Learning in anticipation of reward and punishment: Perspectives across the human lifespan

Matthew J. Betts[1,2,#], Anni Richter[3,#], Lieke de Boer[4], Jana Tegelbeckers[5], Valentina Perosa[1], Rumana Chowdhury[6], Raymond J. Dolan[7,8], Constanze Seidenbecher[3,12], Björn H. Schott[3,9,10,12], Emrah Düzel[1,2,11,12], Marc Guitart-Masip[4,8,*], Kerstin Krauel[5,12,*]

[1]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany
[2]Institute of Cognitive Neurology and Dementia Research, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
[3]Department of Neurochemistry and Molecular Biology, Department of Behavioral Neurology, Leibniz Institute for Neurobiology, Magdeburg, Germany
[4]Aging Research Centre, Karolinska Institute, Stockholm, Sweden
[5]Department of Child and Adolescent Psychiatry and Psychotherapy, Otto von Guericke University Magdeburg, Magdeburg, Germany
[6] Leeds Teaching Hospital NHS Trust, Leeds, UK
[7]The Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London, WC1N 3BG, UK
[8]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London
[9]Department of Psychiatry, University Medicine Göttingen, Göttingen, Germany
[10]Department of Neurology, University of Magdeburg, Magdeburg, Germany
[11]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, UK
[12]Center for Behavioral Brain Sciences, University of Magdeburg, Magdeburg, Germany

[#, *]Denotes equal contribution

Corresponding author: Matthew J. Betts

Key words: action, reward, Pavlovian bias, instrumental learning, lifespan

Abstract

Pavlovian biases influence the interaction between action and valence by coupling reward seeking to action invigoration and punishment avoidance to action suppression. In this study we used an orthogonalised go/no-go task to investigate learning in 247 individuals across the human lifespan (7-80 years) to demonstrate that all participants, independently of age, demonstrated an influence of Pavlovian control. Computational modeling revealed peak performance in young adults was attributable to greater sensitivity to both rewards and punishment. However in children and adolescents an increased bias towards action but not reward sensitivity was observed. In contrast, reduced learning in midlife and older adults was accompanied with decreased reward sensitivity and especially punishment sensitivity. These findings reveal distinct learning capabilities across the human lifespan that cannot be probed using conventional *go*/reward *no-go*/punishment style paradigms that have important implications in life-long education.

Introduction

Adaptive behavioral choices maximize reward and minimize punishment. In order to achieve this optimization, humans are equipped with two broad classes of mechanisms. Firstly, a Pavlovian controller directly ties affectively important outcomes together with learned predictions of these outcomes, to valence-dependent stereotyped behavioral responses via a conceived hard-wired mechanism. Secondly, a more flexible, instrumental controller learns choices on the basis of contingent consequences (Dayan and Balleine, 2002). Generally these controllers favor the same choices rendering learning fast and efficient. However in some circumstances, Pavlovian influences may impair instrumental learning by prescribing the opposite of the instrumental controller leading to sub-optimal behavioral choices (Dayan *et al.*, 2006; Guitart-Masip *et al.*, 2010).

A number of studies have shown that the success of instrumental learning critically depends on whether reward or avoidance of punishment is paired with action or inhibition. Using a go/no-go task that independently dissociates, i.e. orthogonalizes, action and valence, young adults demonstrate striking asymmetry in instrumental learning by being better at learning to emit a behavioral response in anticipation of reward and better at withholding a response in anticipation of punishment (Guitart-Masip *et al.*, 2012*b*; Cavanagh *et al.*, 2013, Chowdhury *et al.*, 2013*b*; Guitart-Masip *et al.*, 2014; Richter *et al.*, 2014). Computational modeling approaches in young adults have shown this asymmetry in instrumental action learning is due to a Pavlovian coupling between action and valence expectation where the strength of this bias relates to failure in learning the conflicting conditions: withholding a response in anticipation of reward and emitting a response in anticipation of punishment (Guitart-Masip *et al.*, 2012*b*; Cavanagh *et al.*, 2013; Guitart-Masip *et al.*, 2013). However a recent study has shown that this coupling may also be due to instrumental mechanisms (Swart *et al.*, 2017). Learning success in this go/no-go task requires flexibility, inhibition and the ability to use feedback and to detect reward contingencies. All of these abilities may critically rely on prefrontal cortex (PFC) dependent executive functions, which develop during adolescence but also decline in older age (Kray *et al.*, 2004; Zelazo *et al.*, 2004). Whilst most of the previous studies using the aforementioned task have investigated young adults, little attention has focused on how the processes underlying instrumental

learning and potential conflict with Pavlovian control may change during differential periods across the lifespan— namely during childhood and adolescence, young, midlife and older age.

It has been widely stated that adolescents are highly sensitive to reward, which may contribute to increased risky behavior during this developmental period (Casey *et al.*, 2008). However it has also been suggested that this reward sensitivity may be adaptive by promoting learning and exploration — critical for transition into adulthood (Spear, 2000; Casey, 2015). A recent study demonstrated that adolescents learn to preferentially seek rewards rather than to avoid punishments, whereas young adults learn both behaviors equally well (Davidow *et al.*, 2016). However, previous studies have not dissociated reward sensitivity from action learning, and it remains thus unclear if this interpretation may be confounded by action requirements or to what extent changes in reward sensitivity may influence the strength of coupling between action and valence. This is particularly relevant in light of differential functional and anatomical development of limbic regions, such as the striatum and cognitive control regions during adolescence (Blakemore and Robbins, 2012; Shulman *et al.*, 2016). Such, asymmetrical development may translate into differential Pavlovian and instrumental strategies used by children and adolescents compared to those employed in adulthood.

The human brain also undergoes substantial change during normal aging, which has been associated with numerous cognitive changes (Bäckman *et al.*, 2006; Lindenberger, 2014). However it is not known how age-related differences in Pavlovian and instrumental control may impact behavioral inflexibility in older adults or whether such age-related changes are already evident in midlife adulthood. Previous work has shown that administration of the dopamine precursor L-DOPA enhances the Pavlovian influence of potential reward (Rutledge *et al.*, 2015) and may restore reward prediction errors in old age (Chowdhury *et al.*, 2013*a*). Coupled with the known age-related decline in the integrity of the dopaminergic system (Karrer *et al.*, 2017), a loss of functional dopamine may lead to a decrease in Pavlovian control in older age. Alternatively, previous studies have shown that the PFC is involved in overcoming the Pavlovian bias in young adults (Guitart-Masip *et al.*, 2012*b*; Cavanagh *et al.*, 2013). Thus, decreased functionality of the PFC as a result of normal aging could also lead to increased Pavlovian biases in older adults.

The objective of this study was to explore how acquisition of optimal/adaptive behavioral choices is differentially altered across the lifespan using an established go/no-go task that orthogonalizes action and valence. In addition we used computational modeling to investigate whether Pavlovian congruency is already evident in healthy children and adolescents in support of a hard-wired bias and compared behavior to young, midlife and older adults. Finally we assessed the extent to which action and valence are coupled across the lifespan and whether flexible learning to orthogonalize these two axes of behavioral control may vary with age.

Results

247 healthy individuals from four age groups, namely children and adolescents (age: 7 to 16, n = 68), young adults (age: 18 to 30, n = 77), midlife adults (age: 31 to 60, n = 58), and older adults (age 61 to 80, n = 44) performed a previously established valenced go/no-go probabilistic learning task (Guitart-Masip *et al.*, 2012*b*). Participants had to learn through trial and error which of four fractal cues, preceding an easy visual target detection task, indicated the need (1) to respond to obtain a monetary reward (*go to win*), (2) to respond to avoid a monetary loss (*go to avoid losing*), (3) to withhold a response to obtain a monetary reward (*no-go to win*) and (4) to withhold a response to avoid a monetary loss (*no-go to avoid losing*) (see Figure 1). The outcome was probabilistic whereby correct feedback was given in 80% of win/loss conditions. To characterize the interaction of action and valence across the lifespan we investigated acquisition of and overall accuracy in each of the four conditions and analyzed differential influences of Pavlovian and instrumental learning using computational modeling.
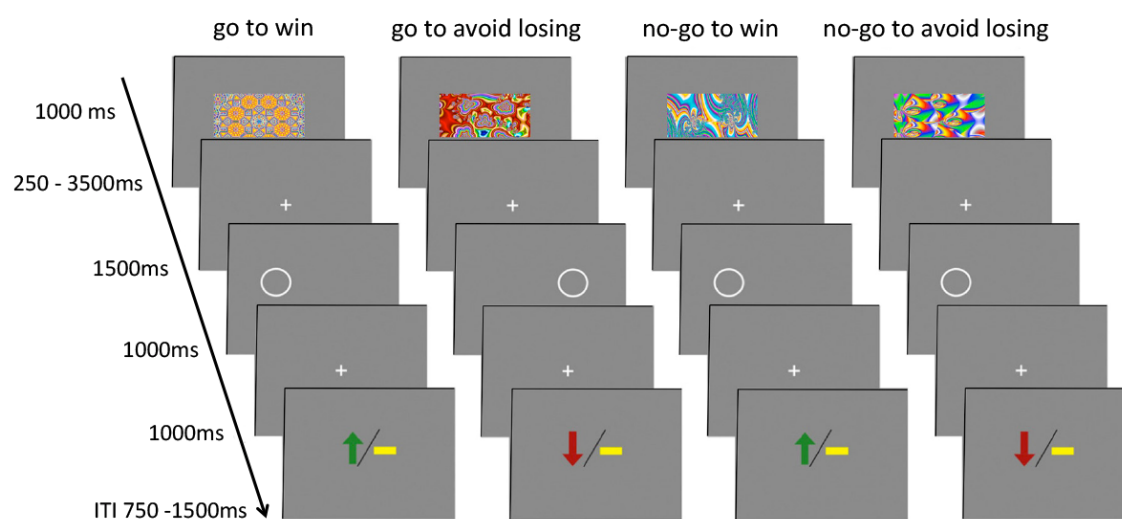

Figure 1: Probabilistic monetary go/no-go task. Participants had to learn over the course of 60 trials per condition, which fractal image was associated with responding or withholding a response to achieve a successful outcome (win or avoid losing). Responses indicated whether the circle presented after the fractal was located on the right or left side. Correct feedback was only provided in 80% of the trials. Abbreviations: ITI – inter-trial interval.

Pavlovian bias across the lifespan

Performance as defined by percentage of correct (optimal) choices was assessed using a 4-way analysis of covariance (ANCOVA) with action (go/no-go), valence (win/lose) and time ($1^{st}/2^{nd}$ half) as within-subject factors, age group (children and adolescents/young adults/midlife/older adults) as a between-subject factor and gender as a covariate. As in previous studies (Guitart-Masip *et al.*, 2012*b*, 2014; Richter *et al.*, 2014) we observed main effects of time ($F_{1,243} = 17.18$, p < 0.0001) and action ($F_{1,243} = 38.67$, p < 0.0001) as well as an action x time interaction ($F_{1,243} = 8.27$, p =0.004), and an action x valence interaction ($F_{1,243} = 14.21$, p < 0.0005). Subjects demonstrated an increase in performance from the first to the second half of the experiment ($t_{246} = -14.90$, *p* < 0.0001) and performed better in conditions requiring a *go* choice than in trials requiring a *no-go* choice ($t_{246} = 15.22$, *p* < 0.0001). Participants demonstrated an initial bias toward *go* responses (1^{st}(*go-nogo*)>2^{nd}(*go-nogo*): $t_{246} = 7.23$, *p* < 0.0001), and showed greater accuracy for go choices when the outcome was a reward (*go to win* > *go to avoid losing*: $t_{246} = 10.45$, *p* < 0.0001) and for no-go choices when the outcome was avoidance of losses (*no-go to avoid losing > no-go to win*: $t_{246} = 7.82$, *p* < 0.0001).

Of particular interest in this study, were age-related differences in performance across groups (see Figure 2 and SI Table 1 for statistics). We observed a main effect of group (*p* < 0.0001), and interactions of group with action (*p* < 0.0001), valence (*p* = 0.033) and time (*p* = 0.001). Moreover a threefold interaction of action x time x group (*p* = 0.040) was found demonstrating a differential ability to learn action responses across the lifespan. However, no action x valence x group interaction was observed ($F_{3,243} = 2.08$, *p* = 0.10) suggesting human learning and decision making is influenced by Pavlovian control irrespective of age (Figure 2). No significant interactions with gender were observed (all p ≥ 0.2).
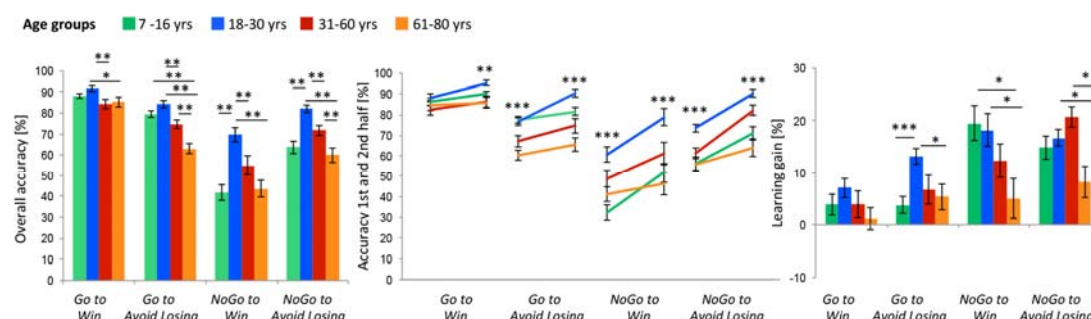
Figure 2: Overall behavioral performance across age groups. On the left, mean performance accuracy (%) ± standard error (SEM) across conditions in children and adolescents (7 to 16 years, n=69), young adults (18 to 30 years, n=77), midlife adults (31 to 60 years, n=58) and older adults (61 to 80 years, n=44) for all trials. In the middle panel, line graphs show mean performance for responses in the first (30 trials) and second half of trials (last 30 trials) across groups and on the right, bar plots demonstrate learning gain between first and second 30 trials across groups. ***(p<0.001), **(p<0.001), *(p<0.008; Holm-Bonferroni-corrected for 6 tests), indicate significant differences compared to young adults (independent samples t-tests).

To compare differences in performance across age groups, we performed independent samples t-tests and applied Holm-Bonferroni correction for six tests. We focused on overall accuracy, initial performance (first half) and ability to learn across trials (learning gain (2nd - 1st half) as shown in SI Table 1.

Young adults showed superior overall performance across all conditions compared to any other age group (all p<0.0001) and higher learning gain compared to older adults (p<0.0001). In *go trials,* children and adolescents demonstrated comparable performance to young adults whilst midlife and older adults demonstrated a significant age-related decline in *go* accuracy (all p<0.001). However, children and adolescents demonstrated significantly reduced learning gain for *go* trials compared to young adults (p=0.001). In *no-go* trials, young adults demonstrated superior overall performance compared to all other age groups irrespective of valence (all p<0.001). Taken together, children and adolescents demonstrated high performance in go trials comparable to young adults but not in no-go trials, indicating a predominant preference for action over withdrawal responses. In comparison, midlife and older adults demonstrated a significant age-related decline in both go and no-go trials compared to younger adults.

Interestingly assessment of learning gain revealed that children and adolescents were able to overcome this initial action bias through learning, demonstrating comparable

learning gain to young adults for no-go trials (Figure 2). Midlife adults also showed poor initial *no-go* accuracy compared to younger adults (all p<0.001), but could increase performance in the second half of the experiment, significantly more so than older adults (p=0.001). In comparison, older adults demonstrated poor initial *no-go* accuracy (all p≤0.007) compared to young and midlife adults and considerable inflexibility in learning no-go responses across trials compared to all other groups (all p<0.004). Lastly older adults demonstrated the worst performance in punishment trials compared to all other age groups (all p≤0.001).

For completeness, reaction times for *go* responses are reported in the supplementary information (SI Figure 1). Their interpretation warrants caution as participants were explicitly instructed to respond accurately, while speed was not emphasized. Generally young and older adults demonstrated the fastest and slowest responses respectively (all p<0.0001) while children/adolescents and midlife adults did not differ (p=1.0).

Parameterizing learning and biases using computational modeling

To identify instrumental and Pavlovian components of the observed asymmetry during learning, six nested reinforcement learning (RL) models were fitted to the behavioral data (see SI Materials and Methods) using the expectation maximization approach as previously described (Huys *et al.*, 2011, Guitart-Masip *et al.*, 2012*b*). All six computational models were fit to the data using a single distribution for all participants. This fitting procedure was, therefore, blind to the existence of different groups with putatively different parameter values. Our computational modeling approach demonstrated that the marked asymmetry in learning (i.e. superior performance in *go to win* and *no-go to avoid losing* compared to *go to avoid losing* and *no-go to win*) could be attributed to an interaction between instrumental and Pavlovian control mechanisms (Guitart-Masip *et al.*, 2012*b*). The best account of the data was provided by the model including a static *action bias*, *Pavlovian bias*, *reward* and *punishment sensitivity, learning rate* and an *irreducible noise* parameter consistent with previous studies using this task (see Table 1) (Guitart-Masip *et al.*, 2012*b*; Cavanagh *et al.*, 2013; Guitart-Masip *et al.*, 2014).

Table 1: Integrated Bayesian Information Criterion (iBIC) for all tested models

Model no.	Model parameters	No. of parameters	Likelihood	Pseudo-R^2	iBIC
1	ε, ρ	2	-28166	0.32	56,376
2	ε, ρ, ξ	3	-27977	0.32	56,019
3	ε, ρ, ξ, b	4	-25653	0.38	51,394
4	ε, $ρ_{win}$, $ρ_{lose}$, ξ, b	5	-25064	0.39	50,239
5	ε, $ρ_{win}$, $ρ_{lose}$, ξ, b, $π_{fluct}$	6	-24680	0.40	49,493
6	**ε, $ρ_{win}$, $ρ_{lose}$, ξ, b, $π_{constant}$**	**6**	**-24519**	**0.41**	**49,170**

The winning model statistics are highlighted in bold font: **ε**, learning rate; **ρ**win, weighting of reward on win trials; **ρ**lose, weighting of punishments on lose trials; **b**, go bias; **π**, Pavlovian bias; **ξ**, irreducible noise. iBIC, integrated Bayesian information criterion. Smaller values indicate a better model fit.

We found associative learning in children and adolescents was inherently driven by an action bias that was significantly greater than that observed in all adult age groups (Kruskal-Wallis with Wilcoxon rank-sum *post-hoc* tests; $χ^2$ (3) = 15.42, all p ≤ 0.01, children and adolescents > young adults, midlife adults, older adults). Effect sizes determined using Cohen's *d* between children/adolescents and all adult age groups were d ≥ 0.5. Parameters indicative of instrumental learning such as *sensitivity to reward* ($χ^2$ (3) = 41.37; p < .0001), *sensitivity to punishment* ($χ^2$ (3) = 39.33; p < 0.0001) and *learning rate* ($χ^2$ (3) = 18.37, p < 0.001) followed an inverted U-shaped distribution, peaking during young adulthood (Figure 3). Children and adolescents demonstrated comparable *sensitivity to reward*, *punishment* and *learning rate* compared midlife adults (all p≥ 0.2) but increased *sensitivity to reward* and *punishment* (Wilcoxon rank-sum *post-hoc* tests; both Z ≤ -3.35, p ≤ 0.001; effect size d ≥ 0.5) but not *learning rate* (p=0.67) compared to older adults.

Both midlife and older adults showed a comparable age-related decline in *reward sensitivity* (both Z ≤ -4.02, p < 0.0001; effect size d ≥ 0.6) and *learning rate* (both Z ≤ -2.9, p ≤ 0.003; effect size d ≥ 0.5) compared to young adults, however older adults showed the lowest estimates for *sensitivity to punishment* (Z = -5.69, p < 0.0001; effect size d = 1.0, compared to young adults) as summarized in Figure 3. The modeling approach revealed no significant difference in the *Pavlovian bias* across groups (p = 0.14). For the *irreducible noise* parameter values ranged between 0 and 1 where 1 indicates no distortion to the softmax rule and 0 indicates a flat softmax i.e. random choices regardless of value differences. We observed an age-related difference in the *irreducible*

noise parameter (χ^2 (3) = 28.82; p < 0.0001) whereby values were highest in young adults and lowest in older adults demonstrating younger adults' performance was more tightly captured by the winning model.
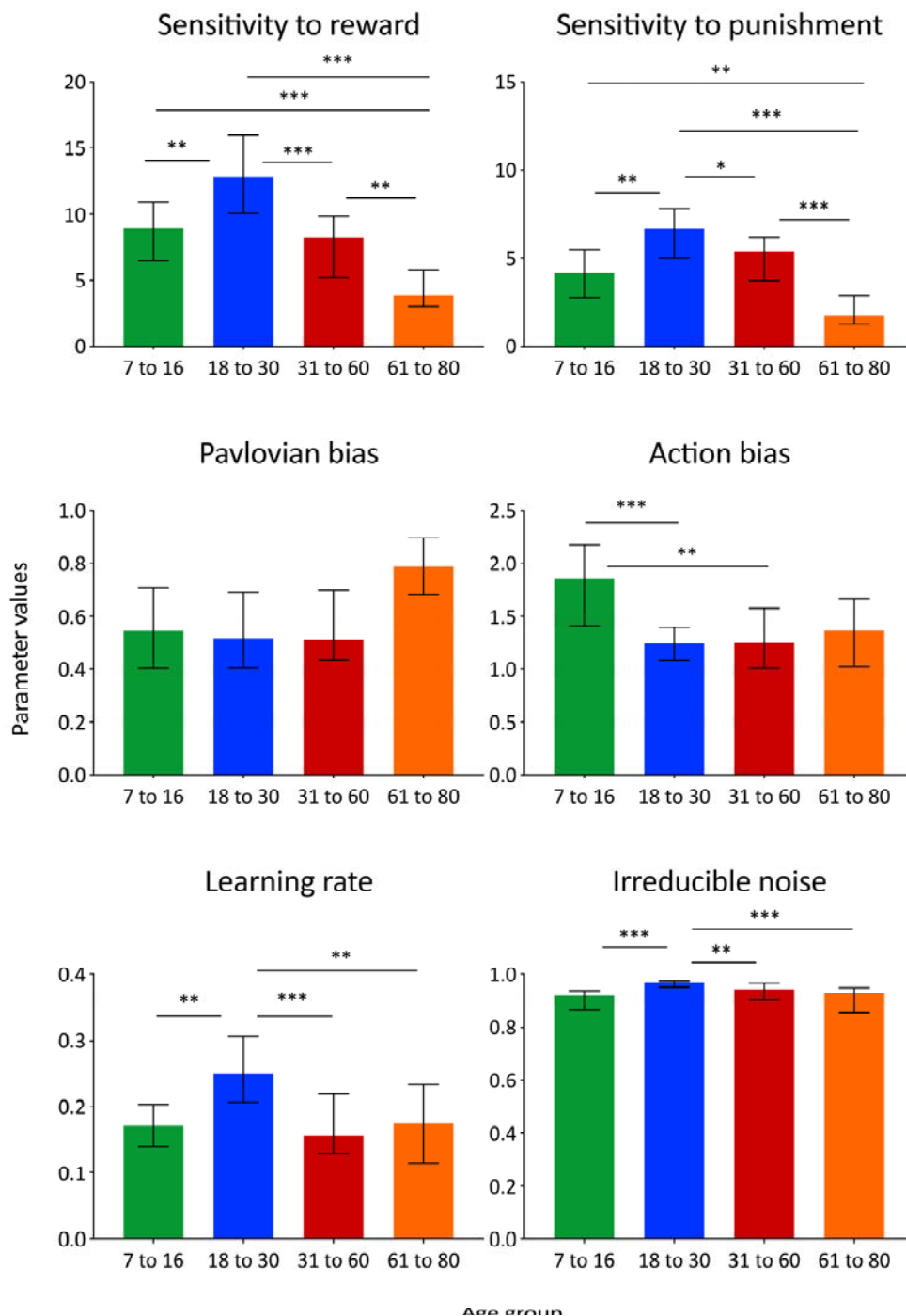

Figure 3: Modeled behavioural performance across the lifespan. Modeling parameters (median (5-95th percentile)) derived from the winning model plotted across age groups. ***(p<0.001), **(p<0.001), *(p<0.008; Holm-Bonferroni-corrected for 6 tests), indicate significant differences between age groups (Wilcoxon ranked-sum tests).

Discussion

In this study we reveal, that individual performance in a valanced go/no-go task across the lifespan (7-80 years) is influenced by Pavlovian control independent of age. Furthermore, the ability to successfully orthogonalize action and valence was characterized by an inverted U-shape distribution with peak performance observed in young adults. Computational modeling revealed that superior performance in younger adults compared to all other age groups, was attributable to greater sensitivity to outcomes (both to reward and punishment) coupled with a low *action bias*. In contrast reduced performance in children and adolescents was attributable to an increased bias towards action but not increased reward sensitivity. In midlife and older adults, an age-related decline in performance was attributable to a decrease in *learning rate*, *reward sensitivity* and especially *punishment sensitivity*. Taken together this study reveals novel age-related discrepancies that cannot be probed using typical *go*/reward *no-go*/punishment style paradigms.

This study set out to investigate the influence of Pavlovian control on instrumental learning responses coupled to action and valence using a go/no-go task. Our results revealed that the participants, independently of age, exhibited an influence of Pavlovian control whereby they were better at initiating an action to gain a reward (*go to win*) compared to punishment (*go to avoid losing*) but also withdrawing an action to avoid punishment (*no-go to avoid losing*) compared to gaining a reward (*no-go to win*). The striking asymmetry in performance across conditions observed here is consistent with previous studies using the same task (Guitart-Masip *et al.*, 2012*b*; Cavanagh *et al.*, 2013, Chowdhury *et al.*, 2013*b*; Guitart-Masip *et al.*, 2014; Richter *et al.*, 2014; de Boer *et al.*, 2019). Furthermore, computational modeling in young adults has previously shown this pattern of behavior can be captured by a model incorporating a Pavlovian bias, where the strength of this bias is related to impaired learning of the conflicting conditions: *no-go to win* and *go to avoid losing* (Guitart-Masip *et al.*, 2012*b*). Here we extend this work to demonstrate the same model can effectively capture learning behavior across the human lifespan. Interestingly we observed no significant difference in the *Pavlovian bias* across the lifespan and instead found performance to be influenced by age-related differences in the ability of the instrumental system to learn the appropriate choice (go

or no-go) for each fractal image as quantified by differences in *learning rate* and *sensitivity* to both *reward* and *punishment*.

While children and adolescents demonstrated Pavlovian responding consistent with all other age groups, we observed an underlying preference for action responses regardless of valence which goes against the interpretation that there is an overall increase in reward sensitivity during adolescence and rather demonstrates this is only true for rewards coupled to action. In fact, our computational modelling show that children and adolescents show decreased *reward sensitivity* when compared to younger adults. As this is the first study to our knowledge to dissociate or orthogonalize action and valence learning in children and adolescents, our results suggest that increased reward sensitivity reported in previous studies (Galvan *et al.*, 2006; Van Leijenhorst *et al.*, 2010; Somerville *et al.*, 2011; Cohen *et al.*, 2016; Palminteri *et al.*, 2016) may be confounded by preference towards action responses. Furthermore these findings highlight the importance of controlling for action tendencies when investigating reward learning in children and adolescents. Previous work in young adults using the same *go/no-go* task has shown that activity in inferior frontal gyrus (IFG), a region known to be involved in action inhibition (Aron *et al.*, 2014), is associated with *no-go* learning and successful instrumental control (Guitart-Masip *et al.*, 2012*b*). Thus protracted development of prefrontal projections to subcortical regions as previously described (Ziegler *et al.*, 2017) may lead to less top-down regulation and increased action bias in children and adolescents.

Reports on changes in instrumental learning across the lifespan have so far focused nearly exclusively on the comparison of young and older adults. Thus, the investigation of midlife adults signifying the transition phase between young and late adulthood has been comparatively neglected. In our study, midlife adults showed significantly compromised performance in *go* conditions compared to younger adults but not to the extent of older adults demonstrating middle age is a pivotal period across the lifespan. The significant age-related decrease in *reward* but also *punishment sensitivity* in midlife adults may reflect the developmental tasks required at this stage of life. In young adults, self-focused goals such as obtaining autonomy or starting a professional career require increased approach motivation and reward responsiveness. However in midlife the

focus shifts to taking care or maximizing the benefit of others (e.g. partner, children, ageing parents) and disengaging from unattainable goals allowing for lower approach motivation (Windsor *et al.*, 2012).

In older adults, we found poorer overall performance in all task conditions, but also considerable inability to learn across trials compared to young adults. A study by Schott *et al.*, (2007) suggested that, compared to young adults, older adults exhibited profoundly reduced mesolimbic activation during reward anticipation, although did activate the ventral striatum during reward feedback. Similarly, Chowdhury *et al.*, (2013*a*) demonstrated that older adults do not show a representation of expected value in the ventral striatum when performing a probabilistic reward learning under basal conditions and that an expected value representation was only observed after boosting the dopaminergic system with L-DOPA. Furthermore, older adults performing a probabilistic reward learning task show an attenuation of value anticipation in the ventromedial prefrontal cortex (vmPFC) that predicts performance in the probabilistic learning task (de Boer *et al.*, 2017). These findings suggest that whilst general reward processing may be intact in older adults, they are impaired in learning the predictive value of probabilistic reward cues. In fact, previous studies have shown that reduced learning in older adults is associated with deficits in the integration and updating of reward information when rewards are uncertain and delivered from probabilistic outcomes (Eppinger *et al.*, 2008; Hämmerer *et al.*, 2011). Additionally, Samanez-Larkin *et al.*, (2007) showed a particularly strong age-related impairment of ventral striatal loss anticipation in older adults. Taken together, these findings are compatible with our current observation of decreased instrumental learning in older adults, which was attributable to an age-related decrease in *reward* but particularly *punishment sensitivity* in the learning model.

A striking finding in our data is that the impairment of the instrumental learning in older adults was especially manifest as decreased performance in the go conditions. The dopamine system is involved in generating active motivated behaviour (Niv *et al.*, 2007; Salamone and Correa, 2012) and instrumental learning through reward-prediction errors (Schultz, 2010). Dopamine depletion leads to decreased motor activity and/or reduced motivation to gain rewards (Palmiter, 2008; Salamone and Correa, 2012). Similarly, previous findings using a variant of the same *go/no-go* task have shown that L-

DOPA administration invigorated instrumental responding regardless of valence (Guitart-Masip *et al.*, 2012*a*). Therefore, an age-related decline in dopaminergic function as previously described (Bäckman *et al.*, 2006; Karrer *et al.*, 2017), could modulate motivation or vigor of actions independently of valence and may explain the overall decrease in *go* performance observed in older compared to younger adults.

Determining the impact of an aging dopaminergic system on performance in the valanced go/no-go task, however, is not straightforward. Most previous studies support the notion that dopamine facilitates the action by valence interaction during learning (see however Guitart-Masip 2014). A study investigating a genetic variant linked to dopamine D2 receptor expression also highlights a modulatory role for genetic variability within the dopaminergic system in individual learning differences of action-valence interactions (Richter *et al.*, 2014). Another study has shown that boosting dopamine with methylphenidate increases the action by valence interaction in participants with high working memory capacity, a proxy for higher dopamine synthesis capacity (Swart *et al.*, 2017). Finally, a recent PET imaging study has shown that the strength of the action by valence interaction scales with the availability of dopamine D1 receptors in the dorsal striatum independent of age (de Boer *et al.*, 2019). Based on this evidence, one would have predicted a decreased Pavlovian bias in older adults. However, learning the correct contingencies of the orthogonalized go/no-go task may critically rely on high-level cognitive functions, thus lifespan differences reported here may also relate to interindividual differences in working memory and long-term memory. These cognitive functions may also be compromised as a result of age-related decline in grey and white matter integrity (Draganski *et al.*, 2011; Samanez-Larkin *et al.*, 2012, Chowdhury *et al.*, 2013*b*; Callaghan *et al.*, 2014; Acosta-Cabronero *et al.*, 2016; Steiger *et al.*, 2016; van de Vijver *et al.*, 2016) which may influence the ability of the instrumental system to learn the task contingencies as indexed by *learning rate*, *reward* and *punishment sensitivity*. Therefore, the effect of decreased dopamine function on the strength of the Pavlovian system may be shadowed by the effects of an age-related decrease in executive functions or instrumental abilities related to structural decline.

No significant age-related differences in the *Pavlovian bias* was observed across age groups, although a marked increase in the *Pavlovian bias* was observed in older adults. The substantial heterogeneity observed across all age groups also indicates that the

strength of the *Pavlovian bias* across the lifespan cannot be entirely accounted for by age. It has been recently argued that an instrumental bias may also influence learning responses coupled to action and valence (Swart *et al.*, 2017) implying a facilitated learning of go responses leading to reward, as well as an impaired unlearning of no-go responses leading to punishment. Thus, it will be interesting in future work to determine how age-related differences in instrumental biases also influence action-valence learning across the human lifespan.

Finally, a general limitation of the study is that we can only infer developmental and age-related differences from a cross-sectional study. These effects should not be assumed to represent underlying causal relationships, nor can we comment on lifespan trajectories. Future longitudinal studies will be needed to address these questions.

Our results demonstrate a dichotomy between prepotent biases that influence learning at either end of the lifespan with a predominant preference for action responses in children/adolescents compared to reduced instrumental learning from both reward and punishment in older age. Collectively, our results emphasize the importance of orthogonally manipulating action requirements and outcome valence to further understand instrumental learning capabilities across different stages of the human lifespan. Such characteristics may underline important evolutionary conserved mechanisms i.e. heightened action learning in adolescents necessary to facilitate active exploration and independence into adulthood or alternatively adaptation to maintain decision-making abilities despite declining learning ability in old age.

Materials and Methods

Participants

Overall, 247 individuals between the age of 7 to 80 participated in the current study and were assigned to one of four age groups: *children and adolescents* (age: 7 to 16, n = 69, 48 (69.6%) males), *young adults* (age: 18 to 30, n = 77, 45 (58.4%) males), *midlife adults* (age: 31 to 60, n = 58, 24 (41.4%) males), and *older adults* (age 61 to 80, n = 44, 20 (45.5%) males) using age ranges in line with previously reported lifespan studies (Tymula *et al.*, 2013; Yang *et al.*, 2016). It was ensured either by a standardized telephone interview or personal clinical interview that none of the participants were affected by a present or past neurological or psychiatric illness, alcohol, or drug abuse or were using centrally acting medication. Cognitive abilities were explicitly assessed in children, adolescents and older adults to ascertain they had intact global cognitive performance (for details see SI Materials and Methods). Adult participants were only included if they had finished compulsory education (minimum 12 years). All participants received detailed oral and written information about the study and gave written consent. For minors, informed consent from children and adolescents as well as their parents was required for participation. The study was approved by the local ethics committee of the University of Magdeburg, Faculty of Medicine, and followed the ethical standards of the Declaration of Helsinki.

Task and Procedure

All participants had to learn which of four fractal cues, preceding an easy visual target detection task, indicated the need (1) to respond to gain a reward (*go to win*), (2) to respond to avoid losing (*go to avoid losing*), (3) to withhold a response to gain a reward (*no-go to win*), and (4) to withhold a response to avoid losing (*no-go to avoid losing*). After display of the fractal cue (1000 ms), participants were presented with the target detection task (1500 ms). During the visual target detection task, participants were presented with a circle either on the right or left side of the screen and had to decide whether they should indicate (*go*) the target side or refrain from pressing a button (*no-go*). For the *go* conditions, they had to emit a button press indicating the side of the target within 1000 ms. Following the circle, participants obtained one of the following feedbacks (1000 ms): a green up-pointing arrow indicating a win of 30 cent in children and adolescents or 50 cent in the adult groups, a red down-pointing arrow indicating a

loss of 30/50 cents, or a yellow horizontal bar representing neither win nor loss. Feedback was probabilistic, thus, in the win conditions 80% of correct choices and 20% of incorrect choices were rewarded. In the lose conditions, 80% of correct choices and 20% of incorrect choices successfully avoided loss. Participants were informed and instructed about the probabilistic nature of the task beforehand.

The task consisted of 240 trials (60 trials for each of the four conditions, presented in a randomized fashion in four runs) and lasted approximately 35 minutes. Before the task, participants were asked to complete 10 practice trials in which only the target detection circles were presented to familiarize themselves with the appropriate buttons on the computer keyboard and to obtain an overall feel for the speed of the task without exposure to any of the fractal cues used in the main task. The possible win/loss per trial was 0,50 €. Children and adolescents received reimbursement and reward in the form of gift vouchers (5 euros) for a local shopping center on completion of the task. Adults received the exact amount they won on completion of the task whereas for children and adolescents, earnings were rounded to 5 or 10 euros gain. Stimuli were presented and responses recorded using the Cogent 2000 toolbox (http://www.vislab.ucl.ac.uk/cogent.php) running on MATLAB (Version 2009b; Mathworks).

Behavioral data analysis

For Behavioral data analysis SPSS Advanced Statistics v21 (IBM Corporation, Armonk, NY, USA) was used. To test whether a Pavlovian bias was evident in all age groups, mean accuracy rates (%) were used in a four-factorial ANCOVA for repeated measures with action (go vs. no-go), valence (win vs. avoid losing) and time (1st vs. 2nd half) as within-subject factors and age group (children and adolescents vs. young adults vs. midlife adults vs. old adults) as between-subject factor, setting gender as covariate of no interest. Independent samples t-tests were performed to compare performance across age groups using the Holm-Bonferroni correction for six tests.

Reinforcement learning models

We fitted choice behavior to a set of 6 nested reinforcement learning (RL) models incorporating different RL hypothesis. The base model was a Q-learning algorithm (Sutton and Barto, 1998) that used a Rescorla-Wagner update rule to independently

track the action value of each choice given each fractal image ($Q_t(go)$ and $Q_t(nogo)$), with a learning rate (ε) as a free parameter. In the model, the probability of choosing one action on trial t was a sigmoid function of the difference between the action values scaled by a slope parameter that was parameterized as sensitivity to reward. This basic model was initially augmented with an irreducible action noise parameter also known as a lapse rate (ξ) (Talmi *et al.*, 2008) and then further expanded by adding a static bias parameter to the value of the go action (*b*). The model was then augmented by adding a fixed Pavlovian value of 1 to the value of the go action as soon as the first reward was encountered for win cues, and a fixed Pavlovian value of -1 to the value of the go action as soon as the first punishment was encountered for loss cues. This fixed Pavlovian value was weighted by a further free parameter (Pavlovian parameter) into the value of the go action (π). Note that this definition of the Pavlovian value is different from the definition in previous studies that have used this task (Guitart-Masip *et al.*, 2012*b*; Cavanagh *et al.*, 2013; de Boer *et al.*, 2019), as model comparison demonstrated it a better fit than a variable Pavlovian value updated on a trial-by-trial basis (see Table 1). The state (action independent) values for each fractal image were updated on every trial using a Rescorla-Wagner update rule with the same learning rate as the update of the action values. Finally, the model including the static action bias and the Pavlovian bias were augmented by including different sensitivities for reward and punishment. Full equations and a description of all considered models are provided in the Supplemental Information.

Model fitting procedure and comparison

As in previous reports (Huys *et al.*, 2011, Guitart-Masip *et al.*, 2012*b*) we used a hierarchical Type II Bayesian (or random effects) procedure using maximum likelihood to fit simple parameterized distributions for higher-level statistics of the parameters. Since the values of parameters for each subject are 'hidden', this employs the Expectation-Maximization (EM) procedure. For each iteration, the posterior distribution over the group for each parameter is used to specify the prior over the individual parameter fits on the next iteration. All six computational models were fit to the data using a single distribution for all participants. This fitting procedure was, therefore, blind to the existence of different groups with putatively different parameter values. Before inference, all parameters except the action bias were suitably transformed to enforce constraints (log for sensitivity to reward and punishment and Pavlovian

parameter and inverse sigmoid transforms for learning rate and irreducible noise. Six modeling parameters were extracted for each individual, namely *sensitivity to reward, sensitivity to punishment, Pavlovian bias, action bias, learning rate* and *irreducible noise.*

Models were compared using the integrated Bayesian Information Criterion (iBIC) as previously described (Huys *et al.*, 2011, Guitart-Masip *et al.*, 2012*b*). Small iBIC values indicate a model that fits the data better after penalizing for the number of data points associated with each parameter. Comparing iBIC values is akin to a likelihood ratio test (Kass and Raftery, 1995). Note that the iBIC penalizes those versions of the model fit that use four distributions for each parameter. Finally, for each modeling parameter, Kruskal-Wallis tests were conducted to identify age-related differences between groups. Significant interactions were followed up by subsequent group comparisons *using* Wilcoxon rank-sum tests. Holm Bonferroni-correction (p< 0.05 for 6 tests) was used to correct for the effect of multiple comparisons. All tests were performed two-tailed.

REFERENCES

Acosta-Cabronero J, Betts MJ, Cardenas-Blanco A, Yang S, Nestor PJ. In Vivo MRI Mapping of Brain Iron Deposition across the Adult Lifespan. J Neurosci 2016; 36: 364–374.

Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one decade on. Trends Cogn Sci 2014; 18: 177–185.

Bäckman L, Nyberg L, Lindenberger U, Li S-C, Farde L. The correlative triad among aging, dopamine, and cognition: Current status and future prospects. Neurosci Biobehav Rev 2006; 30: 791–807.

Blakemore S-J, Robbins TW. Decision-making in the adolescent brain. Nat Neurosci 2012; 15: 1184–1191.

de Boer L, Axelsson J, Chowdhury R, Riklund K, Dolan RJ, Nyberg L, et al. Dorsal striatal dopamine D1 receptor availability predicts an instrumental bias in action learning. Proc Natl Acad Sci 2019; 116: 261–270.

de Boer L, Axelsson J, Riklund K, Nyberg L, Dayan P, Bäckman L, et al. Attenuation of dopamine-modulated prefrontal value signals underlies probabilistic reward learning deficits in old age. eLife 2017; 6: e26424.

Callaghan MF, Freund P, Draganski B, Anderson E, Cappelletti M, Chowdhury R, et al. Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. Neurobiol Aging 2014; 8: 1862-1872

Casey BJ. Beyond Simple Models of Self-Control to Circuit-Based Accounts of Adolescent Behavior. Annu Rev Psychol 2015; 66: 295–319.

Casey BJ, Getz S, Galvan A. The adolescent brain. Dev Rev 2008; 28: 62–77.

Cavanagh JF, Eisenberg I, Guitart-Masip M, Huys Q, Frank MJ. Frontal Theta Overrides Pavlovian Learning Biases. J Neurosci 2013; 33: 8541–8548.

Chowdhury R, Guitart-Masip M, Lambert C, Dayan P, Huys Q, Düzel E, et al. Dopamine restores reward prediction errors in old age. Nat Neurosci 2013; 16: 648–653.

Chowdhury R, Guitart-Masip M, Lambert C, Dolan RJ, Düzel E. Structural integrity of the substantia nigra and subthalamic nucleus predicts flexibility of instrumental learning in older-age individuals. Neurobiol Aging 2013; 34: 2261–2270.

Cohen AO, Breiner K, Steinberg L, Bonnie RJ, Scott ES, Taylor-Thompson KA, et al. When Is an Adolescent an Adult? Assessing Cognitive Control in Emotional and Nonemotional Contexts. Psychol Sci 2016; 27: 549–562

Davidow JY, Foerde K, Galván A, Shohamy D. An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. Neuron 2016; 92: 93–99.

Dayan P, Balleine BW. Reward, motivation, and reinforcement learning. Neuron 2002; 36: 285–298.

Dayan P, Niv Y, Seymour B, D. Daw N. The misbehavior of value and the discipline of the will. Neural Netw 2006; 19: 1153–1160.

Draganski B, Ashburner J, Hutton C, Kherif F, Frackowiak RSJ, Helms G, et al. Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). NeuroImage 2011; 55: 1423–1434.

Eppinger B, Kray J, Mock B, Mecklinger A. Better or worse than expected? Aging, learning, and the ERN. Neuropsychologia 2008; 46: 521–539.

Galvan A, Hare TA, Parra CE, Penn J, Voss H, Glover G, et al. Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents. J Neurosci 2006; 26: 6885–6892.

Guitart-Masip M, Barnes GR, Horner A, Bauer M, Dolan RJ, Duzel E. Synchronization of Medial Temporal Lobe and Prefrontal Rhythms in Human Decision Making. J Neurosci 2013; 33: 442–451.

Guitart-Masip M, Bunzeck N, Stephan KE, Dolan RJ, Duzel E. Contextual Novelty Changes Reward Representations in the Striatum. J Neurosci 2010; 30: 1721–1726.

Guitart-Masip M, Chowdhury R, Sharot T, Dayan P, Duzel E, Dolan RJ. Action controls dopaminergic enhancement of reward representations. Proc Natl Acad Sci 2012; 109: 7511–7516.

Guitart-Masip M, Economides M, Huys QJM, Frank MJ, Chowdhury R, Duzel E, et al. Differential, but not opponent, effects of l-DOPA and citalopram on action learning with reward and punishment. Psychopharmacology (Berl) 2014; 231: 955–966.

Guitart-Masip M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. Go and no-go learning in reward and punishment: Interactions between affect and effect. NeuroImage 2012; 62: 154–166.

Hämmerer D, Li S-C, Müller V, Lindenberger U. Life Span Differences in Electrophysiological Correlates of Monitoring Gains and Losses during Probabilistic Reinforcement Learning. J Cogn Neurosci 2011; 23: 579–592.

Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, et al. Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. PLoS Comput Biol 2011; 7: e1002028.

Karrer TM, Josef AK, Mata R, Morris ED, Samanez-Larkin GR. Reduced dopamine receptors and transporters but not synthesis capacity in normal aging adults: a meta-analysis. Neurobiol Aging 2017; 57: 36–46.

Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc 1995; 90: 773.

Kray J, Eber J, Lindenberger U. Age differences in executive functioning across the lifespan: The role of verbalization in task preparation. Acta Psychol (Amst) 2004; 115: 143–165.

Lindenberger U. Human cognitive aging: Corriger la fortune? Science 2014; 346: 572–578.

Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl) 2007; 191: 507–520.

Palminteri S, Kilford EJ, Coricelli G, Blakemore S-J. The Computational Development of Reinforcement Learning during Adolescence. PLOS Comput Biol 2016; 12: e1004953.

Palmiter RD. *Dopamine Signaling in the Dorsal Striatum Is Essential for Motivated Behaviors*. Ann N Y Acad Sci 2008; 1129: 35–46.

Richter A, Guitart-Masip M, Barman A, Libeau C, Behnisch G, Czerney S, et al. Valenced action/inhibition learning in humans is modulated by a genetic variant linked to dopamine D2 receptor expression. Front Syst Neurosci 2014; 8: 140

Rutledge RB, Skandali N, Dayan P, Dolan RJ. Dopaminergic Modulation of Decision Making and Subjective Well-Being. J Neurosci 2015; 35: 9811–9822.

Salamone JD, Correa M. The Mysterious Motivational Functions of Mesolimbic Dopamine. Neuron 2012; 76: 470–485.

Samanez-Larkin GR, Gibbs SEB, Khanna K, Nielsen L, Carstensen LL, Knutson B. Anticipation of monetary gain but not loss in healthy older adults. Nat Neurosci 2007; 10: 787–791.

Samanez-Larkin GR, Levens SM, Perry LM, Dougherty RF, Knutson B. Frontostriatal White Matter Integrity Mediates Adult Age Differences in Probabilistic Reward Learning. J Neurosci 2012; 32: 5333–5337.

Schott BH, Niehaus L, Wittmann BC, Schutze H, Seidenbecher CI, Heinze H-J, et al. Ageing and early-stage Parkinson's disease affect separable neural mechanisms of mesolimbic reward processing. Brain 2007; 130: 2412–2424.

Schultz W. Dopamine signals for reward value and risk: basic and recent data. Behav Brain Funct 2010; 6: 24.

Shulman EP, Smith AR, Silva K, Icenogle G, Duell N, Chein J, et al. The dual systems model: Review, reappraisal, and reaffirmation. Dev Cogn Neurosci 2016; 17: 103–117.

Somerville LH, Hare T, Casey BJ. Frontostriatal Maturation Predicts Cognitive Control Failure to Appetitive Cues in Adolescents. J Cogn Neurosci 2011; 23: 2123–2134.

Spear LP. The adolescent brain and age-related behavioral manifestations. Neurosci Biobehav Rev 2000; 24: 417–463.

Steiger TK, Weiskopf N, Bunzeck N. Iron Level and Myelin Content in the Ventral Striatum Predict Memory Performance in the Aging Brain. J Neurosci 2016; 36: 3552–3558.

Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, Mass: MIT Press; 1998

Swart JC, Froböse MI, Cook JL, Geurts DE, Frank MJ, Cools R, et al. Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. eLife 2017; 6: 1–36.

Talmi D, Seymour B, Dayan P, Dolan RJ. Human Pavlovian Instrumental Transfer. J Neurosci 2008; 28: 360–368.

Tymula A, Rosenberg Belmaker LA, Ruderman L, Glimcher PW, Levy I. Like cognitive function, decision making across the life span shows profound age-related changes. Proc Natl Acad Sci 2013; 110: 17143–17148.

Van Leijenhorst L, Moor BG, Op de Macks ZA, Rombouts SARB, Westenberg PM, Crone EA. Adolescent risky decision-making: Neurocognitive development of reward and control regions. NeuroImage 2010; 51: 345–355.

van de Vijver I, Ridderinkhof KR, Harsay H, Reneman L, Cavanagh JF, Buitenweg JIV, et al. Frontostriatal anatomical connections predict age- and difficulty-related differences in reinforcement learning. Neurobiol Aging 2016; 46: 1–12.

Windsor TD, Pearson EL, Butterworth P. Age group differences and longitudinal changes in approach–avoidance sensitivity: Findings from an 8-year longitudinal study. J Res Personal 2012; 46: 646–654.

Yang YC, Boen C, Gerken K, Li T, Schorpp K, Harris KM. Social relationships and physiological determinants of longevity across the human life span. Proc Natl Acad Sci 2016; 113: 578–583.

Zelazo PD, Craik FIM, Booth L. Executive function across the life span. Acta Psychol (Amst) 2004; 115: 167–183.

Ziegler G, Ridgway GR, Blakemore S-J, Ashburner J, Penny W. Multivariate dynamical modelling of structural change during development. NeuroImage 2017; 147: 746–762.
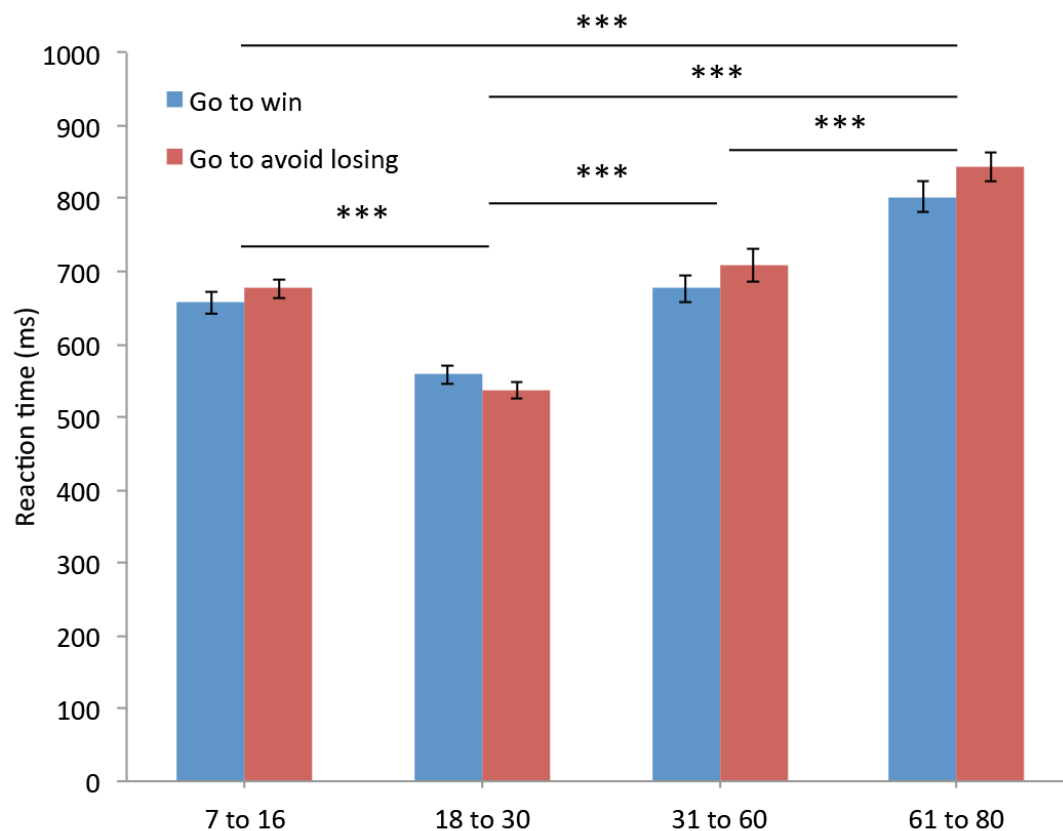
Supporting Information

SI Table 1: Go/no-go performance across age groups

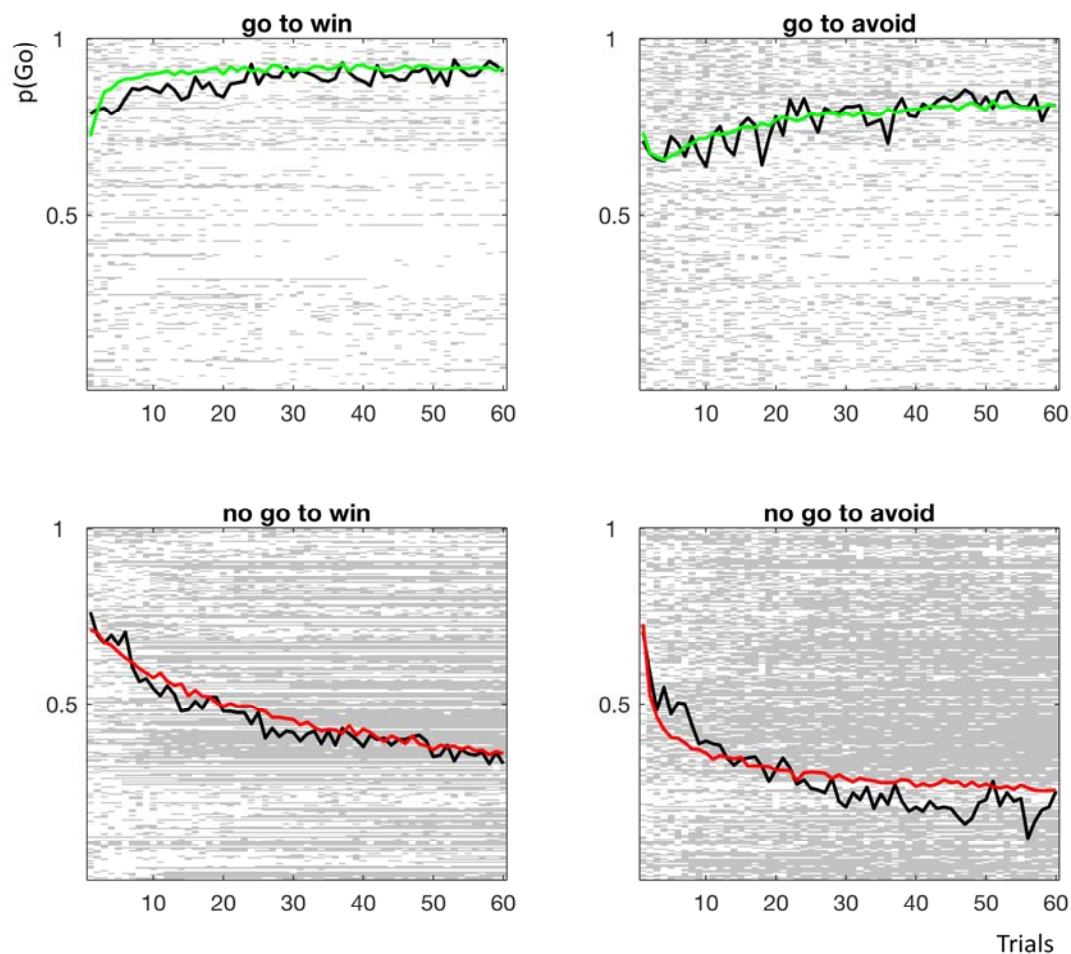|	7 - 16 vs 18 – 30	7 - 16 vs 31 – 60	7 - 16 vs 61 – 80	18 - 30 vs 31 – 60	18 - 30 vs 61 – 80	31 - 60 vs 61 - 80
main effect of age group $F_{3,243}$ = 19.85, p < 0.0001						
overall accuracy	t_{143} = -5.96, p< 0.0001	n.s.	n.s.	t_{133} = 4.13, p<0.0001	t_{119} = 7.36 p<0.0001	t_{100} = 2.86 p=0.008
age group x action $F_{3,243}$ = 13.07; p < 0.0001						
go accuracy	n.s.	n.s.	t_{110} = 4.54, p<0.0001	t_{133} = 3.60, p<0.001	t_{119} = 6.11, p<0.0001	n.s.
nogo accuracy	t_{143} = -6.66, p<0.0001	n.s.	n.s.	t_{133} = 4.02, p<0.001	t_{119} = 6.64, p<0.0001	t_{100} = 2.75, p=0.007
age group x valence $F_{3,243}$ = 2.96, p = 0.033						
Reward	t_{143} = -5.46, p<0.0001	n.s.	n.s.	t_{133} = 3.41, p=0.001	t_{119} = 5.04, p<0.0001	n.s.
Punishment	t_{143} = -4.78, p<0.0001	n.s.	t_{110} = 3.30, p=0.001	t_{133} = 4.36, p<0.0001	t_{119} = 7.97, p<0.0001	t_{100} = 3.81, p<0.001
age group x time $F_{3,243}$ = 6.01, p = 0.001						
overall learning gain	n.s.	n.s.	n.s.	n.s.	t_{119} = 4.69, p<0.0001	n.s.
age group x action x time $F_{3,243}$ = 2.69; p = 0.047						
go learning gain	t_{143} = -3.29, p=0.001	n.s.	n.s.	n.s.	t_{119} = 3.07, p=0.003	n.s.
nogo learning gain	n.s.	n.s.	t_{110} = 3.0, p=0.003	n.s.	t_{119} = 3.59, p<0.001	t_{100} = 2.94, p=0.004
go 1st half	n.s.	n.s.	t_{110} = 4.17, p<0.0001	n.s.	t_{119} = 4.24, p<0.0001	n.s.
go 2nd half	t_{143} = -3.57, p<0.001	n.s.	t_{110} = 3.93, p<0.001	t_{133} = 3.92, p=0.0001	t_{119} = 6.69, p<0.0001	n.s.
nogo 1st half	t_{143} = -6.95, p<0.0001	t_{124} = -2.90, p=0.004	n.s.	t_{133} = 3.56, p=0.001	t_{119} = 5.33, p<0.0001	n.s.
nogo 2nd half	t_{143} = -5.68, p<0.0001	n.s.	n.s.	t_{133} = 3.33, p=0.001	t_{119} = 6.83, p<0.0001	t_{100} = 3.27, p=0.001

Results from independent samples t-tests. Only values from significant results are displayed. Holm-Bonferroni correction was applied for six tests per variable.

SI Figure 1: Reaction times for Go responses across the lifespan

Mean (± S.E.M) reaction times (ms) plotted for *go to win* and *go to avoid losing* responses for all age groups. Whilst participants were instructed that the accuracy of their response was more important than speed, a significant difference in *go* response times was observed between groups, whereby young and older adults demonstrated the fastest and slowest responses respectively. ***(p<0.0001) indicates significant differences between groups (1-way ANOVA with Bonferroni *post-hoc* tests).

SI Figure 2: Observed and modeled learning across the lifespan.

Modeling parameters from the winning model were used to generate simulated choice data. The simulated group mean probability of performing a go response on each trial is plotted in colored lines (green for go conditions, where go is the correct response; red for no-go conditions, where no-go is the correct response). The mean for all participants' actual performance is plotted in black lines, reflecting the proportion of actual go responses on each trial. In the plot area, each row represents choice behavior for each participant (n=247) corresponding to a total of 247 pixels per trial. A white pixel illustrates that a participant chose a go response on that trial whilst a gray pixel represents a no-go response.

SI Materials and Methods

Assessment of cognitive function and medical history

Participants from the age of 7 to 16 were recruited from a pool of typically developing children and adolescents at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, or through advertisements in a local newspaper. Participants and their parents were interviewed with the German adaptation (Delmo et al., 2000) of the Revised Schedule for Affective Disorders and Schizophrenia for School-Age Children: Present and Lifetime Version (K-SADS-PL – DSM-IV; Kaufman et al., 1997). Diagnostic tests were selected according to age: Intelligence was assessed using the German adaptation of the *Culture Fair Intelligence Test – Scale 20* (age ≥ 9 years; CFT-20-R; Weiss, 1997) or *Scale 1* (age < 9 years; CFT 1-R, Weiss & Osterland, 2013). *The d2 – Attention Endurance Test* (age ≥ 9 years; Brickenkamp, 2002) or the *bp-Test* (subtest from a developmental test battery for elementary school children; *Basisdiagnostik umschriebener Entwicklungsstörungen im Grundschulalter*; age < 9 years; Esser et al., 2008) were used to measure attentional performance. The Youth-Self-Report (YSR, age > 10 years) and the Child-Behavior-Checklist (CBCL; Achenbach, 1991a,b) were included as further clinical measures. None of the participants received a clinical diagnosis on participation or reported any history of neurological disorders.

All young, midlife and older adults were recruited at the Leibniz Institute for Neurobiology and Institute of Cognitive Neurology and Dementia Research in the University of Magdeburg. All older adults were screened to ensure intact global cognitive performance using a brief neuropsychological battery comprising mini-mental state examination, *Stroop test* in German language and Logical memory test parts I, and II, from the Wechsler memory scale. Individuals with depression were excluded using the Becks Depression Inventory II. Older adults' alertness and divided attention was assessed using the *Test of Attentional Performance* (TAP). Any individuals known to have had neurological or major psychiatric illness, myocardial infarction, significant cardiovascular history or diabetes mellitus were not eligible for participation.

Computational modeling of the behavioral data

As in previous experiments where this task has been used (Cavanagh et al., 2013; Guitart-Masip et al., 2012, 2014), we fit six nested models to the observed behavioral data in order to test different instrumental and Pavlovian reinforcement-learning hypothesis. In all models, expected values $Q(a, s)$ on each trial t where calculated for each action $a \in \{0,1\}$ on each state $s \in \{1,2,3,4\}$, where the actions can be go and no go and states are the four experimental conditions of our task. $Q(a, s)$ was based on a simple Rescorla-Wagner or delta rule update equation that was implemented each time and outcome was observed:

$$Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \varepsilon\left(\rho r_t - Q_{t-1}(a_t, s_t)\right) \qquad (1)$$

where ε is the learning rate. Reinforcements entered the equation through $r_t \in \{-1, 0, 1\}$ and ρ is a free parameter that determined the effective size of reinforcements. In four models (RW, RW+noise, RW+noise+bias and RW+noise+bias+Pav) there was only one value of ρ per subject, meaning that loss of a reward was equally as aversive as obtaining a punishment. The two remaining models (RW(rew/pun)+noise+bias, RW(rew/pun)+noise+bias+Pav) included different values of the parameter ρ for reward and punishment trials.

A squashed softmax function (Sutton and Barto, 1998) was used to calculate the probability of selection each action on a given state:

$$p(a_t|s_t) = \left[\frac{exp\left(w_t(a_t, s_t)\right)}{\sum_{a'} exp\left(w_t(a_t, s_t)\right)}\right](1 - \xi) + \frac{\xi}{2} \qquad (2)$$

where $w(a_t, s_t)$ reflects the propensity of selecting action a in state s, and ξ is the irreducible noise, which was kept at 0 for one of the models (RW), but was free to vary between 0 and 1 for all other models. Different models varied in the way $w(a_t, s_t)$ was constructed. In the simplest models (RW and RW+noise), $w_t(a_t, s_t) = Q_t(a_t, s_t)$.

Further models added extra factors to the action propensities. For models that contained a bias parameter, the action weight was modified to include a static bias parameter b:

$$w_t(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b & \text{if } a = \text{go} \\ Q_t(a_t, s_t) & \text{else} \end{cases} \quad (3)$$

For the model including a Pavlovian factor (RW+noise+bias+Pav and RW(rew/pun)+noise+bias+Pav), the action weight consisted of three components:

$$w_t(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b + \pi V_t(s_t) & \text{if } a = \text{go} \\ Q_t(a_t, s_t) & \text{else} \end{cases} \quad (4)$$

where π was again a free parameter. The Pavlovian value V_t was determined by the first experienced reinforcing feedback on state s. For "win" states, this Pavlovian value was set to 1 after the first trial on which a win outcome was experienced, and for "loss" states, this Pavlovian values was set to -1 after the first trial on which a loss outcome was experienced. Thus, for the `avoid loss' conditions, in which the $V(s)$ would be non-positive, the Pavlovian parameter inhibited the go tendency in proportion to the negative value $V(s)$ of the stimulus, while it similarly promoted the tendency to go in conditions in the 'win' conditions.

REFERENCES

Achenbach TM (1991a): Manual of the Youth Self-Report and 1991 Profile. Burlington: Department of Psychiatry, University of Vermont.

Achenbach TM (1991b): Manual for the Child Behavior Checklist/ 4–18 and 1991 Profile. Burlington: Department of Psychiatry, University of Vermont.

Brickenkamp R (2002): Test d2 – Aufmerksamkeits-Belastungs-Test. Goettingen: Hogrefe.

Cavanagh, J.F., Eisenberg, I., Guitart-Masip, M., Huys, Q., and Frank, M.J. (2013). Frontal Theta Overrides Pavlovian Learning Biases. J. Neurosci. *33*, 8541–8548.

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dolan, R.J., and Düzel, E. (2013). Structural integrity of the substantia nigra and subthalamic nucleus predicts flexibility of instrumental learning in older-age individuals. Neurobiol Aging. 2013 Oct;34(10):2261-70.

Delmo C, Weiffenbach O, Gabriel M, Bölte S, Marchio E, Poustka F (2000) Fragebogen für Affektive Störungen und Schizophrenie für Kinder im Schulalter (6-18 Jahre). Frankfurt: Klinik für Psychiatrie und Psychotherapie des Kindes- und Jugendalters.

Esser G, Wyschkon A, Ballaschk K (2008) BUEGA - Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter. Goettingen: Hogrefe.

Guitart-Masip, M., Huys, Q.J.M., Fuentemilla, L., Dayan, P., Duzel, E., and Dolan, R.J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. NeuroImage *62*, 154–166.

Guitart-Masip, M., Economides, M., Huys, Q.J.M., Frank, M.J., Chowdhury, R., Duzel, E., Dayan, P., and Dolan, R.J. (2014). Differential, but not opponent, effects of l-DOPA and citalopram on action learning with reward and punishment. Psychopharmacology (Berl.) *231*, 955–966.

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N (1997): Schedule for affective disorders and schizophrenia for school-age children - Present and lifetime version (K-SADS-PL): Initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 36:980–988.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: an introduction (Cambridge, Mass: MIT Press).

Weiss RH, Osterland J (2013) CFT 1-R. Goettingen: Hogrefe.

Weiss RH (2008): Grundintelligenztest Skala 2 – Revision (CFT 20-R). Goettingen: Hogrefe.