review. The neurobiology of pathological gambling and drug addiction: An overview and new findings. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3181–3189 (2008).
12. Wiehler, A. & Peters, J. Reward-based decision making in pathological gambling: The roles of risk and delay. *Neurosci. Res.* **90**, 3–14 (2015).
13. Miedl, S. F., Büchel, C. & Peters, J. Cue-induced craving increases impulsivity via changes in striatal value signals in problem gamblers. *J. Neurosci.* **34**, 4750–4755

(2014).

14. Huys, Q. J. M., Maia, T. V & Frank, M. J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **19**, 404–413 (2016).
15. Farrell, S. & Lewandowsky, S. *Computational modeling of cognition and behavior*. *Computational Modeling of Cognition and Behavior* (Cambridge University Press, 2018). doi:10.1017/CBO9781316272503.
16. Lempert, K. M., Steinglass, J. E., Pinto, A., Kable, J. W. & Simpson, H. B. Can delay discounting deliver on the promise of RDoC? *Psychol. Med.* **49**, 190–199 (2019).
17. Miedl, S. F., Peters, J. & Büchel, C. Altered neural reward representations in pathological gamblers revealed by delay and probability discounting. *Arch. Gen. Psychiatry* **69**, 177–186 (2012).
18. Bickel, W. K., Moody, L. & Quisenberry, A. Computerized Working-Memory Training as a Candidate Adjunctive Treatment for Addiction. *Alcohol Res. Curr. Rev.* (2014).
19. Sutton & Barto. *Reinforcement Learning: An Introduction*. (MIT Press, 1998).
20. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).
21. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
22. Kool, W., Cushman, F. A. & Gershman, S. J. When Does Model-Based Control Pay Off? *PLoS Comput. Biol.* **12**, e1005090 (2016).
23. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goaldirected control. *Elife* **5**, 1–24 (2016).
24. Wyckmans, F. *et al.* Reduced model-based decision-making in gambling disorder. *Sci. Rep.* **9**, 19625 (2019).
25. American Psychological Association. Clinical Practice Guideline for the Treatment of Posttraumatic Stress Disorder (PTSD). *Washington, DC APA, Guidel. Dev. Panel Treat. Posttraumatic Stress Disord. Adults.* 139 (2017) doi:10.1162/jocn.
26. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Res. Rev.* **18**, 247–291 (1993).
27. Berridge, K. C. & Robinson, T. E. Liking, wanting, and the incentive-sensitization theory of addiction. *Am. Psychol.* **71**, 670–679 (2016).
28. Volkow, N. D. *et al.* Cocaine cues and dopamine in dorsal striatum: Mechanism of

craving in cocaine addiction. *J. Neurosci.* **26**, 6583–6588 (2006).

29. Carter, B. L. & Tiffany, S. T. Meta--analysis of cue--reactivity in addiction research. *Addiction* (1999).
30. Starcke, K., Antons, S., Trotzke, P. & Brand, M. Cue-reactivity in behavioral addictions: A meta-analysis and methodological considerations. *J. Behav. Addict.* **7**, 227–238 (2018).
31. Courtney, K. E., Schacht, J. P., Hutchison, K., Roche, D. J. O. & Ray, L. A. Neural substrates of cue reactivity: Association with treatment outcomes and relapse. *Addict. Biol.* **21**, 3–22 (2016).
32. Dixon, M. R., Jacobs, E. A., Sanders, S. & Carr, J. E. Contextual Control of Delay Discounting by Pathological Gamblers. *J. Appl. Behav. Anal.* **39**, 413–422 (2006).
33. Wagner, B., Mathar, D. & Peters, J. Gambling environment exposure increases temporal discounting but improves model-based control in regular slot-machine gamblers . *bioRxiv* (2021) doi:https://doi.org/10.1101/2021.07.15.452520.
34. Genauck, A. *et al.* Cue-induced effects on decision-making distinguish subjects with gambling disorder from healthy controls. *Addict. Biol.* **25**, 1–10 (2020).
35. Tolliver, B. K. *et al.* Impaired cognitive performance in subjects with methamphetamine dependence during exposure to neutral versus methamphetamine-related cues. *Am. J. Drug Alcohol Abuse* **38**, 251–259 (2012).
36. Potenza, M. N. *et al.* Gambling Urges in Pathological Gambling. *Arch. Gen. Psychiatry* **60**, 828 (2003).
37. van Holst, R. J., de Ruiter, M. B., van den Brink, W., Veltman, D. J. & Goudriaan, A. E. A voxel-based morphometry study comparing problem gamblers, alcohol abusers, and healthy controls. *Drug Alcohol Depend.* **124**, 142–148 (2012).
38. Brevers, D., Sescousse, G., Maurage, P. & Billieux, J. Examining Neural Reactivity to Gambling Cues in the Age of Online Betting. *Curr. Behav. Neurosci. Reports* **6**, 59–71 (2019).
39. Brevers, D. *et al.* Impulsive Action but Not Impulsive Choice Determines Problem Gambling Severity. *PLoS One* **7**, (2012).
40. Crockford, D. N., Goodyear, B., Edwards, J., Quickfall, J. & El-Guebaly, N. Cue-induced brain activity in pathological gamblers. *Biol. Psychiatry* **58**, 787–795 (2005).
41. Kober, H. *et al.* Brain Activity during Cocaine Craving and Gambling Urges: An fMRI Study. *Neuropsychopharmacology* **41**, 628–637 (2016).
42. Limbrick-Oldfield, E. H. *et al.* Neural substrates of cue reactivity and craving in

gambling disorder. *Transl. Psychiatry* **7**, e992 (2017).

43. Braithwaite, J. J., Watson, D. G., Jones, R. & Rowe, M. A Guide for Analysing Electrodermal Activity (EDA)& Skin Conductance Responses (SCRs)for Psychological Experiments. *Psychophysiology* **2013**, (2013).
44. Ratcliff, R. & McKoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
45. Ballard, I. C. & McClure, S. M. Joint modeling of reaction times and choice improves parameter identifiability in reinforcement learning models. *J. Neurosci. Methods* **317**, 37–44 (2019).
46. Shahar, N. *et al.* Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Comput. Biol.* **15**, e1006803 (2019).
47. Fontanesi, L., Gluth, S., Spektor, M. S. & Rieskamp, J. A reinforcement learning diffusion decision model for value-based decisions. *Psychon. Bull. Rev.* **26**, 1099–1121 (2019).
48. Pedersen, M. L., Frank, M. J. & Biele, G. The drift diffusion model as the choice rule in reinforcement learning. *Psychon. Bull. Rev.* **24**, 1234–1251 (2017).
49. Wagner, B., Clos, M., Sommer, T. & Peters, J. Dopaminergic modulation of human inter-temporal choice: a diffusion model analysis using the D2-receptor-antagonist haloperidol. *J. Neurosci.* **40**, 7936–7948 (2020).
50. Miletić, S., Boag, R. J. & Forstmann, B. U. Mutual benefits: Combining reinforcement learning with sequential sampling models. *Neuropsychologia* **136**, (2020).
51. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (Arlington, VA: Author., 2013).
52. Chamberlain, S. R. & Sahakian, B. J. The neuropsychiatry of impulsivity. *Curr. Opin. Psychiatry* **20**, 255–261 (2007).
53. Nordin, C. & Eklundh, T. Altered CSF 5-HIAA Disposition in Pathologic Male Gamblers. *CNS Spectr.* **4**, 25–33 (1999).
54. Spanagel, R., Herz, A. & Shippenberg, T. S. Opposing tonically active endogenous opioid systems modulate the mesolimbic dopaminergic pathway. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2046–2050 (1992).
55. Robinson, T. E. & Berridge, K. C. The incentive sensitization theory of addiction: Some current issues. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3137–3146 (2008).
56. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative

reward prediction error signal. *Neuron* **47**, 129–141 (2005).

57. Berridge, K. C. From prediction error to incentive salience: Mesolimbic computation of reward motivation. *Eur. J. Neurosci.* **35**, 1124–1143 (2012).
58. Ayano, G. Common Neurotransmitters : Criteria for Neurotransmitters , Key Locations , Classifications and Functions. **4**, 91–95 (2017).
59. Stahl, S. M. *Stahl's Essential Psychopharmacology: Neuroscientific Basis and Practical Applications*. (Cambridge University Press, 2008).
60. Lodish, H., Berk, A. & Zipursky, S. Neurotransmitters, Synapses, and Impulse Transmission. in *Molecular Cell Biology* (W.H. Freeman, 2000).
61. Breedlove, S., Watson, N. & Fraser, S. *Biological psychology: an introduction to behavioral, cognitive, and clinical neuroscience*. (Sinauer Associates, 2013).
62. Arias-Carrián, O., Stamelou, M., Murillo-Rodríguez, E., Menéndez-Gonzlez, M. & Pöppel, E. Dopaminergic reward system: A short integrative review. *Int. Arch. Med.* **3**, 1–6 (2010).
63. Beaulieu, J. M. & Gainetdinov, R. R. The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacol. Rev.* **63**, 182–217 (2011).
64. Vallone, D., Picetti, R. & Borrelli, E. Structure and function of dopamine receptors. *Neurosci. Biobehav. Rev.* **24**, 125–132 (2000).
65. Missale, C., Russel Nash, S., Robinson, S. W., Jaber, M. & Caron, M. G. Dopamine receptors: From structure to function. *Physiol. Rev.* **78**, 189–225 (1998).
66. Jaber, M., Robinson, S. W., Missale, C. & Caron, M. G. ~ Pergamon Review Dlopamine Receptors and Brain Function. *Neuropharmacology* **35**, 1503–1519 (1996).
67. Verhoeff, N. P. L. G. Radiotracer imaging of dopaminergic transmission in neuropsychiatric disorders. *Psychopharmacology (Berl).* **147**, 217–249 (1999).
68. Piccini, P. Neurodegenerative movement disorders: The contribution of functional imaging. *Curr. Opin. Neurol.* **17**, 459–466 (2004).
69. Harris-Warrick, R. M. & Johnson, B. R. Checks and balances in neuromodulation. *Front. Behav. Neurosci.* **4**, 1–9 (2010).
70. Grace, A. A. The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction* **95**, (2000).
71. Goto, Y., Otani, S. & Grace, A. A. The Yin and Yang of dopamine release: a new perspective. *Neuropharmacology* **53**, 583–587 (2007).
72. Schultz, W. Responses of midbrain dopamine neurons to behavioral trigger stimuli in

the monkey. *J. Neurophysiol.* **56**, 1439–1461 (1986).

73. Schultz, W. & Romo, R. Dopamine neurons of the monkey midbrain: Contingencies of responses to stimuli eliciting immediate behavioral reactions. *J. Neurophysiol.* **63**, 607–624 (1990).
74. Zhang, L., Doyon, W. M., Clark, J. J., Phillips, P. E. M. & Dani, J. A. Controls of tonic and phasic dopamine transmission in the dorsal and ventral striatum. *Mol. Pharmacol.* **76**, 396–404 (2009).
75. Arias-Carrión, O. & Pöppel, E. Dopamine, learning, and reward-seeking behavior. *Acta Neurobiol. Exp. (Wars).* **67**, 481–488 (2007).
76. Smith, Y. & Villalba, R. Striatal and extrastriatal dopamine in the basal ganglia: An overview of its anatomical organization in normal and Parkinsonian brains. *Mov. Disord.* **23**, 534–547 (2008).
77. Bissonette, G. B. & Roesch, M. R. Development and function of the midbrain dopamine system: What we know and what we need to. *Genes, Brain Behav.* **15**, 62–73 (2016).
78. Obeso, J. A. *et al.* Functional organization of the basal ganglia: Therapeutic implications for Parkinson's disease. *Mov. Disord.* **23**, 548–559 (2008).
79. Zaghloul, K. A. *et al.* Human substantia nigra neurons encode unexpected financial rewards. *Science (80-.).* **323**, 1496–1499 (2009).
80. Adamantidis, A. R. *et al.* Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *J. Neurosci.* **31**, 10829–10835 (2011).
81. Cools, R. & D'Esposito, M. Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol. Psychiatry* **69**, e113–e125 (2011).
82. Cools, R. Chemistry of the Adaptive Mind: Lessons from Dopamine. *Neuron* **104**, 113–131 (2019).
83. Wise, R. A. Forebrain substrates of reward and motivation. *J. Comp. Neurol.* **493**, 115–121 (2005).
84. Glimcher, P. W. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15647–15654 (2011).
85. Clark, L., Boileau, I. & Zack, M. Neuroimaging of reward mechanisms in Gambling disorder: an integrative review. *Mol. Psychiatry* **24**, 674–693 (2019).
86. Kim, K. M. *et al.* Optogenetic mimicry of the transient activation of dopamine neurons by natural reward is sufficient for operant reinforcement. *PLoS One* **7**, 1–8 (2012).

87. Leyton, M. & Vezina, P. On cue: striatal ups and downs in addictions. *Biol Psychiatry* **72**, (2012).
88. Ross, S. & Peselow, E. The neurobiology of addictive disorders. *Clin. Neuropharmacol.* **32**, 269–276 (2009).
89. Pessiglione, M. *et al.* How the Brain Translates Money into Force: A Neuroimaging Study of Subliminal Motivation. *Science* **316**, 904–907 (2007).
90. Olds, J. & Milner, P. Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of Rat Brain. *J. Comp. Physiol. Psychol.* **47**, 419–427 (1954).
91. Pavlov, I. Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. (1927).
92. Schultz, W. Reward prediction error. *Curr. Biol.* **27**, R369–R371 (2017).
93. Schultz, W. Dopamine signals for reward value and risk: Basic and recent data. *Behav. Brain Funct.* **6**, 1–10 (2010).
94. Schultz, W. Dopamine reward prediction-error signalling: A two-component response. *Nat. Rev. Neurosci.* **17**, 183–195 (2016).
95. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045 (2006).
96. D'Ardenne, K., McClure, S., Nystrom, L. & Cohen, L. BOLD Responses Reflecting Dopaminergic Signals in the Human Ventral Tegmental Area. *Science* (2008).
97. Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete Coding of Reward Dopamine Neurons. *Science (80-.).* **299**, 1898–1902 (2003).
98. Redish, A. D. Addiction as a computational process gone awry. *Science (80-.).* **306**, 1944–1947 (2004).
99. Volkow, N. D., Fowler, J. S. & Wang, G. J. The addicted human brain: Insights from imaging studies. *J. Clin. Invest.* **111**, 1444–1451 (2003).
100. Everitt, B. J. & Robbins, T. W. Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nat. Neurosci.* **8**, 1481–1489 (2005).
101. Hare, T. A., Camerer, C. F. & Rangel, A. Self-control in decision-making involves modulation of the vmPFC valuation system. *Science (80-.).* **324**, 646–649 (2009).
102. Kringelbach, M. The hedonic brain: A functional neuroanatomy of human pleasure. in *Pleasure of the brain* 202–221 (Oxford University Press, 2010).
103. Berridge, K. C. & Kringelbach, M. Pleasure systems in the brain. *Neuron* **86**, 646–664

(2015).

104. Smith, K. S., Berridge, K. C. & Aldridge, J. W. Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proc. Natl. Acad. Sci. U. S. A.* **108**, (2011).
105. Hickey, C. & Peelen, M. Neural Mechanisms of Incentive Salience in Naturalistic Human Vision. *Neuron* **85**, 512–518 (2015).
106. Berridge, K. C. & Aldridge, J. W. Decision utility, the brain and pursuit of hedonic goals. *Soc. Cog.* **26**, 621–646 (2008).
107. Wassum, K. M., Ostlund, S. B., Balleine, B. W. & Maidment, N. T. Differential dependence of Pavlovian incentive motivation and instrumental incentive learning processes on dopamine signaling. *Learn. Mem.* **18**, 475–483 (2011).
108. Giroux, I. *et al.* Gambling exposure in virtual reality and modification of urge to gamble. *Cyberpsychol. Behav. Soc. Netw.* **16**, 224–231 (2013).
109. Bickel, W. K., Koffarnus, M. N., Moody, L. & Wilson, A. G. The behavioral- and neuro-economic process of temporal discounting: A candidate behavioral marker of addiction. *Neuropharmacology* **76 Pt B**, 518–527 (2014).
110. Alessi, S. & Petry, N. M. Pathological gambling severity is associated with impulsivity in a delay discounting procedure. *Behav. Processes* **64**, 345–354 (2003).
111. Hogarth, L. & Chase, H. W. Parallel goal-directed and habitual control of human drug-seeking: Implications for dependence vulnerability. *J. Exp. Psychol. Anim. Behav. Process.* **37**, 261–276 (2011).
112. Jasinska, A. J., Stein, E. A., Kaiser, J., Naumer, M. J. & Yalachkov, Y. Factors modulating neural reactivity to drug cues in addiction: A survey of human neuroimaging studies. *Neurosci. Biobehav. Rev.* **38**, 1–16 (2014).
113. Wilson, S., Sayette, M. & Fiez, J. Prefrontal responses to drug cues: a neurocognitive analysis. *Nat. Neurosci.* **7**, 211–214 (2004).
114. Meng, Y. jing *et al.* Reward pathway dysfunction in gambling disorder: A meta-analysis of functional magnetic resonance imaging studies. *Behav. Brain Res.* **275**, 243–251 (2014).
115. Goudriaan, A. E., De Ruiter, M. B., Van Den Brink, W., Oosterlaan, J. & Veltman, D. J. Brain activation patterns associated with cue reactivity and craving in abstinent problem gamblers, heavy smokers and healthy controls: An fMRI study. *Addict. Biol.* **15**, 491–503 (2010).
116. Noori, H. R., Cosa Linan, A. & Spanagel, R. Largely overlapping neuronal substrates

of reactivity to drug, gambling, food and sexual cues: A comprehensive meta-analysis. *Eur. Neuropsychopharmacol.* **26**, 1419–1430 (2016).

117. Tremblay, A. M., Desmond, R. C., Poulos, C. X. & Zack, M. Haloperidol modifies instrumental aspects of slot machine gambling in pathological gamblers and healthy controls. *Addict. Biol.* **16**, 467–484 (2011).
118. Grüsser, S. M., Plöntzke, B. & Albrecht, U. Pathologisches glücksspiel. Eine empirische untersuchung des verlangens nach einem stoffungebundenen suchtmittel. *Nervenarzt* **76**, 592–596 (2005).
119. SHARPE, L., TARRIER, N., SCHOTTE, D. & SPENCE, S. H. The role of autonomic arousal in problem gambling. *Addiction* **90**, 1529–1540 (1995).
120. Schacht, J. P., Anton, R. F. & Myrick, H. Functional neuroimaging studies of alcohol cue reactivity: A quantitative meta-analysis and systematic review. *Addict. Biol.* **18**, 121–133 (2013).
121. Sjoerds, Z., Van Den Brink, W., Beekman, A. T. F., Penninx, B. W. J. H. & Veltman, D. J. Cue reactivity is associated with duration and severity of alcohol dependence: An fMRI study. *PLoS One* **9**, (2014).
122. Sescousse, G., Barbalat, G., Domenech, P. & Dreher, J. C. Imbalance in the sensitivity to different types of rewards in pathological gambling. *Brain* **136**, 2527–2538 (2013).
123. Van Holst, R. J., Veltman, D. J., Bchel, C., Van Den Brink, W. & Goudriaan, A. E. Distorted expectancy coding in problem gambling: Is the addictive in the anticipation? *Biol. Psychiatry* **71**, 741–748 (2012).
124. Sescousse, G. *et al.* Amplified Striatal Responses to Near-Miss Outcomes in Pathological Gamblers. *Neuropsychopharmacology* **41**, 2614–2623 (2016).
125. Gelskov, S. V., Madsen, K. H., Ramsøy, T. Z. & Siebner, H. R. Aberrant neural signatures of decision-making: Pathological gamblers display cortico-striatal hypersensitivity to extreme gambles. *Neuroimage* **128**, 342–352 (2016).
126. Gamito, P. *et al.* Eliciting nicotine craving with virtual smoking cues. *Cyberpsychology, Behav. Soc. Netw.* **17**, 556–561 (2014).
127. Tiffany, S. T. A critique of contemporary urge and craving research: Methodological, psychometric, and theoretical issues. *Adv. Behav. Res. Ther.* **14**, 123–139 (1992).
128. Dixon, M. R., Marley, J. & Jacobs, E. A. Delay Discounting by Pathological Gamblers. *J. Appl. Behav. Anal.* 449–458 (2003).
129. Bickel, W. K. & Marsch, L. A. Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction* **96**, 73–86 (2001).

130. Andrade, L. F. & Petry, N. M. Delay and probability discounting in pathological gamblers with and without a history of substance use problems. *Psychopharmacology (Berl).* **219**, 491–499 (2012).
131. Kirby, K. N. One-year temporal stability of delay-discount rates. *Psychon. Bull. Rev.* **16**, 457–462 (2009).
132. Ohmura, Y., Takahashi, T., Kitamura, N. & Wehr, P. Three-month stability of delay and probability discounting measures. *Exp. Clin. Psychopharmacol.* **14**, 318–328 (2006).
133. Bruder, L. R., Scharer, L. & Peters, J. Reliability assessment of temporal discounting measures in virtual reality environments. *Sci. Rep.* 1–16 (2021) doi:10.1038/s41598-021-86388-8.
134. Anokhin, A. P., Golosheykin, S., Grant, J. D. & Heath, A. C. Heritability of delay discounting in adolescence: A longitudinal twin study. *Behav. Genet.* **41**, 175–183 (2011).
135. Figner, B. *et al.* Lateral prefrontal cortex and self-control in intertemporal choice. *Nat. Neurosci.* **13**, 538–539 (2010).
136. Potenza, M. N. *et al.* An fMRI Stroop Task Study of Ventromedial Prefrontal Cortical Function in Pathological Gamblers. *Am. J. Psychiatry* **160**, 1990–1994 (2003).
137. Kim, B. K. & Zauberman, G. Can Victoria's Secret change the future? A subjective time perception account of sexual-cue effects on impatience. *J. Exp. Psychol. Gen.* **142**, 328–335 (2013).
138. Li, X. The effects of appetitive stimuli on out-of-domain consumption impatience. *J. Consum. Res.* **34**, 649–656 (2008).
139. Mackillop, J. Integrating behavioral economics and behavioral genetics: Delayed reward discounting as an endophenotype for addictive disorders. *J. Exp. Anal. Behav.* **99**, 14–31 (2013).
140. Myerson, J., Green, L. & Warusawitharana, M. Area Under the Curve As a Measure of Discounting. *J. Exp. Anal. Behav.* **76**, 235–243 (2001).
141. Mazur, J. E. An adjusting procedure for studying delayed reinforcement. **1987**, 55–73 (1987).
142. Green, L., Myerson, J. & Macaux, E. W. Temporal discounting when the choice is between two delayed rewards. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1121–1133 (2005).
143. Wiehler, A. & Peters, J. Diffusion modeling reveals reinforcement learning

impairments in gambling disorder that are linked to attenuated ventromedial prefrontal cortex value representations. *bioRxiv* 2020.06.03.131359 (2020) doi:10.1101/2020.06.03.131359.

144. Peters, J. & D'Esposito, M. The drift diffusion model as the choice rule in inter-temporal and risky choice: A case study in medial orbitofrontal cortex lesion patients and controls. *PLoS Comput. Biol.* **16**, 1–26 (2020).
145. Voon, V., Reiter, A., Sebold, M. & Groman, S. Model-Based Control in Dimensional Psychiatry. *Biol. Psychiatry* **82**, 391–400 (2017).
146. Kahneman, D. A. A perspective on judgment and choice: Mapping bounded rationality. *Am. Psychol.* **58**, 697–720 (2003).
147. Fudenberg, D. & Levine, D. A dual self model of impulse control. *Am. Econ. Rev.* **96**, 1449–76 (2006).
148. Sloman, S. The empirical case for two systems of reasoning. *Psychol. Bull.* **119**, 3–22 (1996).
149. Maia, T. V. Two-factor theory, the actor-critic model, and conditioned avoidance. *Learn. Behav.* **38**, 50–67 (2010).
150. Balleine, B. W. & O'Doherty, J. P. Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* **35**, 48–69 (2010).
151. Daw, N., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–11 (2005).
152. Doll, B. B., Simon, D. A. & Daw, N. D. The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* **22**, 1075–1081 (2012).
153. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585–595 (2010).
154. Peters, J. & Büchel, C. Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. *J. Neurosci.* **29**, 15727–15734 (2009).
155. Deserno, L. *et al.* Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1595–1600 (2015).
156. Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D. & Dolan, R. J. Disruption of Dorsolateral Prefrontal Cortex Decreases Model-Based in Favor of

Model-free Control in Humans. *Neuron* **80**, 914–919 (2013).

157. Wunderlich, K., Smittenaar, P. & Dolan, R. J. Dopamine Enhances Model-Based over Model-Free Choice Behavior. *Neuron* **75**, 418–424 (2012).
158. Voon, V. *et al.* Disorders of compulsivity: A common bias towards learning habits. *Mol. Psychiatry* **20**, 345–352 (2015).
159. Sebold, M. *et al.* When Habits Are Dangerous: Alcohol Expectancies and Habitual Decision Making Predict Relapse in Alcohol Dependence. *Biol. Psychiatry* **82**, 847–856 (2017).
160. Sebold, M. *et al.* Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology* **70**, 122–131 (2014).
161. Doñamayor, N., Strelchuk, D., Baek, K., Banca, P. & Voon, V. The involuntary nature of binge drinking: goal directedness and awareness of intention. *Addict. Biol.* **23**, 515–526 (2018).
162. Romanczuk-Seiferth, N., Koehler, S., Dreesen, C., Wüstenberg, T. & Heinz, A. Pathological gambling and alcohol dependence: Neural disturbances in reward and loss avoidance processing. *Addict. Biol.* **20**, 557–569 (2015).
163. Antons, S., Brand, M. & Potenza, M. N. Neurobiology of cue-reactivity, craving, and inhibitory control in non-substance addictive behaviors. *J. Neurol. Sci.* **415**, (2020).
164. Gillan, C. M. *et al.* Comparison of the Association between Goal-Directed Planning and Self-reported Compulsivity vs Obsessive-Compulsive Disorder Diagnosis. *JAMA Psychiatry* **77**, 77–85 (2020).
165. Otto, A. R., Skatova, A., Madlon-Kay, S. & Daw, N. D. Cognitive Control Predicts Use of Model-based Reinforcement Learning. *J. Cogn. Neurosci.* **23**, (2015).
166. Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20941–20946 (2013).
167. Toyama, A., Katahira, K. & Ohira, H. A simple computational algorithm of model-based choice preference. *Cogn. Affect. Behav. Neurosci.* **17**, 764–783 (2017).
168. Toyama, A., Katahira, K. & Ohira, H. Reinforcement Learning With Parsimonious Computation and a Forgetting Process. *Front Hum Neurosci.* **13**, 1–16 (2019).
169. Hastings, A. W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 97–109 (1970).
170. Gelman, A. & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci. 7* **4**, 457–472 (1992).

171. Marsman, M. & Wagenmakers, E. J. Three Insights from a Bayesian Interpretation of the One-Sided P Value. *Educ. Psychol. Meas.* **77**, 529–539 (2017).
172. Beard, E., Dienes, Z., Muirhead, C. & West, R. Using Bayes factors for testing hypotheses about intervention effectiveness in addictions research. *Addiction* **111**, 2230–2247 (2016).
173. Plummer, M. A Program for analysis of Bayesian graphical models. *Work. Pap.* (2003).
174. R Core Team. R: A Language and Environment for Statistical Computing. (2013).
175. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, (2017).
176. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *J.R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 583–639 (2002).
177. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
178. Critchley, H. D. Electrodermal responses: What happens in the brain. *Neuroscientist* **8**, 132–142 (2002).
179. Shibasaki, M. & Crandall, C. G. Mechanisms and controllers of eccrine sweating in humans. *Front. Biosci. - Sch.* **2 S**, 685–696 (2010).
180. Schandry, R. *Biologische Psychologie*. (Beltz, 2011).
181. Tranel, D. (2000). Electrodermal activity in cognitive neuroscience: Neuroanatomical and neuropsycho- logical correlates. In R. D. Lane & L. Nadel (Eds.), Cognitive neuroscience of emotion. in *Cognitive neuroscience of emotion.* (Oxford University Press, 2000).
182. Damasio, A. R. *Descartes' error: Emotion, reason, and the human brain.* (Grosset/ Putnam, 1994).
183. Critchley, H. D., Elliott, R., Mathias, C. J. & Dolan, R. J. Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. *J. Neurosci.* **20**, 3033–3040 (2000).
184. Critchley, H. D., Mathias, C. J. & Dolan, R. J. Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron* **29**, 537–545 (2001).
185. Martin-Soelch, C., Linthicum, J. & Ernst, M. Appetitive conditioning: Neural bases and implications for psychopathology. *Neurosci. Biobehav. Rev.* **31**, 426–440 (2007).
186. Benedek, M. & Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80–91 (2010).

187. Kao, Y.-H., Chao, P. & Wey, C.-L. Design and Validation of a New PPG Module to Acquire High-Quality Physiological Signals for High-Accuracy Biomedical Sensing. *IEEE J. Sel. Top. Quantum Electron.* **25**, (2019).
188. Eisenberg, I. W. *et al.* Applying novel technologies and methods to inform the ontology of self-regulation. *Behav. Res. Ther.* **101**, 46–57 (2018).
189. Enkavi, A. Z. *et al.* Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 5472–5477 (2019).
190. Odum, A. L. Delay discounting: Trait variable? *Behav. Processes* **87**, 1–9 (2011).
191. Juliano, L. M. & Brandon, T. H. Reactivity to instructed smoking availability and environmental cues: evidence with urge and reaction time. *Exp. Clin. Psychopharmacol.* **6(1)**, (1998).
192. Sayette, M. A. *et al.* The effects of cue exposure on reaction time in male alcoholics. *J. Stud. Alcohol* 629–633 (1994).
193. Vollstädt-Klein, S. *et al.* Validating incentive salience with functional magnetic resonance imaging: Association between mesolimbic cue reactivity and attentional bias in alcohol-dependent patients. *Addict. Biol.* **17**, 807–816 (2012).
194. Havermans, R. C., Mulkens, S., Nederkoorn, C. & Jansen, A. The efficacy of cue exposure with response prevention in extinguishing drug and alcohol cue reactivity. ,. *Behav. Interv. Pract. Resid. Community-Based Clin. Programs* **22(2)**, 121–135 (2007).
195. Bordnick, P. S., Traylor, A. C., Graap, K. M., Copp, H. L. & Brooks, J. Virtual reality cue reactivity assessment: A case study in a teen smoker. *Appl. Psychophysiol. Biofeedback* **30**, 187–193 (2005).
196. Choi, J. S. *et al.* The effect of repeated virtual nicotine cue exposure therapy on the psychophysiological responses: A preliminary study. *Psychiatry Investig.* **8**, 155–160 (2011).
197. Maples-Keller, J. L., Yasinski, C., Manjin, N. & Rothbaum, B. O. Virtual Reality-Enhanced Extinction of Phobias and Post-Traumatic Stress. *Neurotherapeutics* **14**, 554–563 (2017).
198. Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J. & Owen, A. M. The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage* **50**, 1313–1319 (2010).
199. Frederiksen, J. G. *et al.* Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surg. Endosc.* **34**, 1244–1252 (2020).

200. Paas, F. Training Strategies for attaining transfer of Problemsolving skill in statistics: A cognitive load task. *J. Educ. Psychol.* **84**, 429–434 (1992).
201. Collins, J. *et al.* Measuring cognitive load and insight: A methodology exemplified in a virtual reality learning context. *Proc. - 2019 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2019* 351–362 (2019) doi:10.1109/ISMAR.2019.00033.
202. Andersen, S. A. W., Mikkelsen, P. T., Konge, L., Cayé-Thomasen, P. & Sorensen, M. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *Laryngoscope* **126**, E74–E79 (2015).

Abbreviations

anterior cingulate cortex (ACC)
area under the curve (AUC)
autonomous nervous system (ANS)
beats per minute (BPM)
caudate nucleus (CN)
continuous decomposition analysis (CDA)
cyclic adenosine monophosphate (cAMP)
dopamine (DA)
dorsolateral prefrontal cortex (dlPFC)
drift diffusion model (DDM)
electrodermal activity (EDA)
functional magnetic resonance imaging (fMRI)
gambling disorder (GD)
guanine nucleotide binding protein (G-protein)
hierarchical generalized linear model (HGLM)
impulse response function (IRF)
Incentive-sensitization theory (IST)
larger but delayed rewards (LL)
Markov Chain Monte Carlo (MCMC)
model-free (MF)
model-based (MB)
nucleus accumbens (NAcc)
number of spontaneous non-specific SCRs (nSCRs)
orbitofrontal cortex (OFC)
photoplethysmogram (PPG)
prefrontal cortex (PFC)
reinforcement learning (RL)
response times (RT)
retrorubral field (RRF)
reward prediction error (RPE)
skin conductance level (SCL)
skin conductive responses (SCRs)
smaller immediate reward (SS)
South Oaks Gambling Screen (SOGS)
state-action values (Q-values)
substance-use-disorders (SUD)
substantia nigra pars compacta (SNc)
temporal difference (TD)
Two-Step task (TST)
ventral striatum (VS)
ventral tegmental area (VTA)
ventromedial prefrontal cortex (vmPFC)
virtual reality (VR)
Watanabe-Akaike Information Criterion (WAIC)