Aus dem
CharitéCentrum für Neurologie, Neurochirurgie und Psychiatrie (CC15)
Klinik für Psychiatrie und Psychotherapie, Charité Campus Mitte
Direktor: Prof. Dr. med. Dr. phil. Andreas Heinz

# Habilitationsschrift
# Veränderungen von Lernen und Entscheidungsfindung bei Abhängigkeit und Psychose

Zur Erlangung der Lehrbefähigung
für das Fach Experimentelle Psychiatrie

vorgelegt dem Fakultätsrat der Medizinischen Fakultät
Charité-Universitätsmedizin Berlin

von
**Dr. med. Heiner Stuke**

**Eingereicht: 10/2023**
**Dekan: Prof. Dr. Joachim Spranger**

## Inhaltsverzeichnis

## Abkürzungsverzeichnis

| | |
|---|---|
| AUDIT | alcohol use disorders identification test |
| CAPS | Cardiff Anomalous Perceptions Scale |
| CSF | continuous flash suppression |
| DALY | disability-adjusted life years |
| DLPFC | dorsolateral prefrontal cortex |
| EEG | Elektroenzephalographie |
| fMRT | funktionelle Magnetresonanztomographie |
| ICD | international classification of diseases |
| irisa | syndrome of impaired response inhibition and salience attribution |
| MPFC | medial prefrontal cortex |
| MRT | Magnetresonanztomographie |
| PDI | Peters Delusion Inventory |
| PET | Positronen-Emissions-Tomographie |
| ROI | regions of interest |
| VTA | ventral tegmental area |

# 1. Einleitung

## *1.1 Überblick*

Psychische Erkrankungen sind häufig und haben oft schwerwiegende Konsequenzen für Betroffene, Angehörige und die Gesellschaft als Ganzes: Neue Berechnungen gehen davon aus, dass mehr als ein Achtel der durch Krankheit verlorenen nach Lebensqualität adjustierten Lebensjahre (disability-adjusted life years, DALY) auf das Konto psychischer Erkrankungen gehen (Vigo, Thornicroft, & Atun, 2016). Zwei der in diesem Zusammenhang bedeutendsten Erkrankungen sind Abhängigkeiten und psychotische Erkrankungen. Abhängigkeiten werden zusammen genommen für 4 % der DALY weltweit verantwortlich gemacht (Rehm, Taylor, & Room, 2006), die Schizophrenie als schwerwiegendste psychotische Erkrankung ist diesen Berechnungen zufolge alleine für 1,1 % der DALY verantwortlich (Theodoridou & Rössler, 2010). Jahrzehnte psychologischer und neurowissenschaftlicher Forschung haben unser Wissen um die diesen Erkrankungen zugrundeliegenden Mechanismen auf verschiedenen Ebenen erweitert. Das Resultat davon sind zunehmend elaborierte Theorien darüber, wie die Verarbeitung von Informationen, Lernen und Entscheidungsfindung sich verändern, wenn Patient*innen eine Abhängigkeitserkrankung oder eine psychotische Erkrankung entwickeln. Die vorliegenden Veränderungen werden hierbei, inspiriert durch Fortschritte in den Computerwissenschaften und einem verstärkten interdisziplinären Austausch, zunehmend in einer der Informatik entlehnten Terminologie gefasst, was eine Quantifizierung der Defizite erleichtert. Durch Fortschritte in der Hirnbildgebung mit Methoden wie der funktionellen Magnetresonanztomographie (fMRT) oder der Positronen-Emissions-Tomographie (PET), sowie in Tiermodellen und genetischen Assoziationsstudien können diese kognitiven Prozesse außerdem besser mit neurobiologischen Prozessen in Verbindung gebracht werden. Es gibt allerdings, wie in dieser Schrift herausgearbeitet werden soll, ungeklärte Fragen, die einer Translation der grundlagenwissenschaftlichen Erkenntnisse in die klinische Psychiatrie bislang im Wege stehen.

Die vorliegende Habilitationsschrift skizziert im ersten Abschnitt zunächst grundsätzlich pathologische Veränderungen von Lernen und Entscheidungsfindung bei Abhängigkeit und psychotischen Erkrankungen sowie die Rolle des Dopaminsystems darin. Die beiden

Erkrankungen werden anschließend einzeln dargestellt und die den eigenen Arbeiten zugrundeliegenden Theorien zu Veränderungen in Lernen und Entscheidungsfindung und den entsprechenden neurobiologischen Korrelaten bei diesen Erkrankungen zusammengefasst. Schließlich werden damit verbundene offene Fragestellungen, die die Motivation für die eigenen Arbeiten darstellten, abgeleitet. In sechs eigenen Arbeiten wurden darauf aufbauend spezifische Hypothesen dieser Theorien empirisch überprüft, teils mittels fMRT, teils mit statistischer Modellierung von Verhaltensdaten. Diese eigenen Arbeiten werden im zweiten Abschnitt dargestellt. Der dritte Abschnitt diskutiert die Ergebnisse der eigenen Arbeiten kritisch im Kontext der motivierenden theoretischen Modelle und entsprechender Vorstudien und skizziert Perspektiven für die Translation in die klinische Praxis und für weitere Forschungsschritte.

## *1.2 Lernen, Entscheidungsfindung und Dopamin: Ein neurowissenschaftlicher Rahmen zum Verständnis psychischer Erkrankungen wie Abhängigkeit und Psychose*

Aus Erfahrungen zu lernen und Entscheidungen basierend auf dem gelernten Wissen zu treffen, zählt zu den grundsätzlichen Leistungen des Gehirns. Veränderungen von Lernen und Entscheidungsfindung werden dementsprechend als zentral für die Entstehung psychischer Erkrankungen wie Schizophrenie oder Abhängigkeit betrachtet. Wie im folgenden Abschnitt dargestellt, wird hierbei bei Abhängigkeitserkrankungen grundsätzlich eine gesteigerte Reaktion auf und gesteigertes Lernen aus substanzbezogenen Belohnungen postuliert, begleitet durch eine reduzierte Reaktion auf und reduziertes Lernen aus nicht-substanzbezogenen Belohnungen, sowie durch defizitäre kognitive Kontrollprozesse (Abschnitt 1.2). Bei psychotischen Erkrankungen wird, wie im übernächsten Abschnitt ausgeführt wird, davon ausgegangen, dass die Integration von durch Lernprozesse gewonnenen Vorannahmen und neuen Informationen gestört ist, was fundamentale Veränderungen der Informationsverarbeitung und darauf basierender Entscheidungsfindung mit sich bringt (Abschnitt 1.3).

Dopamin ist ein entscheidender Neurotransmitter beim Lernen aus neuen Informationen. Auch die Veränderungen von Lernen und Informationsverarbeitung bei Abhängigkeit und Psychose

werden maßgeblich auf Veränderungen in dopaminergen Prozessen zurückgeführt und die Dopaminhypothesen von Abhängigkeit und Psychose zählen dementsprechend zu den einflussreichsten Theorien zur Erklärung dieser Krankheitsbilder in der biologischen Psychiatrie (Heinz, 2002; Howes & Nour, 2016; Keiflin & Janak, 2015; Maia & Frank, 2017). Zum Verständnis der krankheitsspezifischen Veränderungen im Dopaminsystem soll zunächst die generelle Neuroanatomie rekapituliert werden, bevor die entsprechenden pathologischen Veränderungen bei Abhängigkeit und Psychose im einzelnen betrachtet werden.

Bisher sind fünf verschiedene Subtypen von Dopaminrezeptoren beschrieben worden (D1 – D5 Rezeptoren), die alle metabotrop sind und zur Bildung von Botenstoffen führen, die sekundäre Zellsignalwege aktivieren oder blockieren (Baik, 2013). Dopaminrezeptoren, vor allem die D1 und D2 Subtypen, sind im Gehirn weit verbreitet und entsprechend umfassend ist die Rolle des Dopaminsystems in psychomotorischen, emotionalen und kognitiven Prozessen (Klein et al., 2019) sowie bei psychischen Erkrankungen. Neuroanatomisch wird das Dopaminsystem in das nigrostriatale und das mesolimbische System unterteilt. Ersteres umfasst Dopamin-Neuronen in der Substantia nigra, die in das dorsale Striatum projizieren und spielt vor allem eine Rolle bei der Koordination von Bewegungen (Dysfunktionen oder Verlust von Neuronen in diesem System führen entsprechend zu parkinsinoiden Bewegungsstörungen). Bei psychischen Erkrankungen von größerer Bedeutung ist das mesolimbische System, in dem dopaminerge Mittelhirnneuronen aus dem ventralen tegmentalen Areal (VTA) über die mesokortikale Bahn in den präfrontalen Kortex und über die mesolimbische Bahn in das ventrale Striatum (nucleus accumbens) projizieren.

Es ist seit langem bekannt, dass Dopamin eine wichtige Rolle bei Verstärkungslernen und damit bei der Motivation von Handlungen spielt: Dies betrifft sowohl die Kopplung von Reiz und Belohnung in operanten Konditionierungsparadigmen als auch die Motivation für die Suche nach Belohnungen (Collins & Saunders, 2020). Dieser Funktion entsprechend werden die meisten dopaminergen Neuronen durch unerwartete primäre Belohnungen wie Nahrung stark unmittelbar aktiviert. Die bahnbrechenden Studien von Wolfram Schultz und anderen zeigten jedoch, dass dopaminerge Neuronen nicht durch Belohnungen an sich aktiviert werden, sondern durch einen "reward prediction error", der der Differenz zwischen der erwarteten und der erhaltenen Belohnung entspricht. Wenn also eine Belohnung größer als erwartet ausfällt (positiver reward prediction error), werden die dopaminergen Neurone stark aktiviert, während

sie bei erwarteten Belohnungen kaum aktiviert und bei hinter den Erwartungen zurückbleibenden Belohnungen sogar gehemmt werden können (Schultz, Dayan, & Montague, 1997). Inzwischen ist ein enger Zusammenhang zwischen dem Aktivierungsmuster dopaminerger Neuronen und bestimmten Charakteristika von Belohnungen auf mehreren Ebenen gezeigt worden, unter anderem eine Korrelation mit der Wahrscheinlichkeit, Größe und zeitlichen Latenz der Belohnung (Bromberg-Martin, Matsumoto, & Hikosaka, 2010). Während der Zusammenhang zwischen dopaminerger Aktivität und reward prediction errors inzwischen experimentell gut untermauert ist, weisen neuere Studien auf eine Reaktivität von Subpopulationen dopaminerger Neuronen auf Reize hin, die nicht unmittelbar belohnend, aber auf andere Arten relevant sind (und damit eine adaptive Neuroplastizität erfordern). Dies umfasst sowohl aversive Reize, als auch solche, die auf einem frühen sensorischen Prozessierungsstadium als „alarmierend“ eingestuft werden, zum Beispiel, weil sie ungewöhnlich, überraschend oder intensiv sind (Bromberg-Martin et al., 2010; Lammel, Lim, & Malenka, 2014). Um diese vielfältigen Funktionen auszuüben, moduliert Dopamin auf verschiedene Arten die Aktivierung anderer Transmittersysteme, sowohl im Sinne einer dauerhaft veränderten synaptischen Verbindung, als auch im Sinne einer kurzfristigen belohnungsassoziierten Steigerung der Aktivität (Reynolds & Wickens, 2002; Speranza, di Porzio, Viggiano, de Donato, & Volpicelli, 2021).

Zusammenfassend und vereinfachend lässt sich sagen, dass, obwohl innerhalb des mesolimbischen Systems eine Binnendifferenzierung verschiedener Populationen dopaminerger Neurone mit spezifischen Aktivierungsmustern existiert, seine Neurone grundsätzlich durch belohnende, relevante, überraschende und alarmierende Reize aktiviert werden und die Aktivierung anderer Systeme im Sinne eines Lern- und Aufmerksamkeitssignals modulieren (Bromberg-Martin et al., 2010; Schultz, 2007). Krankheitsspezifische Veränderungen in diesem System können dementsprechend dazu führen, dass Reize als relevant verarbeitet werden, die im gesunden Zustand nicht als relevant erscheinen. Dies kann, wie in den folgenden beiden Abschnitten detaillierter ausgeführt wird, bei der Abhängigkeit eine „Hyperrelevanz“ drogenbezogener Reize umfassen und in der Psychose eine übersteigerte Interpretation eigentlich nicht relevanter Reize, die die Entstehung von Halluzinationen und Wahn begünstigt.

### *1.3 Veränderungen von Lern- und Entscheidungsprozessen bei Abhängigkeit*

Die Abhängigkeit wird gemäß International Classification of Diseases (ICD-10, World Health Organisation, 2016) anhand von sechs Kriterien definiert:

- starker Wunsch und/oder Zwang, die Substanz zu konsumieren
- verminderte Kontrollfähigkeit bezüglich des Beginns, der Menge und/oder der Beendigung der Einnahme
- körperliche Entzugssymptome
- Toleranzentwicklung (Wirkverlust) bzw. Dosissteigerung
- erhöhter Zeitaufwand, um die Substanz zu beschaffen oder sich von den Folgen des Konsums zu erholen, verbunden mit der Vernachlässigung anderer Interessen
- fortgesetzter Konsum trotz Folgeschäden

Aus diesen Kriterien wird deutlich, dass bei Patient*innen mit Abhängigkeit sowohl ein massives Suchtverlangen (craving) vorliegt, als auch eine reduzierte Kontrollfähigkeit bezüglich des Substanzkonsums. Hierauf aufbauend wurde die vor allem in der Bildgebungsforschung einflussreiche dual system Theorie der Abhängigkeit entwickelt. Diese erklärt für Abhängigkeit charakteristische Verhaltensweisen wie Suchtverlangen, Kontrollverlust und fortgesetztem Konsum trotz Folgeschäden durch ein Ungleichgewicht zwischen automatisch-impulsiven und bewusst-reflektiven Modi der Verhaltenssteuerung (McClure & Bickel, 2014). Die den fortgesetzten Substanzkonsum bedingenden Veränderungen werden demzufolge in zwei unterschiedlichen „Systemen“ lokalisiert, von denen eines mit der automatischen Entstehung von Verlangen in Verbindung gebracht wird (Belohnungsverarbeitung) und das andere mit einer reflektiven Regulierung von Verlangen z.B. mit Blick auf längerfristige Ziele wie Abstinenz oder Erhalt der Gesundheit (kognitive Kontrolle).

Die 1993 erstmals postulierte incentive salience Theorie konkretisiert die Veränderungen der Belohnungsverarbeitung (d.h., im automatisch-impulsiven System) als eine Kernpathologie bei Abhängigkeitserkrankungen (Robinson & Berridge, 1993, 2001). Dieser Theorie zufolge entwickelt sich eine Abhängigkeit als Folge von Neuroadaptationen, die durch wiederholten Drogenkonsum induziert werden. Es wird vermutet, dass das mesolimbische Dopamin-System, das an der Verarbeitung von Belohnungsreizen beteiligt ist (s. Abschnitt 1.2), allmählich für

substanzbezogene Stimuli sensibilisiert wird (d.h., stärker auf diese reagiert) und für alternative Belohnungsreize desensibilisiert wird (d.h., schwächer auf diese reagiert). Dies manifestiert sich subjektiv als massives Verlangen gegenüber substanzbezogenen Reizen und einer relativen Anhedonie anderen Belohnungen gegenüber. Gehirnstrukturen, die am limbischen Belohnungssystem beteiligt sind, umfassen die Amygdala, die ventral tegmental area (VTA), den Hippocampus, den Inselcortex, das ventrale Pallidum, das ventrale Striatum / Nucleus accumbens, das anteriore Cingulum und den medialen präfrontalen Cortex (MPFC) (Koob & Le Moal, 2001; Noori, Cosa Linan, & Spanagel, 2016; Tang, Fellows, Small, & Dagher, 2012). Spezifische Meta-Analysen von fMRT-Studien bestätigen eine verstärkte Aktivierung dieser Areale durch Bilder substanzbezogene Reize (Lin et al., 2020; Schacht, Anton, & Myrick, 2013) sowie eine reduzierte Aktivierung durch nicht-substanzbezogene Belohnungsreize (Hommer, Bjork, & Gilman, 2011; Lin et al., 2020) bei den im Rahmen dieser Habilitation untersuchten Abhängigkeitserkrankungen, der Alkoholabhängigkeit und der Nikotinabhängigkeit.

Neben den oben beschriebenen Veränderungen in Belohnungsprozessen wird in dual system Theorien ein zweites verändertes System postuliert, das als bewusst-reflektives System an der Kontrolle und Suppression automatisch ausgelöster Belohnungsprozesse unter Einbeziehung langfristiger Ziele und Vorsätze (wie z.B. Abstinenz oder Konsumreduktion) beteiligt ist (Baler & Volkow, 2006; McClure & Bickel, 2014). Wiederholter Konsum addiktiver Substanzen führt in diesem System zu einer reduzierten Inhibitionsfähigkeit, vor allem substanzbezogenen Belohnungen gegenüber, die zu dem abhängigkeitsdefinierenden Kontrollverlust beiträgt. Diesem „syndrome of impaired response inhibition and salience attribution“ (irisa) korrespondieren demzufolge auf neurofunktioneller Ebene vor allem Veränderungen im Präfrontalkortex. Der laterale Teil des Präfrontalkortex ist von kritischer Bedeutung für komplexe Prozesse der Selbstkontrolle, der Inhibition von automatischen und impulsiven Verhaltensweisen und der Verhaltensplanung mit Blick auf langfristige Ziele (Tanji & Hoshi, 2008), während im medialen Präfrontalkortex verschiedene Zielstellungen über unterschiedliche Zeitskalen integriert und für Entscheidungsfindung kombiniert werden. Passend zu dieser Theorie zeigen Bildungebungsstudien eine reduzierte Funktion dieser inihibitorischer Prozesse mit Beteiligung des lateralen Präfrontalkortex und eine Zunahme von Belohnungs- und Bedeutungssignalen im medialen Präfrontalkortex im Zusammenhang mit Suchtverlangen und Intoxikation (Goldstein & Volkow, 2011).

Der bisherige Forschungsstand zu Lern- und Entscheidungsprozessen bei Abhängigkeit lässt allerdings einige zentrale Fragen ungeklärt: So wurden Veränderungen in Entscheidungsprozessen und kognitiver Kontrolle bislang vor allem bei suchtunabhängigen Entscheidungsaufgaben untersucht (wie monetary delay discounting tasks, bei denen Entscheidungen zwischen einem unmittelbar auszahlbaren Geldbetrag und einem später auszahlbaren, aber etwas größeren Geldbetrag getroffen werden). Der Zusammenhang zwischen diesen unspezifischeren Maßen von Entscheidungsfindung und tatsächlichen abhängigkeitstypischen Entscheidungen für Substanzkonsum ist unklar und die Frage, welche Rolle Veränderungen im Belohnungs- und Kontrollsystem bei tatsächlichen Entscheidungen für Substanzkonsum spielen, wird in der unter 2.1 dargestellten Studie untersucht. Darüber hinaus wurden in den bisherigen Studien zu suchtassoziierten Veränderungen im Belohnungssystem suchtspezifische und alternative Belohnungsreize lediglich mit neutralen Reizen und nicht direkt miteinander verglichen, was eine direkte Überprüfung der incentive salience Theorie (Sensitivierung gegenüber suchtbezogenen Reizen und Desensitivierung gegenüber alternativen Belohnungsreizen) ermöglichen würde. Dieser direkte Vergleich wurde in der unter 2.2 dargestellten Studie geleistet. Schließlich wurde der Einfluss von langfristigen Zielstellungen (wie z.B. der Vermeidung von suchtassoziierten körperlichen Schäden) auf das Konsumverlangen und seine neurofunktionellen Marker bisher nicht direkt untersucht. In der unter 2.3 dargestellten Studie wurde diese Lücke geschlossen, indem der Einfluss von aversiven Reizen, die die körperlichen Folgeschäden von Nikotinkonsum zeigen, auf die Aktivierung von Belohnungs- und Kontrollsystem während der nachfolgenden Präsentation von suchtbezogenen Reizen untersucht wurde. Diese Studie adressiert damit die Frage nach den Wirkmechanismen von „Schockbildern“ von Rauchfolgen, d.h., ob diese zu einer Aktivierung von Kontrollprozessen und / oder zu einer Reduktion von Suchtverlangen und assoziierten Aktivierungen im Belohnungssystem führen. Eine Aufklärung der Wirkmechanismen könnte von Bedeutung sein für den optimierten Einsatz solcher Stimuli in public health Programmen zur Rauchprävention (Hammond, 2011) und in der individuellen Entwöhnungsbehandlung (Pang et al., 2021).

### *1.4 Veränderungen von Lern- und Entscheidungsprozessen bei Psychosen*

Psychose ist ein Kernsyndrom schwerwiegender psychiatrischer Erkrankungen wie der Schizophrenie. Sie umfasst verschiedene Symptome, unter anderem unbegründete

Überzeugungen (Wahn) und Wahrnehmungen ohne ursächlichen Reiz (Halluzinationen). Sowohl in kognitiven als auch in neurobiologischen Theorien ist es eine zentrale Herausforderung, die Vielgestaltigkeit psychotischer Erfahrungen auf wenige (oder auch nur eine) „Kernpathologien“ zurückführen können (Heinz et al., 2019). Einflussreiche, durch Konzepte der Bayes-Statistik inspirierte Theorien erklären psychotische Symptome durch ein Ungleichgewicht in der Gewichtung von Vorannahmen und aktuellen Informationen (im folgenden als Bayesianische Psychosemodelle bezeichnet, (Corlett et al., 2019; Fletcher & Frith, 2009; Sterzer et al., 2018)). Diesen Theorien zufolge besteht eine zentrale Herausforderung für das Gehirn darin, aus eingehenden sensorischen Informationen von relativ schlechter Qualität Rückschlüsse auf den Zustand der Außenwelt zu ziehen. Bewusste Wahrnehmung wird demzufolge nicht in erster Linie als sensorische Abbildung interpretiert, sondern als Resultat eines Konstruktionsprozesses, der auf internen Modellen beruht, die frühere Erfahrungen (Vorannahmen) mit akutellen sensorischen Informationen integrieren (Aggelopoulos, 2015). Es wird hypothetisiert, dass das Gehirn hierbei auf probabilistische Vorannahmen zurückgreift (dem Konzept der prior probability in der Bayes-Statistik entsprechend (Barlow, 1990; Zeki & Chen, 2020)). Die Nutzung von Vorannahmen bei der Interpretation uneindeutiger sensorischer Signale wurde in den Neurowissenschaften vielfach belegt (de Lange, Heilbron, & Kok, 2018). Insbesondere werden solche Interpretationen „bevorzugt“, die entweder häufig (Series & Seitz, 2013) oder emotional bedeutsam (Mather & Sutherland, 2011) sind. So kann beispielsweise eine verrauschte und damit uneindeutige akustische Information je nach Vorannahme entweder als Rauschen oder als Stimme interpretiert werden. Eine Veränderung in diesen Vorannahmen kann entsprechend zur Entstehung von Halluzinationen beitragen (Corlett et al., 2019).

Es wird außerdem (ebenfalls in Übereinstimmung zur Bayes-Statistik) angenommen, dass Vorannahmen kontinuierlich durch Vorhersagefehler aktualisiert werden. Dies bedeutet, dass kontinuierliches Lernen als Reaktion auf überraschende (nicht zur Vorannahme passende) Informationen die flexible Anpassung von Vorannahmen ermöglicht. In diesem Sinne könnten psychotische Symptome als Resultat von maladaptivem Lernen bezeichnet werden, das auftritt, wenn irrelevante Informationen aufgrund veränderter Vorhersagefehlersignale als überraschend und relevant angesehen werden und damit zu einer Korrektur der Vorannahmen führen. Ein klassisches Experiment, um übersteigertes Lernen aus neuen Informationen zu

objektivieren, ist beispielsweise der sogenannte beads task (Phillips & Edwards, 1966): Hierbei sehen die Proband*innen zwei Urnen, die zwei verschiedene Arten von Perlen in unterschiedlichem Verhältnis beinhalten (zum Beispiel könnte eine Urne 85 % blaue Perlen und 15 % rote Perlen enthalten, während die andere umgekehrt 15 % blaue Perlen und 85 % rote Perlen enthält). Die Teilnehmer*innen haben die Aufgabe, zu schätzen, aus welcher der Urnen eine Reihe von Perlen gezogen wird. Sie können diese Entscheidung basierend auf einer beliebigen Zahl von gezogenen Perlen treffen. Bei Patient*innen mit Psychose wurde eine rasche Entscheidung auf Basis weniger Informationen (einer oder zwei Perlen) festgestellt („jumping to conclusions“, (Dudley, Taylor, Wickham, & Hutton, 2016)). Dies ist ein Beispiel für ein pathologisch verändertes Lernen durch eine erhöhte Zuschreibung von Relevanz zu Informationen bei Patient*innen mit Psychose (aberrant salience, (Howes & Nour, 2016; Kapur, 2003)).

Auch bei diesen auf der Verhaltensebene feststellbaren Veränderungen wird auf neurobiologischer Ebene eine Beteiligung des Dopamin-Systems angenommen. Für eine generelle Beteiligung des Dopaminsystems an der Entstehung psychotischer Symptome sprechen Befunde aus verschiedenen grundlagenwissenschaftlichen und klinischen Forschungslinien: Es ist beispielsweise lange bekannt, dass psychotrope Substanzen, die die striatale Dopaminfreisetzung erhöhen (wie beispielsweise Amphetamine), Psychosen auslösen können (Howes & Kapur, 2009). Umgekehrt ist die Wirksamkeit eines (typischen) Antipsychotikums proportional zu seiner Fähigkeit, D2/3-Rezeptoren zu antagonisieren (Howes & Kapur, 2009; Seeman, Lee, Chau-Wong, & Wong, 1976). Passend dazu zeigen PET-Studien eine erhöhte Dopaminsynthese und -freisetzung im Striatum und in der VTA bei Patient*innen mit Schizophrenie im Vergleich zu Kontrollproband*innen (Howes & Nour, 2016).

Zusammengefasst zeigen Patient*innen mit Schizophrenie spezifische Veränderungen von Lernen und Entscheidungsfindung (u.a. eine pathologisch erhöhte Zuschreibung von Bedeutung zu eigentlich irrelevanten Informationen) und eine erhöhte Dopaminsynthese und -freisetzung, deren pharmakologische Korrektur durch Dopaminantagonisten zur Reduktion von psychotischen Symptomen führt. Es stellt sich damit die Frage nach den Mechanismen, durch die eine verstärkte Dopaminsynthese zur Entstehung von psychotischen Symptomen beitragen kann. Die „aberrant salience“ Theorie stellt diese Verknüpfung zwischen den neurobiologischen und den kognitiven Auffälligkeiten bei Patient*innen mit Schizophrenie her: Wie in Abschnitt

1.2 ausgeführt, reagiert ein großer Teil dopaminerger Neurone vor allem auf Reize, die überraschend, relevant und belohnend sind (Schultz, 2007). Eine Überaktivität dieser Neurone würde entsprechend verschiedenste, auch eigentlich irrelevante, Reize als bedeutsam erscheinen lassen, ein Zustand der der bei Patient*innen mit Psychose oft initial auftretenden „Wahnstimmung“ entspricht (Howes & Nour, 2016; Kapur, 2003). In den oben beschriebenen Bayesianischen Psychosemodellen entspricht dieser „aberrant salience“ ein relatives Übergewicht aktueller Reize gegenüber Vorannahmen. Da ein Feuern dopaminerger Neurone auch als Lernsignal fungiert (Berke, 2018; Wise, 2004), bedingen Aberrationen in ihrer Aktivität ein maladaptives Lernen, d.h. eine fehlerhafte Anpassung von Vorannahmen durch neue Informationen. Diese fehlerhafte Anpassung könnte wiederum zu einer Etablierung und Konsolidierung falscher Vorannahmen führen, was sich klinisch als Verfestigung und Systematisierung von Wahninhalten in späteren Krankheitsstadien niederschlägt (Sterzer et al., 2018).

Die Bayesianischen Psychosemodelle stellen den Versuch eines vereinheitlichenden Modells der Entstehung der vielfältigen Symptome einer Psychose dar und basieren auf etablierten statistischen Theorien zur Informationsverarbeitung. Trotz dieser konzeptionellen Vorteile sind viele Spezifika dieser Modelle allerdings noch ungeklärt und kaum empirisch untersucht. Eine dieser Unklarheiten betrifft die kognitiven Domänen, die von den hypothetisierten Veränderungen der Informationsverarbeitung betroffen sind. Wie oben angedeutet, wurden viele Konzepte der Bayesianischen Psychosemodelle an Hand von Theorien der Wahrnehmung (Aggelopoulos, 2015) entwickelt (Sterzer et al., 2018), werden aber auch auf Phänomene psychosetypischer Verzerrungen beim kognitiv-probabilistischen Schlussfolgern (wie jumping to conclusions) angewendet. Die unter 2.4 vorgestellte eigene Arbeit untersucht in diesem Kontext den Zusammenhang zwischen Neigung zu psychotischen Symptomen und der Nutzung von Vorannahmen und neuen Information in analogen Aufgaben der perzeptuellen und kognitiv-probabilistischen Entscheidungsfindung. Es ist außerdem nicht geklärt, welches Stadium der visuellen Informationsverarbeitung von den durch die Bayesianischen Psychosemodelle beschriebenen Veränderungen betroffen ist. Hierbei ist denkbar, dass diese Veränderungen nur oder hauptsächlich die späte, bewusste Verarbeitungsphase beeinflussen. Alternativ dazu könnten die Veränderungen frühe, automatische sensorische Verarbeitungsphasen betreffen, die den Zugang von Reizen zum Bewusstsein bestimmen. In der

unter 2.5 beschriebenen Studie haben wir zur Beantwortung dieser Frage eine Technik namens „continuous flash suppression“ genutzt, bei der einem Auge ein Zielreiz dargeboten wird, während dem anderen Auge eine dynamische Maske präsentiert wird, die den Zielreiz zunächst aus der bewussten Wahrnehmung ausblendet. Die Zeit, die der unterdrückte Reiz benötigt, um diese interokuläre Suppression zu überwinden, wurde als Index für die Stärke seiner unbewussten Verarbeitung genutzt (Gayet, Van der Stigchel, & Paffen, 2014). Indem wir die Zeit bis zur Überwindung der interkolulären Suppression von Gesichtern mit gerichteten Blick in Beziehung zur Neigung zu psychotischen Symptomen gesetzt haben, konnten wir testen, ob Proband*innen mit starker subklinischer Psychoseneigung eine besonders starke Vorannahme für gerichteten Blick während unbewusster visueller Verarbeitung aufweisen. Schließlich postulieren Bayesianische Psychosemodelle nicht nur eine veränderte *Nutzung*, sondern auch eine veränderte *Bildung und Korrektur* von Vorannahmen, was bislang nur sehr sporadisch empirisch getestet wurde. In der unter 2.6 dargestellten Studie haben wir dieses maladaptive Lernen und Anpassen von Vorannahmen daher in einem computationalen Lernmodell formalisiert und individuell angepasste Parameter dieses Modells in Bezug zu der Neigung der Teilnehmer*innen zu psychotischen Symptomen gesetzt.

Es soll an dieser Stelle außerdem bereits auf eine konzeptionelle Uneindeutigkeit der Bayesianischen Psychosemodelle hingewiesen werden, die sich später in widersprüchlichen empirischen Befunden ausdrückte und eine gravierende Herausforderung für diese Theorien darstellt. Da der spezifische Inhalt der Vorannahmen nicht eindeutig bestimmt ist, lassen sich Halluzinationen beispielsweise als Resultat eines *Übergewichts* von Vorannahmen konzipieren, in dem Sinne, dass uneindeutige sensorische Informationen im Sinne einer starken Vorannahme (von Stimmen oder anderen bedeutungsvollen Inhalten) interpretiert werden (strong prior Hypothese, (Corlett et al., 2019)). Umgekehrt wurden Halluzinationen auch als Resultat einer *unzureichenden* Überformung verrauschter sensorischer Informationen durch Vorannahmen konzipiert, wodurch eigentlich unwahrscheinliche Interpretationen, die sonst durch starke Vorannahmen unterdrückt worden wären, wirksam werden können (weak prior Hypothese, (Adams, Stephan, Brown, Frith, & Friston, 2013)). Wie im Folgenden gezeigt wird, haben beide Hypothesen in eigenen Arbeiten wie auch in Arbeiten anderer Arbeitsgruppen empirische Bestätigung gefunden. Mögliche Konsequenzen für Bayesianische Psychosemodelle werden in der Diskussion vorgeschlagen.

## 2. Eigene Arbeiten

### *2.1 Erhöhte Aktivierung belohnungsassoziierter Areale im Zusammenhang mit Entscheidungen für Alkoholkonsum*

Schädlicher Alkoholkonsum beinhaltet die wiederholte Entscheidung für Alkohol bei konkreten Trinkgelegenheiten trotz damit einhergehender negativer Konsequenzen. Wie in Abschnitt 1.3 dargestellt, gibt es zahlreiche fMRT-Studien, die Veränderungen in der Aktivierung von Arealen v.a. des Belohnungssystems beim (passiven) Betrachten von suchtbezogenen und nicht-suchtbezogenen Stimuli bei Abhängigkeitserkrankungen zeigen. Die Veränderungen in der Aktivierung bestimmter Hirnregionen, die mit *Entscheidungen* für oder gegen Alkoholkonsum einhergehen, wurden hingegen bisher nur sehr wenig untersucht, sind aber von großer Bedeutung für ein Verständnis der neurophysiologischen Veränderungen, die einen fortgesetzten schädlichen Konsum, oft wider besserer Vorsätze begünstigen. Basierend auf der dual system Theorie der Abhängigkeit, haben wir daher in dieser Studie getestet, ob ein hyperaktives Belohnungssystem und/oder ein beeinträchtigtes kognitives Kontrollsystem zu Entscheidungen für Alkoholkonsum beitragen.

Wortgetreu und selbstständig übersetztes Abstract des Originalartikels (Stuke, H, Gutwinski, S, Wiers, C E, Schmidt, T T, Gropper, S, Parnack, J, . . . Bermpohl, F. To drink or not to drink: Harmful drinking is associated with hyperactivation of reward areas rather than hypoactivation of control areas in men. J Psychiatry Neurosci 2016; 41(3): E24-36.):

*„Hintergrund: Die Aufrechterhaltung eines schädlichen Alkoholkonsums kann als eine wiederholte Entscheidung für Alkohol bei konkreten Trinkanlässen betrachtet werden. Diese Entscheidungen werden oft trotz der Absicht getroffen, mit dem Alkoholkonsum aufzuhören oder ihn zu reduzieren. Wir haben untersucht, ob ein hyperaktives Belohnungssystem und/oder ein gestörtes kognitives Kontrollsystem zu solchen ungünstigen Entscheidungen beitragen. Methoden: In dieser fMRT-Studie trafen Männer mit mäßigem bis schädlichem Trinkverhalten, das mit dem Alcohol Use Disorders Identification Test (AUDIT) gemessen wurde, wiederholt Entscheidungen zwischen alkoholischen und nichtalkoholischen Getränken. Auf der Grundlage vorangegangener individueller Bewertungen wurden Entscheidungspaare gebildet, bei denen eine alkoholische Entscheidungsoption von den Teilnehmern als wünschenswerter, aber weniger*

*vorteilhaft angesehen wurde. Durch Korrelation der AUDIT-Ergebnisse mit der Hirnaktivierung während der Entscheidungsfindung konnten wir feststellen, welche Hirnareale bei Männern mit höherem Alkoholkonsum explizit mit Entscheidungen zugunsten von Alkohol in Verbindung stehen.*

*Ergebnisse: Achtunddreißig Männer nahmen an unserer Studie teil. In Bezug auf das Verhalten fanden wir eine positive Korrelation zwischen den AUDIT-Werten und der Anzahl der Entscheidungen für begehrte alkoholische Getränke im Vergleich zu gesünderen nichtalkoholischen Getränken. Die fMRT-Ergebnisse zeigten, dass die AUDIT-Scores positiv mit der Aktivierung von Bereichen, die mit der Belohnungs- und Motivationsverarbeitung in Zusammenhang stehen (d. h. ventrales Striatum, Amygdala, medialer präfrontaler Kortex), korreliert waren, während Entscheidungen zugunsten eines begehrten, ungesünderen alkoholischen Getränks getroffen wurden. Umgekehrt fanden wir keine Hypoaktivierung in Bereichen, die mit Selbstkontrolle in Verbindung gebracht werden (dorsolateraler präfrontaler Kortex). Diese Effekte traten nicht auf, wenn die Teilnehmer ein begehrtes, ungesünderes nicht-alkoholisches Getränk wählten.*

*Limitationen: Die Männer, die an unserer Studie teilnahmen, mussten abstinent sein und würden am Ende des Experiments möglicherweise ein alkoholisches Getränk konsumieren. Daher haben wir manifeste Alkoholabhängigkeit nicht als Einschlusskriterium definiert und uns stattdessen auf weniger stark betroffene Personen konzentriert.*

*Schlussfolgerung: Unsere Ergebnisse deuten darauf hin, dass mit zunehmendem Schweregrad des Alkoholkonsums die Entscheidung für alkoholische Getränke mit einer zunehmenden Aktivität in belohnungsassoziierten neuronalen Systemen und nicht mit einer abnehmenden Aktivität in selbstkontrollassoziierten Systemen verbunden ist.“*

Unsere Ergebnisse in dieser Studie sprechen dafür, dass Entscheidungen für alkoholische Getränke mit zunehmendem Alkoholkonsum eher mit einer zunehmenden Aktivität in belohnungsassoziierten Systemen als mit einer abnehmenden Aktivität in mit kontrollassoziierten Systemen verbunden sind. Die der Hypothese entgegengesetzten Aktivitäts*steigerungen* kontrollassoziierter Areale bei Entscheidungen für Alkoholkonsum mit zunehmender Trinkschwere können eventuell als Resultat gescheiterter Selbstkontrollbemühungen gewertet werden (im Sinne eines mit zunehmender Trinkschwere schwereren Konflikts zwischen Konsumverlangen und Kontrollbemühungen mit Blick auf

bestehende Reduktionsziele). Limitationen und weitere Implikationen der Resultate werden im Abschnitt 3.1 diskutiert.

Research Paper

# To drink or not to drink: Harmful drinking is associated with hyperactivation of reward areas rather than hypoactivation of control areas in men

Heiner Stuke, MD; Stefan Gutwinski, MD; Corinde E. Wiers, PhD; Timo T. Schmidt; Sonja Gröpper, MD; Jenny Parnack, MD; Christiane Gawron, MD; Catherine Hindi Attar, PhD; Stephanie Spengler, PhD; Henrik Walter, MD, PHD; Andreas Heinz, MD, PhD; Felix Bermpohl, MD, PhD

**Background:** The maintenance of harmful alcohol use can be considered a reiterated decision in favour of alcohol in concrete drinking occasions. These decisions are often made despite an intention to quit or reduce alcohol consumption. We tested if a hyperactive reward system and/or an impaired cognitive control system contribute to such unfavourable decision-making. **Methods:** In this fMRI study, men with modest to harmful drinking behaviour, which was measured using the Alcohol Use Disorders Identification Test (AUDIT), repeatedly made decisions between alcoholic and nonalcoholic drinks. Based on prior individual ratings, decision pairs were created with an alcoholic decision option considered more desirable but less beneficial by the participant. By correlating AUDIT scores with brain activation during decision-making, we determined areas explicitly related to pro-alcohol decisions in men with greater drinking severity. **Results:** Thirty-eight men participated in our study. Behaviourally, we found a positive correlation between AUDIT scores and the number of decisions for desired alcoholic drinks compared with beneficial nonalcoholic drinks. The fMRI results show that AUDIT scores were positively associated with activation in areas associated with reward and motivation processing (i.e., ventral striatum, amygdala, medial prefrontal cortex) during decisions favouring a desired, nonbeneficial alcoholic drink. Conversely, we did not find hypoactivation in areas associated with self-control (dorsolateral prefrontal cortex). These effects were not present when participants chose a desired, nonbenefical, nonalcoholic drink. **Limitations:** The men participating in our study had to be abstinent and would potentially consume an alcoholic drink at the end of the experiment. Hence, we did not define manifest alcohol dependence as an inclusion criterion and instead focused on less severely affected individuals. **Conclusion:** Our results indicate that with growing drinking severity, decisions for alcoholic drinks are associated with increasing activity in reward-associated neural systems, rather than decreasing activity in self-control–associated systems.

## Introduction

The maintenance of harmful alcohol use implies reiterated decisions to consume alcohol in concrete drinking occasions. These decisions are often made despite an intention to quit or reduce alcohol consumption. Although there is quite a large body of evidence on neural responsivity to alcohol cues or neural mechanisms of general decision-making capacities in individuals with alcohol use disorders, the neural processes during real drinking decisions remain largely unclear.

Dual-process models of addiction[1,2] state the importance of 2 distinct but interacting systems during decisions for and against alcohol consumption. On the one hand, a reward system (also referred to as an impulsive, motivational, or reflexive system) has been implicated in the immediate emotional assessment of stimuli and automatic (approach) behaviour. On the other hand, a cognitive control system (also referred to as a deliberative or reflective system) that modulates this primary assessment by integration of higher-order considerations, such as long-term effects of a possible decision, has been suggested. In theory, both a hyperactive reward system and an impaired control system may contribute to addictive behaviour. Indeed, behavioural and neuroimaging data suggest alterations in both systems in individuals with substance use disorders.

Alcohol-dependent or heavily drinking individuals show subjective craving[3] and automatic approach tendencies[4,5] when confronted with alcoholic drinks, and a substantial body of literature suggests that such addiction-related behaviour is

**Correspondence to:** H. Stuke, Department of Psychiatry and Psychotherapy, Charité — Universitätsmedizin Berlin, Campus Mitte, Charitéplatz 1, D-10117 Berlin, Germany; heiner.stuke@charite.de

Submitted June 1, 2015; Revised Sept. 20, 2015; Accepted Oct. 21, 2015; Early-released Feb. 23, 2016

DOI: 10.1503/jpn.150203

associated with an overactive reward system. Specifically, fMRI studies have consistently linked alcohol cue reactivity (i.e., brain responses to the presentation of alcohol stimuli) with the amygdala, ventral striatum and ventromedial prefrontal cortex (VMPFC) in both alcohol-dependent patients[6–14] and heavy drinkers.[15–17]

Moreover, alcohol-dependent patients showed activation of the ventral striatum and VMPFC when approaching versus avoiding alcohol compared with fruit juice in a joystick task,[18] and activation of the amygdala and ventral striatum has been reported to correlate with subjective craving in alcohol-dependent patients.[14,18] Thus, hyperactivity in reward-associated neural systems appears to play a role in craving and approach behaviour. Conversely, this enhanced response to alcohol-related stimuli may be accompanied by an attenuated response to nonalcoholic rewarding stimuli.[12,18] This neuroimaging finding is behaviourally paralleled by a loss of interest in activities that are not related to alcohol consumption.

On the other hand, previous findings suggest impaired self-control function in alcohol-dependent or heavily drinking individuals. At the behavioural level, these individuals show a preference for short-term rather than long-term rewards,[19] as well as for riskier decision options.[20] At the neural level, this may correspond to attenuated activity of the second system in dual system models of decision-making. This control system supposedly modifies automatic behaviour by integrating goals related to long-term benefits.

In healthy individuals, dorsolateral prefrontal cortex (DLPFC) activation has been associated with a preference for long-term over short-term rewards,[21,22] whereas disruption of the DLPFC by repetitive transcranial magnetic stimulation (rTMS) has been shown to promote impulsive decision behaviour.[23] In an fMRI study on healthy dieters choosing between a tastier and a healthier food product, decisions in favour of the healthier product were correlated with increased DLPFC activation.[24] In line with this finding, lesions of the DLPFC led to the inability to change dysfunctional decision patterns.[25] In indivdiuals with substance use disorders, neuroimaging studies have shown attenuated DLPFC activity during inhibitory control tasks.[26,27] Furthermore, the DLPFC was more active in smokers when using cognitive strategies to suppress craving.[28] Taken together, these findings suggest that functional and structural alterations in self-control areas could lead to the inability to resist craving despite the intention to quit drinking.

Behavioural and neuroimaging data suggest that alterations in the reward as well as in the control system contribute to addictive behaviour. An overwhelming desire (associated with hyperactivation of reward-associated circuits) as well as impaired control processes (associated with hypoactivation in control-associated areas) may contribute to the maintenance of substance use despite awareness of its harmful consequences. The aforementioned fMRI studies either focused on passive exposure to alcohol-related stimuli (thus studying responsivity of the reward system to alcohol cues, independent of actual decision-making situations) or on general decision-making tasks, such as the Iowa Gambling Task[29] or the Monetary Delayed Discounting Task[30] (thus studying control processes independent of alcohol stimuli). Hence, these studies mainly focused either on reward or control processes in addiction. The present study addressed the question of how both systems interact during real-life drinking decisions and how this interplay is altered with increasing drinking severity.

For this purpose, we used an fMRI task where individuals with widely differing drinking severity decided between alcoholic and nonalcoholic drinks. The decision options were individually designed in a way that participants experienced a conflict between the desire and benefit associated with the respective drinks. We implemented a real-world decision by scheduling scanning sessions on Friday or Saturday evenings and by serving one of the chosen drinks directly after scanning. By this means, the paradigm established by Hare and colleagues[24] was adopted to elucidate the neural mechanisms of decisions for desired, nonbeneficial alcoholic drinks. Specifically, we tested if increased activity of reward areas (hypothesis of overwhelming desire), decreased activity of self-control areas (hypothesis of impaired control processes) or a combination of both promotes harmful pro-alcohol decisions.

## Methods

### *Participants*

We recruited men between 20 and 60 years old through advertising for participation in the study. Exclusion criteria were withdrawal symptoms when abstinent, cannabis consumption 4 weeks before participation and substance dependence other than alcohol and/or nicotine. Participants were told before they enrolled in the study that there would be urine toxicology tests on a random basis. In practice, this random screening was not performed, and we relied on the participants' self-disclosure instead. In addition, to be eligible for participation, individuals were required to have no other DSM-IV Axis-I disorders and no history of head trauma or neurologic disorders. To guarantee a general awareness of health issues, participants were asked about eating habits and health awareness in the screening interview.

Participants were screened for DSM-IV criteria for alcohol abuse and alcohol dependence using the Mini-International Neuropsychiatric Interview (M.I.N.I.).[31] As participants received real drinks at the end of the experiment, we did not include abstinent or immediately treatment-seeking participants to avoid the risk of provoking relapses. After the experiment, all participants were given information on addiction counselling centres and treatment possibilities.

Participants completed the following questionnaires concerning drinking behaviour: the Alcohol Use Disorders Identification Test[32] (AUDIT; assessing harmful drinking on a scale of 0 to 40), the Obsessive Compulsive Drinking Scale[33] (OCDS) and the Alcohol Dependence Scale[34] (ADS). The AUDIT was used as the main variable modelling severity of harmful drinking.

We collected the following additional information to allow strict control over confounding variables and potential psychiatric comorbidities. Handedness was assessed using the Edinburgh Handedness Inventory[35] (EHI), and the Matrix Reasoning Test of the Wechsler Adult Intelligence Scale[36]

(WAIS) was used as a proxy of general intelligence. We assessed depressive symptoms using the Beck Depression Inventory (BDI), anxiety using the State-Trait Anxiety Inventory[37] (STAI) and impulsiveness using the Barratt Impulsiveness Scale[38] (BIS) and the Monetary Choice Questionnaire[39] (MCQ). The Lifetime Drinking History (LDH[40]) was used to assess the participants' drinking behaviour over the lifespan.

The study was approved by the Ethical Committee of the Charité, Universitätsmedizin Berlin. After complete description of the study, written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

### *Experimental setting*

Participants were instructed not to drink anything for 2 hours before the scanning session to ensure a basic level of thirst. Because participants arrived at the scanning site 90 min before the fMRI session to perform the ratings and fill out consent forms and questionnaires, we can say for certain that they did not drink within this timeframe. Moreover, every session was scheduled for evenings before either weekends or public holidays to guarantee drinking willingness. Before the experiment, a minibar with drinks was presented to the participant in a room near the scanning room, and the participants were told that 1 of the decisions made during the experiment would be implemented after the experiment.

### *Ratings*

Prior to scanning, participants rated 120 photographs depicting alcoholic drinks (e.g., beer, wine, liquor) as well as a variety of nonalcoholic drinks (e.g., lemonade, milk, juice) with regard to desire and beneficence. The wording of the 2 questions was (translated from German) "In your honest opinion, how great is your desire to have this drink right now?" for the desire rating and "How beneficial/harmful would it be to have this drink?" for the benefit rating. In both cases, the scale reached from –4 to 4, with 0 as a neutral value (Fig. 1). The drinks were presented using high-resolution colour pictures matched for luminance and size between alcoholic and nonalcoholic items. We used the ratings to create conflicting pairs of a more beneficial and a more desired drink in the decision task. The image set's suitability to create such conflicting pairs was investigated beforehand and optimized in a behavioural pilot study involving 8 participants.

### *Decision task*

In the decision task, 2 images of drinks were presented simultaneously, followed by a fixation cross (Fig. 2). Within a 4-s interval, participants chose (by button press) between 2 nonalcoholic drinks or between an alcoholic and a nonalcoholic drink. The decisions involving an alcoholic drink are hereafter referred to as "alcohol trials," and those with 2 nonalcoholic drinks are refferred to as "nonalcohol" trials. Decision options were presented in such a way that they induced a conflict within the participant between the desire and the benefit associated with the consumption of the respective drinks. That is, based on the prescan ratings, decision options were presented where 1 drink was considered more beneficial and the other more desirable by the participant. Moreover, "close" conflict pairs that differed by only 1 point on both scales were excluded from analysis.

Depending on the participants' prior ratings and the real decision, each trial was subsumed under 1 of 4 conditions that were defined as follows (Fig. 1):

- SA: successful self-control in an alcohol–nonalcohol conflict (e.g., choosing the less desired, more beneficial nonalcoholic item),
- FA: failed self-control in an alcohol–nonalcohol conflict (e.g., choosing the more desired, less beneficial alcoholic item),
- SN: successful self-control in a nonalcohol–nonalcohol conflict (e.g., choosing the less desired, more beneficial nonalcoholic item), and
- FN: failed self-control in a nonalcohol–nonalcohol conflict (e.g., choosing the more desired, less beneficial nonalcoholic item).

The decision task was split into 4 runs of 50 trials each. For 12 of the participants, the ratings did not allow us to create the 200 conflicting stimulus pairs, so fewer trials were tested (range 50–192 decisions per participant). The reason for this was a correlation between desire and benefit ratings in these participants, which led to a reduced number of pairs with conflict between benefit and desire of the drinks and — because only pairs with this conflict were shown to the participant — to a reduced number of decisions. However, a confounding effect of this imbalance is unlikely since the number of trials per participant was not correlated with our variable of interest, the AUDIT scores ($p = 0.77$).

The general functionality of the task and the stimulus set was tested with a proof-of-concept analysis comparing blood-oxygen level–dependent (BOLD) responses between alcohol and nonalcohol trials (FA + SA) > (FN + SN). As expected, this analysis yielded strong effects in the posterior and anterior cingulate cortex and the medial prefrontal cortex (inter alia, family-wise error [FWE]–corrected whole brain analysis). Because of their replicative character, these results are not reported in the Results section.

### *FMRI data acquisition and preprocessing*

We used a Siemens Trio 3 T scanner equipped with a 12-channel head coil to acquire MRI volumes. $T_2$*-weighted gradient-echo echo-planar images (EPI) containing 36 axial slices (3.5 mm thick, interleaved) without interslice gap were acquired with the following imaging parameters: repetition time (TR) 2250 ms, echo time (TE) 30 ms, flip angle 80°, matrix size 64 × 64 and field of view (FOV) 134 mm, resulting in a voxel size of 3.5 × 3.5 × 3.5 mm. Images were acquired in an oblique orientation of 30° to the anterior commissure–posterior commissure line. High resolution $T_1$-weighted structural data were collected for anatomic localization, with TR 900 ms, TE 2.52 ms, matrix size 256 × 256, FOV 256 mm, 192 slices (1 mm thick) and flip angle 9°.

We preprocessed functional scans using SPM8 software.[41] Functional images were corrected for slice-acquisition time (using sinc interpolation), realigned and unwarped. The high-resolution $T_1$ image was coregistered with the mean EPI image and subsequently segmented. Images were normalized using DARTEL and the segmented grey and white matter maps. Finally, images were spatially smoothed with an 8 mm full-width at half-maximum Gaussian kernel.

*First-level analyses*

After preprocessing, individual data analysis was performed using SPM8. For each participant, we used the onsets of presentation of the decision options to generate regressors for the 4 conditions (SA, FA, SN, FN) in an event-related design (see the Decision task section and Fig. 1). We used the realignment parameters of the motion correction as covariates

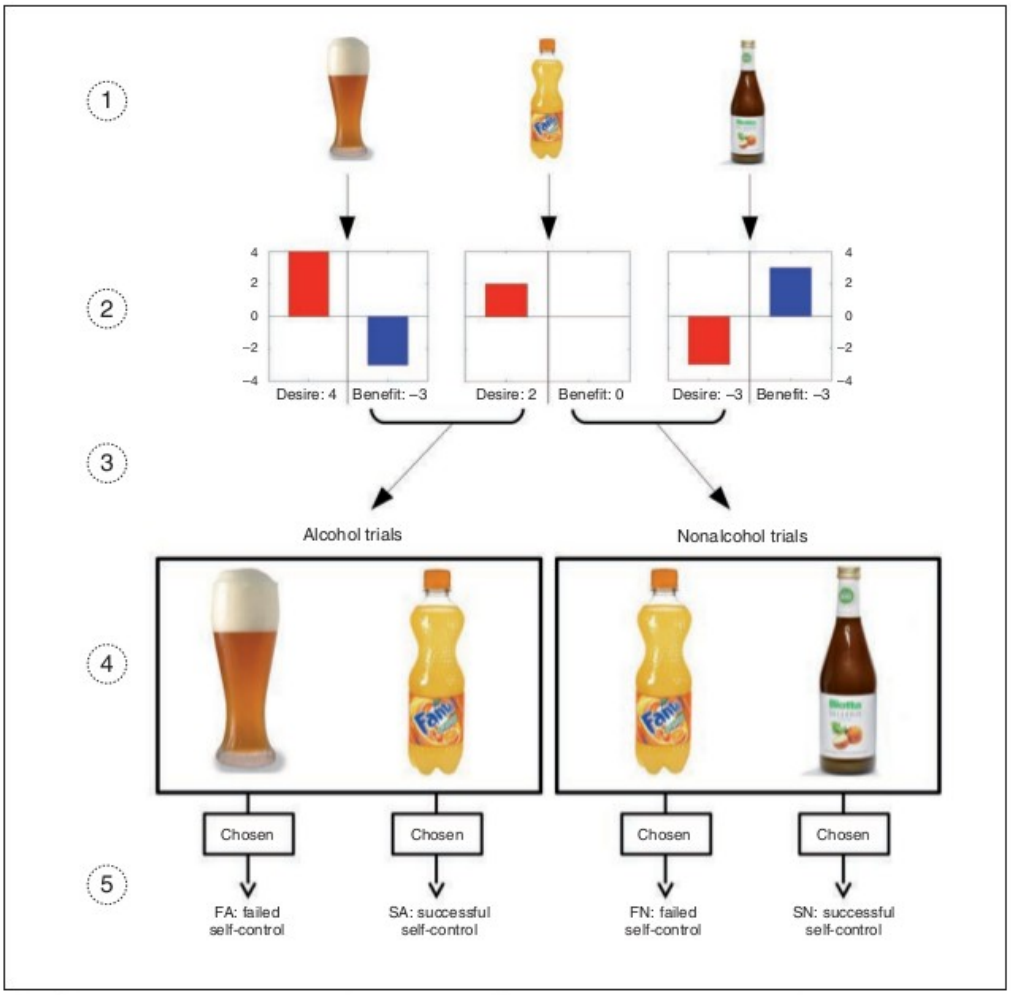


**Fig. 1:** Stimulus set, ratings and conditions of the decision task. (**1**) The stimulus set comprised images of 120 alcoholic and nonalcoholic drinks. (**2**) These drinks were rated by the participant in terms of desire to drink and beneficence of the drink. (**3**) Pairs of drinks inducing a conflict between desire and benefit were generated based on the individual ratings. (**4**) During the fMRI session, the participant chose between the 2 drinks. (**5**) All decisions made by participants during the decision task were assigned to 1 of 4 conditions: successful self-control in an alcohol–nonalcohol conflict (SA; e.g., choosing the nonalcoholic item), failed self-control in an alcohol–nonalcohol conflict (FA; e.g., choosing the alcoholic item), successful self-control in a nonalcohol–nonalcohol conflict (SN; e.g., choosing the less desired, more beneficial nonalcoholic item), and failed self-control in a nonalcohol–nonalcohol conflict (FN; e.g., choosing the less beneficial, more desired nonalcoholic item).

of no interest. Subsequently, specific *t* contrast images (see the Contrast testing section) were created and entered into the second-level group analyses.

*Second-level analyses*

For every contrast image created in the first-level analyses, we performed a group-level correlation analysis between AUDIT scores and contrast-specific brain activation using the Multiple Regression Design of SPM (see the "Contrast testing" section). Because there was an association between AUDIT scores and age ($r = 0.27$, $p = 0.10$), we included age as a covariate of no interest to preclude a confounding influence of age differences. For this analysis, we used a priori regions of interest (ROIs) for small-volume $\alpha$ error adjustment. Based on prior studies on neural correlates of alcohol-related cue reactivity, craving and approach behaviour, we included the amygdala, striatum and MPFC as ROIs to test our hypothesis of overwhelming desire. These ROIs are hereafter referred to as "reward-associated areas," although this wording certainly does not cover all cognitive processes previously proposed for these areas. Conversely, we used the DLPFC as an ROI to test our hypothesis of impaired control processes ("control-associated area"). The striatum, amygdala and MPFC were defined as described by Beck and colleagues[7] using a combination of anatomic hypotheses and previous fMRI findings regarding alcohol cue reactivity. As the DLPFC is anatomically not clearly defined and has not been reported in cue reactivity studies, a functionally defined ROI was downloaded from an online atlas.[42] All imaging results are presented with a significance threshold of $p < 0.05$, small volume–corrected for the amygdala, striatum, MPFC and DLPFC ROIs and using FWE correction to account for multiple testing.

*Contrast testing*

To study how brain activation during pro-alcohol decisions varies with drinking severity, we correlated AUDIT scores with specific BOLD contrasts obtained during the decision task. We aimed to identify 2 types of brain regions: areas whose activation was positively correlated with drinking severity during pro-alcohol decisions (reward-associated areas according to the hypothesis of overwhelming desire) and areas whose activation was negatively correlated with drinking severity (control-associated areas according to the hypothesis of impaired control processes).

To ensure the specificity of our findings for alcohol trials, we used decisions for more desired drinks in nonalcohol trials (FN trials) as a control condition (resulting in the contrast AUDIT × [FA – FN]). To further ensure the specificity for trials with a failure in self-control (i.e., to preclude a sole alcohol effect causing activations in AUDIT × [FA – FN]), we then subtracted the analogous contrast for successful self-control trials. This calculation yielded the interaction contrast AUDIT × [(FA – FN) – (SA – SN)], which represents the impact of growing drinking severity on activation during decisions for the more desired alcoholic drink compared with both decisions for the more desired nonalcoholic drink and decisions against the alcoholic drink. Thus, the contrasts AUDIT × (FA – FN) and AUDIT × [(FA – FN) – (SA – SN)] can be used to test the hypothesis of overwhelming desire (enhanced activation of reward areas during pro-alcohol decisions with growing drinking severity). Analogically, the inverse correlations –AUDIT × (FA – FN) and –AUDIT × [(FA – FN) – (SA – SN)] were computed indicating which areas show decreasing activations during pro-alcohol decisions with growing drinking severity (test for hypothesis of impaired control processes).

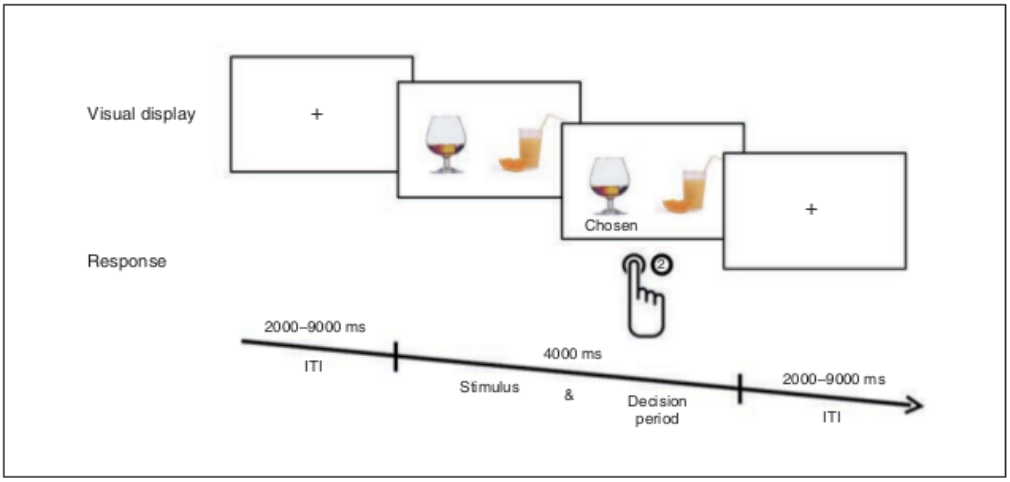


**Fig. 2:** Test sequence in the decision task. Two drinks were presented simultaneously. Participants had to choose 1 of the drinks within 4000 ms by pressing a button. After pressing the button, a fixation cross was presented for a variable intertrial interval (ITI) lasting 2000–9000 ms.

*Behavioural analyses*

For the 4 conditions SA, FA, SN and FN, we calculated the number of trials per condition and subject-wise mean response times. To check the validity of the AUDIT scores, we computed Pearson correlations between AUDIT scores and other alcohol-related measures (OCDS, ADS).

As a proxy of general impulsiveness, we correlated AUDIT scores with the general proportion of failed self-control trials (all failed self-control trials ÷ by all trials). As a measure of tendency to more likely fail in alcohol trials, we computed the ratio of failed self-control rates between alcohol and non-alcohol trials, referred to as "alcohol-specific failed self-control." It was correlated with AUDIT scores to check if this alcohol-specific failed self-control was more likely to occur in participants with more severe drinking.

We compared mean response times (RTs) between the different conditions as another measure of impulsive decision making. Analogous to contrast testing of imaging data, the interaction contrast of response times ([FAReact – FNReact] – [SAReact – SNReact]) was used to ensure the highest possible specificity for failed self-control decisions in favour of alcohol.

## Results

*Participants*

We recruited 44 men to participate in the study. Five of them had to be excluded from the analysis for technical reasons, and 1 was excluded because of an incidental finding, leaving 38 men for data analysis. All participants were right-handed. Seventeen participants fulfilled DSM-IV criteria for alcohol abuse and 2 further fulfilled the criteria for alcohol dependence. Table 1 summarizes the final sample's demographic and behavioural features.

*Behavioural results*

To check the validity of AUDIT measures, we computed correlations between AUDIT, OCDS, ADS and LDH scores. These analyses revealed a significant correlation between AUDIT and OCDS ($r = 0.768$, $t_{36} = 7.19$, $p < 0.001$), AUDIT and ADS ($r = 0.828$, $t_{36} = 8.86$, $p < 0.001$) and AUDIT and alcohol consumption per month ($r = 0.561$, $t_{36} = 4.06$, $p < 0.001$) as well as for the entire life ($r = 0.513$, $t_{36} = 3.59$, $p = 0.001$) as measured

**Table 1: Sample characteristics and behavioural results**

| Characteristic* | No. participants | Range | Mean ± SD | Pearson *R* |
|---|---|---|---|---|
| Age, yr | 38 | 23 to 49 | 32.53 ± 7.13 | 0.27 |
| Age at first drunken stupor, yr | 38 | 12 to 18 | 15.16 ± 1.59 | 0.19 |
| Alcohol Dependence Scale Score | 36 | 25 to 54 | 32 ± 7.04 | 0.83† |
| Alcohol-specific failed self-control (ratio of failed self-control rates between alcohol and nonalcohol trials) | 37 | 0.07 to 3.25 | 1.18 ± 0.56 | 0.41‡ |
| AUDIT score | 38 | 2 to 30 | 11.08 ± 7.05 | — |
| No. drinking d/wk in the last mo | 38 | 0.25 to 6 | 2.98 ± 2.04 | 0.48† |
| No. of drinks per drinking d in the last mo | 38 | 3 to 12 | 8.16 ± 2.95 | 0.54† |
| Barratt Impulsiveness Scale score | 38 | 42 to 171 | 69.43 ± 25.37 | 0.13 |
| Beck Depression Inventory score | 35 | 21 to 119 | 27.94 ± 16.27 | 0.11 |
| Edinburgh Handedness Inventory quotient | 38 | 10 to 100 | 81.75 ± 19.53 | 0.05 |
| Interaction effec in response times [(FA – FN) – (SA –SN)] | 37 | –848.53 to 876.82 | –5.7 ± 401.14 | 0.37‡ |
| IQ | 34 | 70.00 to 115.00 | 96.91 ± 10.87 | 0.14 |
| Lifetime Drinking History alcohol intake per mo, g | 38 | 82 to 9465 | 1762 ± 1774 | 0.56† |
| Lifetime Drinking History total alcohol intake, g | 38 | 4861 to 2 754 299 | 390 481 ± 524 030 | 0.51† |
| Response time for FA trials, ms | 38 | 972.57 to 2354.06 | 1499.08 ± 287.90 | 0.36‡ |
| Response time for FN trials, ms | 38 | 934.76 to 2303.95 | 1536.92 ± 316.19 | 0.19 |
| Response time for SA trials, ms | 37 | 996.96 to 3063.50 | 1811.08 ± 479.95 | –0.22 |
| Response time for SN trials, ms | 38 | 994.33 to 3031.50 | 1864.11 ± 438.21 | 0.20 |
| Monetary Choice Questionnaire — Discounting Index score | 38 | 0.0003 to 69 | 0.019 ± 0.019 | 0.20 |
| Obsessive Compulsive Drinking Scale score | 35 | 2 to 28 | 11.09 ± 6.16 | 0.77† |
| Proportion of failed self-control trials in all trials | 38 | 0.11 to 0.98 | 0.72 ± 0.22 | 0.03 |
| State-Trait Anxiety Inventory score | 38 | 45 to 52 | 48.92 ± 1.81 | 0.01 |
| Total abstinence, mo | 36 | 0 to 7 | 1.45 ± 1.95 | 0.16 |
| Total drinking, yr | 38 | 5 to 34 | 16 ± 7 | 0.30 |
| Education, yr | 38 | 10 to 22 | 16.36 ± 2.79 | –0.15 |

AUDIT = Alcohol Use Disorders Identification Test; FA = failed self-control in an alcohol–nonalcohol conflict; FN = failed self-control in a nonalcohol–nonalcohol conflict; SA = successful self-control in an alcohol–non-alcohol conflict; SD = standard deviation; SN = successful self-control in a nonalcohol–nonalcohol conflict.
*Eighteen participants were smokers and 20 were not ($p = 0.21$, 2-sample $t$ test).
†Significant at a threshold of $p < 0.01$.
‡Significant at a threshold of $p < 0.05$.

with LDH. There was a positive correlation between drinking severity, as reflected in the AUDIT scores, and our behavioural measure of alcohol-specific failed self-control (see the Behavioural analyses section; $r = 0.41$, $t_{36} = 2.70$, $p = 0.012$). That is, with increasing AUDIT scores, participants failed more often in alcohol than in nonalcohol trials. Moreover, with increasing AUDIT scores, participants made significantly faster decisions in alcohol trials than in nonalcohol trials in failed compared with successful self-control (interaction effect for response times AUDIT × [(FAResp – FNResp) – (SAResp – SNResp)] ($r = –0.371$, $t_{36} = –2.40$, $p = 0.024$).

There was no significant correlation between AUDIT scores and EHI scores, intelligence (matrices subtest of WAIS), BDI scores, years of education, STAI scores and impulsiveness (general proportion of failed self-control, BIS, MCQ), excluding these variables as possible confounders (Table 1).

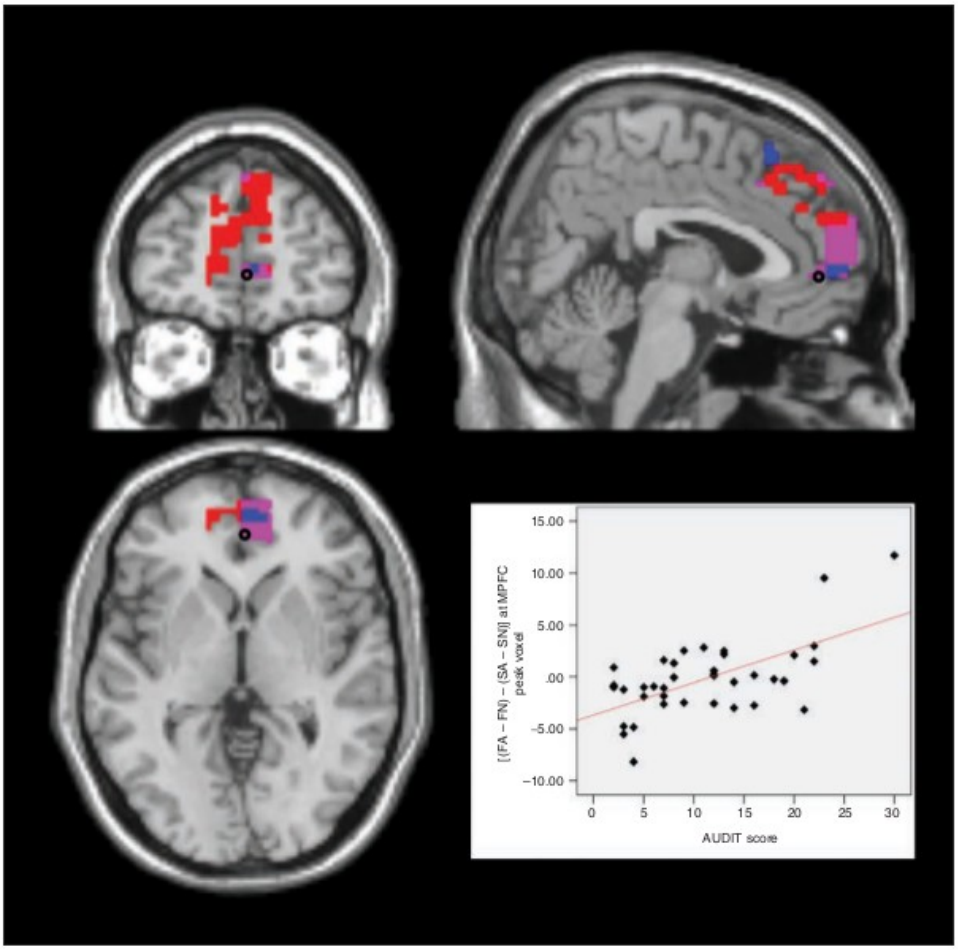


**Fig. 3:** Failed self-control in alcohol trials — medial prefrontal cortex (MPFC). Section views showing significant clusters for AUDIT × (FA – FN) and AUDIT × [(FA – FN) – (SA – SN)] within the MPFC at a threshold of $p < 0.05$, family-wise error–corrected. Custers are presented at a threshold of $p < 0.005$, uncorrected. Red: MPFC cluster with higher activity in participants with higher drinking severity in failed self-control in alcohol compared with nonalcohol trials (AUDIT × [FA – FN]) Blue: MPFC cluster with higher activity in participants with higher drinking severity in failed compared with successful self-control in alcohol compared with nonalcohol trials (AUDIT × [(FA – FN) – (SA – SN)]) Violet: overlap between these 2 clusters. Plot: effect of interaction contrast (FA – FN) – (SA – SN) at the marked peak voxel (Montreal Neurological Institute space: $x, y, z = 4, 49, 0$) plotted subject-wise against AUDIT score. AUDIT = Alcohol Use Disorders Identification Test; FA = failed self-control in an alcohol–nonalcohol conflict; FN = failed self-control in a nonalcohol–nonalcohol conflict; SA = successful self-control in an alcohol–non-alcohol conflict; SN = successful self-control in a nonalcohol–nonalcohol conflict.

*Imaging results*

To study the effect of increasing drinking severity on brain activation during failed self-control in favour of alcohol (pro-alcohol decisions), we correlated AUDIT scores with activation during failed self-control in alcohol compared with failed self-control in nonalcohol trials.

**Hyperactivated areas during pro-alcohol decisions**

According to the hypothesis of overwhelming desire, reward-associated areas should show enhanced activation during pro-alcohol decisions, and this hyperactivation should increase with growing drinking severity.

The corresponding analysis testing positive correlations between drinking severity and brain activation during pro-alcohol decisions (i.e., AUDIT × [FA – FN]) revealed significant results in the bilateral striatum (peak left in Montreal Neurological Institute [MNI] space: $x, y, z = -4, 7, 4$, $t_{35} = 4.34$, $p_{FWE} = 0.013$, extent = 9; peak right: $x, y, z = 35, -18, -7$, $t_{35} = 3.81$, $p_{FWE} = 0.046$, extent = 9; clusters were localized in the ventral striatal parts), in the bilateral MPFC (peak left: $x, y, z = 0, 60, 18$, $t_{35} = 4.29$, $p_{FWE} = 0.018$, extent = 82; peak right: $x, y,$

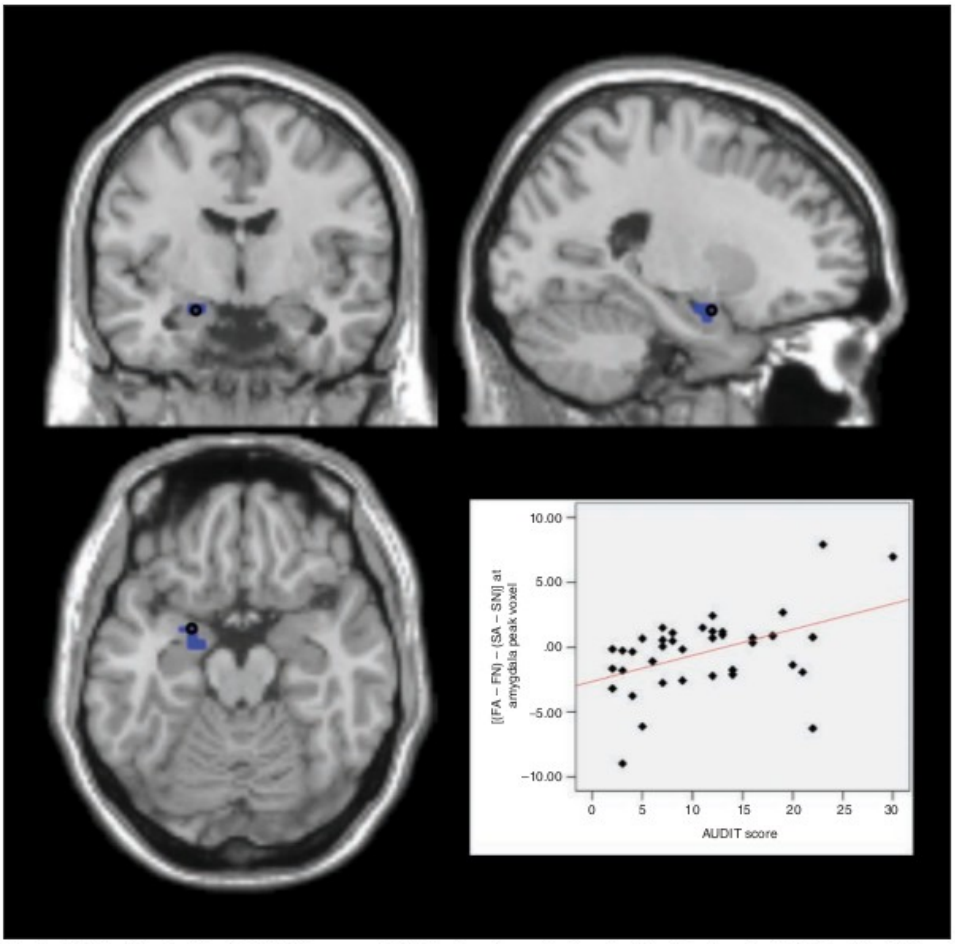


**Fig. 4:** Failed self-control in alcohol trials — amygdala. Section views showing significant clusters for AUDIT × (FA – FN) and AUDIT × [(FA – FN) – (SA – SN)] within the amygdala at a threshold of $p < 0.05$, family-wise error–corrected. Clusters are presented at a threshold of $p < 0.005$, uncorrected. Blue: amygdala cluster with higher activity in participants with higher drinking severity in failed compared with successful self-control in alcohol compared with nonalcohol trials (AUDIT × [(FA – FN) – (SA – SN)]). Plot: effect of interaction contrast (FA – FN) – (SA – SN) at the marked peak voxel (Montreal Neurological Institute space: $x, y, z = -21, 0, -18$) plotted subject-wise against AUDIT score. AUDIT = Alcohol Use Disorders Identification Test; FA = failed self-control in an alcohol–nonalcohol conflict; FN = failed self-control in a nonalcohol–nonalcohol conflict; SA = successful self-control in an alcohol–non-alcohol conflict; SN = successful self-control in a nonalcohol–nonalcohol conflict.

$z$ = 4, 56, 18, $t_{35}$ = 4.71, $p_{FWE}$ = 0.005, extent = 105), and in the left DLPFC (peak: $x$, $y$, $z$ = –18, 18, 60, $t_{35}$ = 4.87, $p_{FWE}$ = 0.002, extent = 50). Notably, these correlations were driven by both a positive AUDIT × FA correlation and a negative AUDIT × FN correlation (Appendix 1, Figs. S1–S3, available at jpn.ca), indicating enhanced activation of reward-associated areas during decisions in favour of alcohol as well as attenuated activation during decisions in favour of desirable nonalcoholic drinks.

To preclude a sole alcohol effect causing the activations in AUDIT × (FA – FN), we then subtracted the analogous activation for successful self-control trials from the above contrast. For the resulting analysis, AUDIT × [(FA – FN) – (SA – SN)], we found significant results in the left amygdala (peak: $x$, $y$, $z$ = –21, 0, –18, $t_{35}$ = 3.64, $p_{FWE}$ = 0.011, extent = 3) and in the left DLPFC (peak: $x$, $y$, $z$ = –28, 11, 63, $t_{35}$ = 4.14, $p_{FWE}$ = 0.014, extent = 9) as well as the bilateral MPFC (peak left: $x$, $y$, $z$ = 0, 56, 4, $t_{35}$ = 4.45, $p_{FWE}$ = 0.012, extent = 56; peak right: $x$, $y$, $z$ = 4, 49, 0, $t_{35}$ = 4.52, $p_{FWE}$ = 0.008, extent = 60). That is, with growing drinking severity, these areas showed increasing activations in failed compared with successful self-control in alcohol compared with nonalcohol trials (Fig. 3, Fig. 4, Fig. 5, Fig. 6).

**Hypoactivated areas during pro-alcohol decisions**

According to the hypothesis of impaired control, the

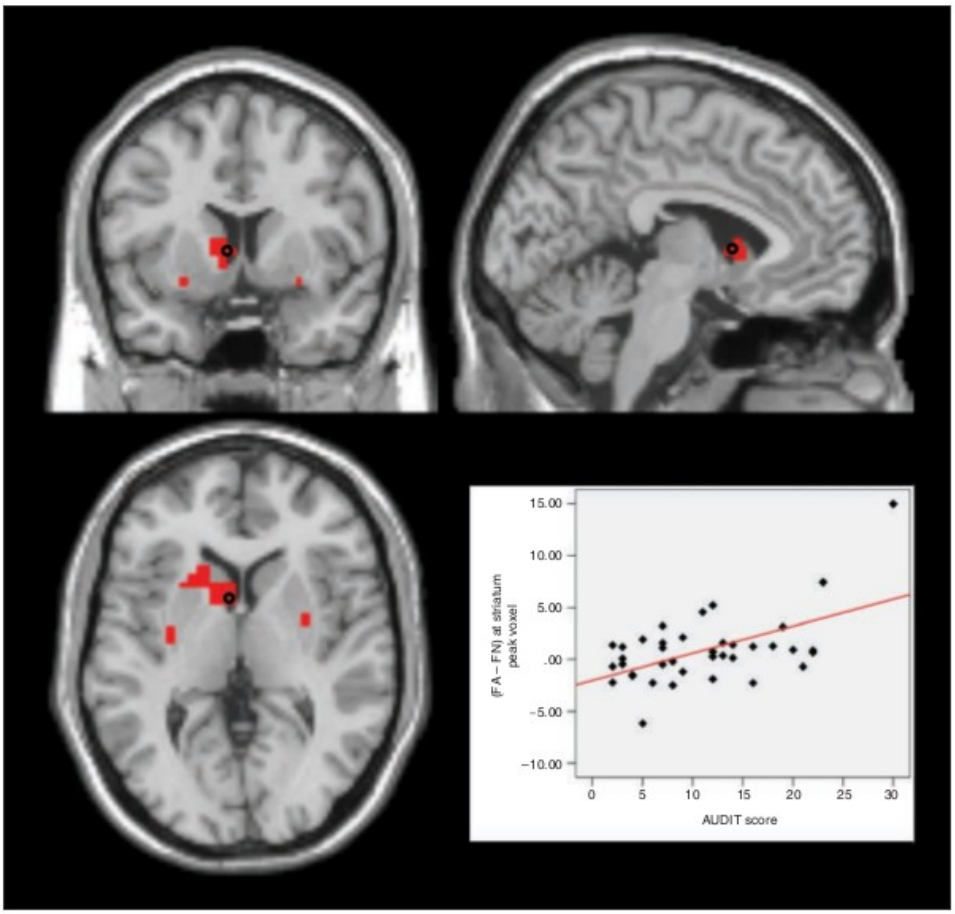


**Fig. 5:** Failed self-control in alcohol trials — striatum. Section views showing significant clusters for AUDIT × (FA – FN) within the striatum at a threshold of $p < 0.05$, family-wise error–corrected. Clusters are presented at a threshold of $p < 0.005$, uncorrected. Red: striatal clusters with higher activity in participants with higher drinking severity in failed self-control in alcohol compared with nonalcohol trials (AUDIT × [FA – FN]).Plot: effect of contrast (FA – FN) at the marked peak voxel (Montreal Neurological Institute space: $x$, $y$, $z$ = –4, 7, 4) plotted subject-wise against AUDIT score. AUDIT = Alcohol Use Disorders Identification Test; FA = failed self-control in an alcohol–nonalcohol conflict; FN = failed self-control in a nonalcohol–nonalcohol conflict.

control-associated areas should show attenuated activation during pro-alcohol decisions, and the activation of these areas should further decrease with growing drinking severity.

The analysis testing negative correlations between drinking severity and brain activation during pro-alcohol decisions (i.e., –AUDIT × [FA – FN]) revealed no significant results, even after lowering the significance threshold to $p < 0.001$, uncorrected. Likewise, the more specific contrast –AUDIT × [(FA – FN) – (SA – SN)] revealed no significant results, even after lowering the threshold to $p < 0.001$, uncorrected. That is, there were no areas showing decreasing activations with growing drinking severity in failed compared with successful self-control in alcohol compared with nonalcohol trials.

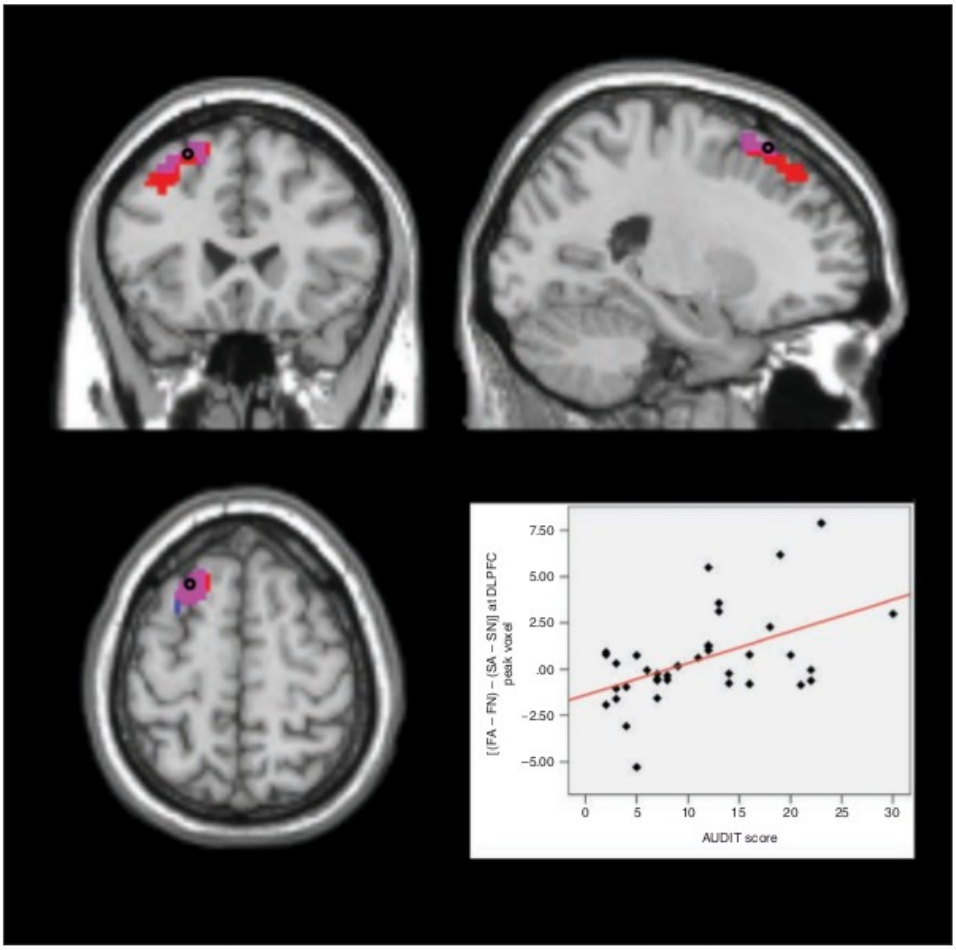


**Fig. 6:** Failed self-control in alcohol trials — dorsolateral prefrontal cortex (DLPFC). Section views showing significant clusters for AUDIT × (FA – FN) and AUDIT × [(FA – FN) – (SA – SN)] within the DLPFC at threshold of $p < 0.05$, family-wise error–corrected. Clusters are presented at a threshold of $p < 0.005$, uncorrected. Red: DLPFC cluster with higher activity in participants with higher drinking severity in failed self-control in alcohol compared with nonalcohol trials (AUDIT × [FA – FN]). Blue: DLPFC cluster with higher activity in participants with higher drinking severity in failed compared with successful self-control in alcohol compared with nonalcohol trials (AUDIT × [(FA – FN) – (SA - SN)]). Violet: overlap between these 2 clusters Plot: effect of interaction contrast (FA – FN) – (SA – SN) at the marked peak voxel (Montreal Neurological Institute space: $x, y, z = –28, 11, 63$) plotted subject-wise against AUDIT score. AUDIT = Alcohol Use Disorders Identification Test; FA = failed self-control in an alcohol–nonalcohol conflict; FN = failed self-control in a nonalcohol–nonalcohol conflict; SA = successful self-control in an alcohol–nonalcohol conflict; SN = successful self-control in a nonalcohol–nonalcohol conflict.

## Discussion

We used fMRI to study the so-called reward and control networks during real-life decisions for and against alcohol. For this purpose, participants with widely differing drinking severity made decisions between more beneficial and more desired alcoholic and nonalcoholic drinks. We found that with increasing drinking severity, participants showed enhanced activations in the bilateral ventral striatum and MPFC as well as in the left amygdala and DLPFC during pro-alcohol decisions (failed self-control in alcohol trials). The specificity of our findings for failed self-control in alcohol trials is documented by the interaction contrast AUDIT × [(FA – FN) – (SA – SN)] that precludes a sole alcohol effect as well as a sole effect of failed self-control. Behaviourally, our fMRI finding was paralleled by an alcohol-related decision bias: with increasing drinking severity, participants failed more frequently and responded significantly faster in alcohol compared with nonalcohol trials.

Earlier studies in individuals with alcohol use disorders have implicated the striatum, amygdala and MPFC in reward processing and have linked activation in these areas to craving and approach behaviour.[7–12,15–17,43] However, to our knowledge, this is the first study to demonstrate that a hyperactivation of these reward-associated areas is associated not only with the development of craving, but also with real decisions in favour of alcoholic drinks.

Besides enhanced responses of the reward system, we hypothesized that areas of the control system would be hypoactive during failed self-control, resulting in pro-alcohol decisions. However, contrary to our hypotheses, we found that these decisions were associated with hyperactivation in the DLPFC, a brain area related to self-control processes.[22-24,26,27,44] This unexpected finding may represent compensatory processes (i.e., enhanced though insufficient self-control efforts in harmful drinkers when confronted with alcohol). This would be in accordance with the clinical observation that individuals with alcohol use disorders tend to choose alcoholic drinks despite their awareness of the risks involved and the intention to quit or reduce drinking. Moreover, similar ineffective hyperactivations of control-associated areas have previously been reported in abstinent alcohol-dependent patients.[45]

In addiction research, there is an ongoing debate on whether harmful decisions for alcohol are due to enhanced responses in reward/motivation areas (overwhelming desire) or to a hypoactive self-control system (impaired control processes).[1,2,46] Our results suggest that decisions for alcohol consumption are linked to a hyperactivation of the reward system (reflected in activations in the striatum, amygdala and MPFC) rather than a hypoactivation of the control system. Notably, we found not only increasing activation of reward areas in pro-alcohol trials with growing drinking severity, but also decreasing activation in nonalcohol trials (Appendix 1). These findings are in line with the "hijacking" hypothesis of the reward system, stating that individuals with addiction show both enhanced responses to addiction-related stimuli and attenuated responses to non–addiction-related rewards.[18] Our findings suggest that both effects may play a role when individuals with harmful drinking behaviour choose between alcoholic and nonalcoholic drinks.

While we refer to the striatum, amygdala and MPFC as reward-associated areas in this article, we acknowledge that for each of these brain areas a variety of distinct psychological functions has been proposed. Although these proposed functions are mostly related to reward-processing, particular functional roles may be considered for each brain region. Specifically, the activation of the ventral striatum has been shown to be related to the occurrence of prediction errors and, therefore, to the guidance of learning processes. Altered activity in the ventral striatum and connectivity with the DLPFC (resulting in altered teaching signals) has been linked to the maintenance of harmful alcohol consumption.[47] Thus, the reported association between drinking severity and activation in the ventral striatum during pro-alcohol decisions may be related to malfunction of prediction error signalling and consequently to altered learning processes.

Our study aimed to transfer the paradigm of Hare and colleagues[24] from decisions between healthier and more desired food items in dieters to the context of (desired but unhealthy) alcohol consumption. Analogous to the study by Hare and colleagues, we distinguished between failed and successful self-control trials. A critical assumption in this type of paradigm is that study participants face a conflict between the desire to consume an attractive but nonbeneficial item and the awareness of the negative consequences of consumption. Because the decision options always consisted of a more desirable and a more beneficial item (as indicated by the participants' individual ratings of the drinks), we believe that participants indeed experienced such conflict in our study; the awareness for nonbeneficial effects of the drinks was reinforced by the prescan ratings that required the participants to reflect on the drinks' harmfulness. Because participants were screened for health considerations during the recruitment for the study and because all participants chose the less desired, more beneficial item in the self-control trials, we assume a general willingness to exert self-control among our study participants. Furthermore, the hyperactivation of the self-control–associated DLPFC indicates enhanced though unsuccessful self-control efforts during decisions for alcohol. In summary, there are good reasons to believe that pro-alcohol decisions in our fMRI study implied reduced self-control. That is, participants chose the desired alcoholic drink, although they were aware of the nonbeneficial effects that the consumption of this particular drink would have on their own health.

### *Limitations*

Participants in our study had to be abstinent at the beginning and would potentially consume an alcoholic drink at the end of the experiment. Because we wanted to avoid inducing withdrawal symptoms or relapse in alcohol-dependent individuals, we did not recruit patients from our department for the study and did not define manifest alcohol dependence as an inclusion criterion. Instead, we focused on less severely affected individuals, assessing drinking severity as a

continuous variable (AUDIT scores). Accordingly, we do not provide categorical comparisons between clinically defined groups (e.g., alcohol-dependent patients v. healthy controls), but rather regression analyses on individuals with a wide range of AUDIT scores. This means our study included indivdiuals showing different severities of alcohol-drinking behaviour, ranging from normal to riskful to abusive to even dependent alcohol consumption. In doing so, we followed current concepts of dependence and abuse that tend to abandon dichotomous classifications (e.g., "addicted" v. "healthy") in favour of a more gradual concept of alcohol use disorders (DSM-5). However, with only 2 participants fulfilling the DSM-IV criteria for alcohol dependence, further research is required to confirm the validity of our results in a larger sample of more severely affected individuals.

Another limitation might be that, especially in small-sized regions of interest like the amygdala and the ventral striatum, we obtained significant results only in a small number of contiguous voxels. Further studies including a larger number of participants might help to also tackle the challenge of achieving larger effect sizes.

Finally, participants were told before they enrolled in the study that there would be urine toxicology tests on a random basis. In practice, this random screening was not performed, and we relied on the participants' self-disclosure instead. Thus, drug consumption among participants cannot completely be excluded.

## Conclusion

Taken together, our data suggest that failed self-control in decisions for alcohol in harmful drinkers is associated with a hyperactive reward system rather than a hypoactive control system. This result is in accordance with clinical findings suggesting that cognitive approaches in psychotherapy attempting to strengthen self-control processes show only moderate effects on relapse rates.[45] The question arises how psychotherapeutic interventions could specifically address the strong automatic, implicit response of the reward system to alcohol-related cues. Cognitive bias modification therapy (CBMT) may represent such a treatment strategy. Recent studies investigated the therapeutic effects of this retraining of automatic approach tendencies and the associated hyperactivation of reward systems. In these studies, CBMT successfully reduced relapse rates 1 year later[5,48] as well as craving-related alcohol cue reactivity in the amygdala.[13] Targeting automatic tendencies rather than control processes may therefore be a promising direction for future therapies in individuals with alcohol use disorders.

**Acknowledgements:** The current project was supported by the German Research Foundation (DFG FOR 16/17 HE2597/14-1 to A. Heinz).

**Affiliations:** From the Department of Psychiatry and Psychotherapy, Charité — Universitätsmedizin Berlin, Berlin, Germany (Stuke, Gutwinski, Wiers, Gröpper, Parnack, Gawron, Attar, Spengler, Walter, Heinz, Bermpohl); the Berlin School of Mind and Brain, Humboldt University Berlin, Berlin, Germany (Gutwinski, Wiers, Walter, Heinz, Bermpohl); and the Neurocomputation and Neuroimaging Unit (NNU), Department of Education and Psychology, Free University Berlin, Berlin, Germany (Schmidt).

**Competing interests:** None declared.

**Contributors:** H. Stuke, S. Gutwinski, C. Wiers, T. Schmidt, H. Walter and F. Bermpohl designed the study. H. Stuke, S. Gröpper, J. Parnack and C. Gawron acquired the data, which H. Stuke, S. Gutwinski, C. Wiers, C. Attar, S. Spengler, A. Heinz and F. Bermpohl analyzed. H. Stuke and F. Bermpohl wrote the article, which all authors reviewed and approved for publication.

## References

1. McClure SM, Bickel WK. A dual-systems perspective on addiction: contributions from neuroimaging and cognitive training. *Ann N Y Acad Sci* 2014;1327:62-78.
2. Noël X, Brevers D, Bechara A. A neurocognitive approach to understanding the neurobiology of addiction. *Curr Opin Neurobiol* 2013;23:632-8.
3. Heinz A, Beck A, Grusser SM, et al. Identifying the neural circuitry of alcohol craving and relapse vulnerability. *Addict Biol* 2009;14:108-18.
4. Schoenmakers T, Wiers R, Field M. Effects of a low dose of alcohol on cognitive biases and craving in heavy drinkers. *Psychopharmacology (Berl)* 2008;197:169-78.
5. Wiers RW, Eberl C, Rinck M, et al. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychol Sci* 2011;22:490-7.
6. Lu L, Hope BT, Dempsey J, et al. Central amygdala ERK signaling pathway is critical to incubation of cocaine craving. *Nat Neurosci* 2005;8:212-9.
7. Beck A, Wustenberg T, Genauck A, et al. Effect of brain structure, brain function, and brain connectivity on relapse in alcohol-dependent patients. *Arch Gen Psychiatry* 2012;69:842-52.
8. Grüsser SM, Wrase J, Klein S, et al. Cue-induced activation of the striatum and medial prefrontal cortex is associated with subsequent relapse in abstinent alcoholics. *Psychopharmacology (Berl)* 2004;175:296-302.
9. Schneider F, Habel U, Wagner M, et al. Subcortical correlates of craving in recently abstinent alcoholic patients. *Am J Psychiatry* 2001;158:1075-83.
10. Vollstädt-Klein S, Loeber S, Kirsch M, et al. Effects of cue-exposure treatment on neural cue reactivity in alcohol dependence: a randomized trial. *Biol Psychiatry* 2011;69:1060-6.
11. Myrick H, Anton RF, Li X, et al. Differential brain activity in alcoholics and social drinkers to alcohol cues: relationship to craving. *Neuropsychopharmacology* 2004;29:393-402.
12. Wrase J, Schlagenhauf F, Kienast T, et al. Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *Neuroimage* 2007;35:787-94.
13. Wiers, CE, Stelzel C, Gladwin TE, et al. Effects of cognitive bias modification training on neural alcohol cue reactivity in alcohol dependence. *Am J Psychiatry* 2015;172:335-43.
14. Myrick H, Anton RF, Li X, et al. Effect of naltrexone and ondansetron on alcohol cue-induced activation of the ventral striatum in alcohol-dependent people. *Arch Gen Psychiatry* 2008;65:466-75.
15. Ihssen N, Cox WM, Wiggett A, et al. Differentiating heavy from light drinkers by neural responses to visual alcohol cues and other motivational stimuli. *Cereb Cortex* 2011;21:1408-15.
16. Claus ED, Ewing SW, Filbey FM, et al. Identifying neurobiological phenotypes associated with alcohol use disorder severity. *Neuropsychopharmacology* 2011;36:2086-96.
17. Filbey FM, Claus E, Audette AR, et al. Exposure to the taste of alcohol elicits activation of the mesocorticolimbic neurocircuitry. *Neuropsychopharmacology* 2008;33:1391-401.
18. Volkow ND, Wang GJ, Fowler JS, et al. Addiction: decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. *BioEssays* 2010;32:748-55.
19. Bechara A, Damasio H. Decision-making and addiction (part I): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia* 2002;40:1675-89.

20. Lane SD, Cherek DR. Analysis of risk taking in adults with a history of high risk behavior. *Drug Alcohol Depend* 2000;60:179-87.
21. McClure SM, Ericson KM, Laibson DI, et al. Time discounting for primary rewards. *J Neurosci* 2007;27:5796-804.
22. Crockett MJ, Braams BR, Clark L, et al. Restricting temptations: neural mechanisms of precommitment. *Neuron* 2013;79:391-401.
23. Figner B, Knoch D, Johnson EJ, et al. Lateral prefrontal cortex and self-control in intertemporal choice. *Nat Neurosci* 2010;13:538-9.
24. Hare TA, Camerer CF, Rangel A. Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 2009;324:646-8.
25. Fellows LK, Farah MJ. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb Cortex* 2005;15:58-63.
26. Salo R, Ursu S, Buonocore MH, et al. Impaired prefrontal cortical function and disrupted adaptive cognitive control in methamphetamine abusers: a functional magnetic resonance imaging study. *Biol Psychiatry* 2009;65:706-9.
27. Bolla K, Ernst M, Kiehl K, et al. Prefrontal cortical dysfunction in abstinent cocaine abusers. *J Neuropsychiatry Clin Neurosci* 2004;16:456-64.
28. Kober H, Mende-Siedlecki P, Kross EF, et al. Prefrontal-striatal pathway underlies cognitive regulation of craving. *Proc Natl Acad Sci U S A* 2010;107:14811-6.
29. Bechara A, Damasio AR, Damasio H, et al. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 1994;50:7-15.
30. Richards JB, Zhang L, Mitchell SH, et al. Delay or probability discounting in a model of impulsive behavior: effect of alcohol. *J Exp Anal Behav* 1999;71:121-43.
31. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview. (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59:22-33.
32. Saunders JB, Aasland OG, Babor TF, et al. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption–II. *Addiction* 1993;88:791-804.
33. Anton RF, Moak DH, Latham P. The Obsessive Compulsive Drinking Scale: a self-rated instrument for the quantification of thoughts about alcohol and drinking behavior. *Alcohol Clin Exp Res* 1995;19:92-9.
34. Skinner HA, Allen BA. Alcohol dependence syndrome: measurement and validation. *J Abnorm Psychol* 1982;91:199-209.
35. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97-113.
36. Climie EA, Rostad K. Test review: Wechsler Adult Intelligence Scale. *J Psychoed Assess* 2011;29:581-6.
37. Spielberger CD. *State-trait anxiety inventory: a comprehensive bibliography*. Consulting Psychologists Press: Palo Alto, USA; 1989.
38. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561-71.
39. Kirby KN, Petry NM, Bickel WK. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *J Exp Psychol Gen* 1999;128:78-87.
40. Skinner HA, Sheu WJ. Reliability of alcohol use indices. The Lifetime Drinking History and the MAST. *J Stud Alcohol* 1982;43:1157-70.
41. Penny WD, Friston KJ, Ashburner JT, et al. *Statistical parametric mapping: the analysis of functional brain images: the analysis of functional brain images*. Academic press; 2011.
42. Shirer WR, Ryali S, Rykhlevskaia E, et al. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex* 2012;22:158-65.
43. Wiers CE, Stelzel C, Park SQ, et al. Neural correlates of alcohol-approach bias in alcohol addiction: The spirit is willing but the flesh is weak for spirits. *Neuropsychopharmacology* 2014;39:688-97.
44. Chanraud S, Martelli C, Delain F, et al. Brain morphometry and cognitive performance in detoxified alcohol-dependents with preserved psychosocial functioning. *Neuropsychopharmacology* 2007;32:429-38.
45. Charlet K, Beck A, Jorde A, et al. Increased neural activity during high working memory load predicts low relapse risk in alcohol dependence. *Addict Biol* 2014;19:402-14.
46. Goldstein RZ, Volkow ND. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci* 2011;12:652-69.
47. Park SQ, Kahnt T, Beck A, et al. Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. *J Neurosci* 2010;30:7749-53.
48. Eberl C, Wiers RW, Pawelczack S, et al. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? *Dev Cogn Neurosci* 2013;4:38-51.

### *2.2 Veränderte Belohnungsprozesse bei Nikotinabhängigkeit*

Wie oben ausgeführt (Abschnitt 1.3), wird eine Veränderung in der Belohnungsverarbeitung als Teil von Abhängigkeitsentwicklung postuliert. Diese Veränderung beinhaltet sowohl eine gesteigerte Reaktivität gegenüber Reizen, die mit der konsumierten Substanz in Verbindung stehen, als auch eine reduzierte Reaktivität gegenüber alternativen (d.h. nicht-substanzbezogenen) Belohnungsreizen. Obwohl diese Zusammenhänge bisher durch einige fMRT-Studien bestätigt wurden (Lin et al., 2020), blieben einige Fragen zum Zusammenhang von Abhängigkeit und Veränderungen der Belohnungsverarbeitung unbeantwortet. So wurden bislang vor allem Raucher*innen ohne Abstinenzwunsch untersucht, was die Relevanz der Ergebnisse für abstinenzmotivierte Raucher*innen in Entwöhnungsprogrammen unklar macht. Darüber hinaus wurden in den bisherigen Studien substanzbezogene oder alternative Belohnungsreize mit neutralen Reizen verglichen oder Raucher*innen mit Nichtraucher*innen ohne neutrale Kontrollbedingung. Dies bedeutet, dass die Interaktion zwischen Raucherstatus und und der Art der Belohnung (substanzbezogen versus alternativ) bislang nicht direkt untersucht wurde. Schließlich wurden meist nicht-individualisierte monetäre Belohnungen als alternative Belohnungskondition verwendet, wodurch die externe Validität eingeschränkt wird. Daher bleiben einige zentrale Fragen dazu, wie sich abstinenzmotivierte Raucher*innen in Bezug auf die Belohnungsverarbeitung von nicht-abhängigen Kontrollproband*innen unterscheiden, noch ungeklärt. Zudem blieb unklar, ob Veränderungen im Belohnungssystem dimensional mit der Schwere der Sucht oder anderen Merkmalen von Raucher*innen zusammenhängen.

Im Rahmen von dual system Theorien wird neben diesen Veränderungen in der Belohnungsverarbeitung eine verringerte kognitive Kontrollfähigkeit gegenüber Belohnungssignalen postuliert, die auf neurophysiologischer Ebene mit defizienter Aktivität im lateralen präfrontalen Kortex in Verbindung gebracht wird und auf der Verhaltensebene mit einer fehlenden Berücksichtigung der negativen Konsequenzen des Konsums einhergeht (Abschnitt 1.3). Diese verminderte Sensibilität für die negativen Konsequenzen des Konsums ist demzufolge nicht nur ein Schlüsselfaktor für die Aufrechterhaltung der Sucht im Allgemeinen, sondern eines ihrer bestimmenden Merkmale (Campbell, 2003). Die Reaktion von an aversiven Reaktionen beteiligten Arealen auf Reize, die die negativen Konsequenzen des Konsums zeigen

(substanzbezogene aversive Reize wie zum Beispiel ein „Raucherbein“) wurde bislang allerdings nicht direkt zwischen abhängigen und nicht-abhängigen Menschen verglichen. Ein solcher Vergleich könnte neurophysiologische Korrelate der beschriebenen Desensibilisierung gegenüber negativen Rauchfolgen bei Nikotinabhängigkeit identifizieren. Um diese Forschungslücken zu schließen, haben wir in dieser Studie die Reaktivität von Belohnungs- und Kontrollsystem auf suchtbezogene und alternative Belohnungsreize, neutrale Reize und suchtbezogene aversive Reize zwischen abstinenzmotivierten Raucher*innen und Kontrollproband*innen verglichen.

Wortgetreu und selbstständig übersetztes Abstract des Originalartikels (Kunas, S L, Stuke, H, Heinz, A, Strohle, A, & Bermpohl, F. Evidence for a hijacked brain reward system but no desensitized threat system in quitting-motivated smokers: An fMRI study. Addiction 2022; 117(3): 701-712. doi:10.1111/add.15651):

*„Hintergrund und Ziele: Mehrere Aspekte der Unterschiede zwischen Personen mit Tabakkonsumstörung (TUD) und Nichtraucher*innen in Bezug auf die Verarbeitung von Belohnungen und Bedrohungen sind noch ungeklärt. Unser Ziel war es, Veränderungen in der Belohnungs- und Bedrohungsverabeitung bei TUD und den Zusammenhang mit Rauchcharakteristika zu untersuchen.*
*Design: Ein Experiment mit funktioneller Magnetresonanztomographie (fMRT) mit between und within subject Faktoren und einem 2 (Gruppen) × 4 (Stimulustyp) faktoriellen Design. Das experimentelle Paradigma umfasste vier Bedingungen: Bilder von (1) Zigaretten dienten als drogenbezogene positive Stimuli, (2) Essen als alternative Belohnungsstimuli, (3) Langzeitfolgen des Rauchens als drogenbezogene aversive Stimuli und (4) neutrale Bilder als Kontrolle.*
*Setting/Teilnehmer*innen: Erwachsene Teilnehmer*innen (n = 38 TUD-Teilnehmer*innen und n = 42 Nie-Raucher*innen) wurden in Berlin, Deutschland, rekrutiert.*
*Messungen: Als Kontraste von primärem Interesse wurden die Interaktionen von Gruppe × Stimulustyp untersucht. Die Signifikanzschwellenkorrektur für multiples Testen wurde mittels family-wise error durchgeführt. Korrelationsanalysen wurden verwendet, um die Assoziation mit Rauchcharakteristika zu testen.*
*Ergebnisse: Die 2 × 2-Interaktion von Raucherstatus und Stimulustyp ergab Aktivierungen im Belohnungssystem des Gehirns auf drogenbezogene positive Reize bei TUD-Proband*innen (between subject effect: P-Werte ≤ 0,036). Als Reaktion auf drogenbezogene aversive Reize*

*zeigten TUD-Proband*innen keine verringerte Aktivierung des aversiven Hirnnetzwerks. Innerhalb der TUD-Gruppe wurde ein signifikanter negativer Zusammenhang zwischen der Reaktion des aversiven Gehirnsystems auf drogenbezogene negative Stimuli (within subject effect: P-Werte ≤ 0,021) und der Anzahl der täglich gerauchten Zigaretten festgestellt (rechte Insula r = -0,386, P = 0,024; linke Insula r = -0,351, P = 0,042; rechter ACC r = -0,359, P = 0,037). Schlussfolgerungen: Moderate Raucher*innen mit Tabakkonsumstörung scheinen im Vergleich zu Nichtraucher*innen eine veränderte Belohnungsverarbeitung drogenbezogener positiver (aber nicht aversiver) Reize im Gehirn zu haben."*

Die Ergebnisse sprechen für die von der incentive salience Theorie postulierten Veränderungen in der Belohnungsverarbeitung bei Raucher*innen (eine stärkere Reaktion auf substanzbezogene Belohnungsreize und eine schwächere Reaktion auf alternative nicht-substanzbezogene Belohnungsreize). Die Ergebnisse sprechen allerdings grundsätzlich nicht für die hypothetisierte reduzierte Reaktion von Raucher*innen auf eine Darstellung der negativen Folgen des Substanzkonsums. Die negative Korrelation zwischen Rauchschwere (Zigaretten pro Tag) und der Reaktion auf  negative Folgen des Substanzkonsums könnte allerdings dafür sprechen, dass es hierbei Subgruppenunterschiede innerhalb der Raucher*innen gibt (in dem Sinne, dass erst bei einer starken Schwere des Konsums die Reaktivität auf seine negativen Folgen tatsächlich reduziert ist).

Received: 23 December 2020 | Accepted: 14 July 2021
DOI: 10.1111/add.15651

RESEARCH REPORT

ADDICTION SSA

# Evidence for a hijacked brain reward system but no desensitized threat system in quitting-motivated smokers: An fMRI study

Stefanie L. Kunas | Heiner Stuke | Andreas Heinz | Andreas Ströhle | Felix Bermpohl

Campus Charité Mitte, Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, Berlin, Germany

**Correspondence**
Stefanie L. Kunas, Campus Charité Mitte, Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Charitéplatz 1, D-10117 Berlin, Germany.
Email: stefanie.kunas@charite.de

## Abstract

**Background and aims:** Several aspects of how quitting-motivated tobacco use disorder (TUD) subjects and never-smokers differ in terms of reward and threat processing remain unresolved. We aimed to examine aberrant reward and threat processes in TUD and the association with smoking characteristics.

**Design:** A between- and within-subjects functional magnetic resonance imaging (fMRI) experiment with a 2 (groups) × 4 (stimulus type) factorial design. The experimental paradigm had four conditions: pictures of (1) cigarettes served as drug-related-positive cues, (2) food as alternative reward cues, (3) long-term consequences of smoking as drug-related-negative cues and (4) neutral pictures as control.

**Setting/participants:** Adult participants ($n = 38$ TUD subjects and $n = 42$ never-smokers) were recruited in Berlin, Germany.

**Measurements:** As contrasts of primary interest, the interactions of group × stimulus-type were assessed. Significance threshold correction for multiple testing was carried out with the family-wise error method. Correlation analyses were used to test the association with smoking characteristics.

**Findings:** The 2 × 2 interaction of smoking status and stimulus type revealed activations in the brain reward system to drug-related-positive cues in TUD subjects (between-subjects effect: $P$-values ≤ 0.036). As a response to drug-related-negative cues, TUD subjects showed no reduced activation of the aversive brain network. Within the TUD group, a significant negative association was found between response of the aversive brain system to drug-related-negative cues (within-subjects effect: $P$-values ≤ 0.021) and the number of cigarettes smoked per day (right insula $r = -0.386$, $P = 0.024$; left insula $r = -0.351$, $P = 0.042$; right ACC $r = -0.359$, $P = 0.037$).

**Conclusions:** Moderate smokers with tobacco use disorder appear to have altered brain reward processing of drug-related-positive (but not negative) cues compared with never smokers.

KEYWORDS
Cue-reactivity, fMRI, quitting motivation, reward processing, threat processing, tobacco use disorder

## INTRODUCTION

Throughout the 20th century, tobacco smoking contributed to the death of approximately 100 million people [1]. It is associated with health consequences such as lung cancer or pulmonary disease [2], which result in premature death in approximately 50% of smokers [3]. This makes smoking the primary cause of preventable deaths [4]. Only approximately 7% of dependent smokers attempting to quit remain abstinent after 12 months [5]. In order to develop new and improve already existing strategies to aid abstinence in quitting-motivated smokers, it is of great importance to understand the mechanisms which underlie tobacco use disorder (TUD) and can provide promising targets for successful smoking cessation interventions.

One previously investigated mechanism involved in the maintenance of TUD is a disruption of reward processing [6,7]. According to incentive–sensitization theory [8,9], addiction and craving develop as a consequence of neuroadaptations induced by repeated consumption of drugs. It is proposed that the mesocorticolimbic brain system, which is involved in the assignment of incentive salience to rewarding stimuli, gradually becomes sensitized to drug-related stimuli and desensitized to non-drug-related alternative rewards [6–9]. Brain structures involved in the cortico–striatal–limbic reward pathway include the amygdala, ventral tegmental area (VTA), hippocampus, ventral pallidum, nucleus accumbens (NAc), medial thalamus and orbitofrontal/medial pre-frontal cortex (mPFC) [10].

The construct of hypersensitivity to drug-associated rewards is supported by several functional magnetic resonance imaging (fMRI) investigations that found heightened activity in mesocorticolimbic areas (e.g. ventral striatum, NAc) in smokers following presentation of drug-related cues compared to healthy controls or neutral cues (e.g. [11–15]). Furthermore, previous studies could demonstrate a reduced activation in smokers as a response to non-drug alternative rewards compared to healthy controls or neutral cues (e.g. [16–21]); for a meta-analysis, see Lin *et al.* [6]. However, many questions still remain unanswered. Non-quitting-motivated smokers were investigated in most studies (e.g. [22,23]), making it more difficult to derive suggestions for smoking cessation programs. In addition, former studies typically compared drug-related or alternative reward cues directly to neutral cues or compared smokers with non-smokers, while not investigating the function of smoking status and reward processing together (e.g. [24,25]). Finally, mainly non-individualized monetary cues were used as alternative rewards (e.g. [26–28]), thus limiting the external validity of studies. Therefore, several aspects of how quitting-motivated TUD subjects and never-smokers differ in terms of reward processing still remain unresolved. Moreover, it is not clear whether such changes are related to the severity of addiction or other characteristics of TUD subjects.

In addition to a potentially 'hijacked' brain reward system, TUD is marked by persistent drug use despite experience or knowledge of its negative consequences. According to Campbell [29], this decreased sensitivity to the negative aspects of consumption is not only a key factor in the maintenance of addiction in general, but one of its defining characteristics. From a theoretical perspective, addiction is marked by a decreased sensitivity to the negative aspects of consumption [30]. Hayes & Northoff [31] identified a core aversion-related brain network associated with the processing of threat stimuli, encompassing cortical and subcortical areas [e.g. amygdala, anterior cingulate cortex (ACC), hippocampus, thalamus, insula, DMPFC, secondary motor cortex]. From a neurofunctional perspective, it can be assumed that this network is also involved in the processing of aversive aspects of drug use and may be altered in subjects suffering from substance use disorders.

To date, only few studies have attempted to elucidate a disruption in the processing of aversive aspects of smoking addiction in regular smokers [32–37]. Dinh-Williams and colleagues [33] showed that non-quitting-motivated chronic smokers display greater activations in regions of the visual association cortex and extended visual system as well as in pre-frontal and limbic brain structures in response to aversive smoking-related images compared to neutral cues. However, they did not include a control group of non-smokers. Therefore, it remains unclear whether quitting-motivated TUD subjects present an aberrant processing of drug-related-negative cues which could constitute an important mechanism underlying the maintenance of TUD.

Summarizing the above, a 'hijacked' reward system and a desensitized aversive system may represent two mechanisms of smoking preservation which are, to date, not sufficiently understood.

To address these issues, we examined quitting-motivated TUD subjects and applied a novel extended cue-reactivity paradigm. The primary aim of this study was to investigate aberrant reward and threat processes in TUD subjects and the association with behavioral smoking characteristics; therefore, we hypothesized that:

1. increased activations elicited by drug-related-positive cues in mesocorticolimbic brain structures in quitting-motivated TUD subjects compared to never-smokers as well as decreased functional activation elicited by alternative rewards;
2. stronger activations in a network characteristic for threat processing in response to drug-related-negative cues (e.g. lung cancer) in never-smokers compared to quitting-motivated TUD subjects; and
3. that heavier and more dependent TUD subjects would show greater activations in mesocorticolimbic brain areas during altered reward processing and a reduced response to drug-related-negative cues in areas related to threat processing.

Additionally, for sensitivity analysis, we investigated general reward and threat processing among both groups, and for the sake of completeness and to replicate findings of previous investigations we examined the effects of the different stimulus types separated for both groups.

## MATERIALS AND METHODS

### Participants

The present study was conducted within the framework of the German Collaborative Research Center (TRR 265: 'Losing and regaining control over drug intake'), funded by the German research foundation (DFG). In total, 82 participants (39 TUD subjects and 43 never-smokers) underwent fMRI scanning. Due to technical issues, 38 TUD subjects (55.26% female) and 42 never-smokers (73.81% female) were included in the present analysis (for a consort flow-chart see Supporting information, Fig. S1). Participants were recruited in Berlin using advertising and flyers. Inclusion criteria for TUD subjects were (a) current DSM-5-TR diagnosis of TUD verified by a structured clinical interview for DSM-5-TR [38]; and (b) aged between 18 and 65 years. Exclusion criteria were (a) comorbid DSM-5-TR mental disorder within the last 12 months; (b) life-time history of any substance-use disorder other than TUD and bipolar or psychotic disorders; (c) current suicidal intent; (d) concurrent psychopharmacological or psychotherapeutic/psychiatric treatment; (e) history of brain injury; and (f) pregnancy. Participants were classified as never-smokers if they had smoked fewer than 10 cigarettes during their life-time. The never-smoker group was free of current or past medical, neurological or mental illness. Healthy controls as well as TUD subjects received financial compensation (€50) for their participation in the study. After the examination all TUD subjects took part in a free, 6-week smoking cessation intervention, as all of them were quitting-motivated. Furthermore, half the participants were randomized to an additional sport intervention. The study was approved by the local ethics committee and all subjects gave written, informed consent prior to participating in the study.

As the primary research question and analysis plan of this study were not pre-registered on a publicly available platform, the results should be considered exploratory.

### Clinical assessments

During the first session, all participants completed the multiple-choice vocabulary test (MWT; range = 0–37) [39] to assess their global level of intelligence, the trait part of the State–Trait Anxiety-Inventory (STAI-T; range = 20–80) [40] and the short version of the General Depression Scale [anxiety and depression scale (ADS-K); range = 0–45] [41]. The Fagerström Test for Nicotine Dependence (FTND; range = 0–10) [42] was used to assess severity of nicotine dependence. Furthermore, information regarding frequency of alcohol use was acquired (drinking days/week). For more details regarding the tests see also Supporting information, Text S1.

### Extended cue-reactivity task

We established a novel extended cue-reactivity task, which was performed during fMRI. We asked participants to abstain from smoking and eating for 3 hours. This duration of abstinence was chosen to ensure a sufficient level of craving for cigarettes, but avoid severe withdrawal in the moderately dependent TUD group at the time of the fMRI scanning. The task was used to study drug-related-positive, drug-related-negative and alternative reward cue-reactivity at the psychological and neural level. The experimental paradigm consisted of four conditions: established photographs displaying cigarette items were used as drug-related-positive cues, pictures of healthy, low-fat, attractive food were used as alternative reward cues, pictures showing long-term consequences of smoking (e.g. bronchial carcinoma) were used as drug-related-negative cues and pictures displaying neutrally valenced items were presented during neutral control conditions. Before the fMRI session, participants rated a set of 144 drug-related-positive, alternative reward and drug-related-negative pictures each. For drug-related-positive and alternative reward pictures, the question 'how strong is your desire to consume this now?' was used. For drug-related-negative cues, the question 'how deterrent do you experience this picture?' was asked, using an eight-point Likert-scale from 'not at all' to 'very much'. For the experiment, the 50% most rewarding/threatening stimuli were automatically selected in order to maximize effects (for an example run and more details regarding the task see Fig. 1, and for more details regarding the selected cues see Supporting information, Text S2).

### Statistical analysis of behavioral data

To examine differences in drug and food craving ratings (as well as in threat ratings) before and within the task the Scheirer–Ray–Hare test was used, as specific assumptions for an analysis of variance were violated (see Supporting information, Text S2). To specify the direction of the effects, we used Mann–Whitney *U*-tests. For analysis of the differences in craving ratings between the two groups at the end of each run, we also used the Mann–Whitney *U*-test. A paired *t*-test was conducted to quantify the impact of drug-related-negative cues on subjective desire for cigarettes; therefore, we calculated the difference between craving ratings at the end of each run when preceded by drug-related-negative cues in comparison to the other categories for the TUD group. Furthermore, the difference between alternative reward and drug-related-positive cues in TUD subjects was examined using a paired *t*-test (see also Supporting information, Text S2).

### fMRI data acquisition and pre-processing

The study was conducted with a 3-Tesla Siemens Magnetom Prisma scanner. Functional images were acquired using T2-weighted

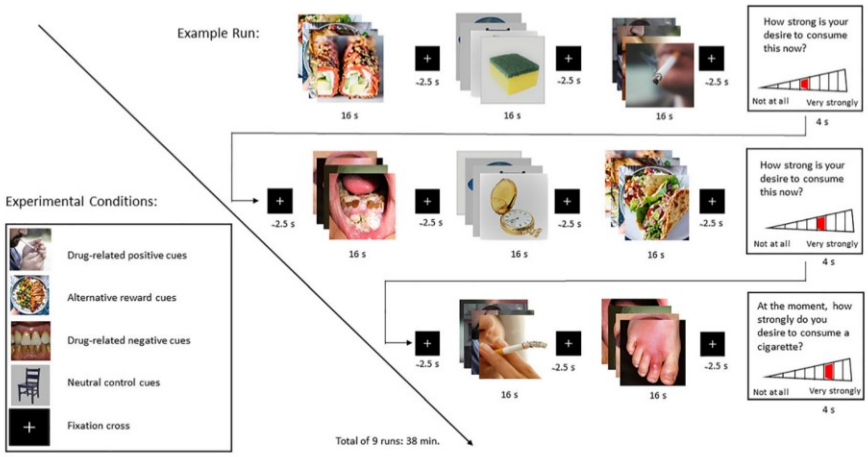


**FIGURE 1** Four pictures of one category were presented per block. Each block lasted 16 seconds and ended with the presentation of a fixation cross [intertrial interval (ITI)], jittered around 2.5 seconds. In each run, two blocks of each of the four categories were presented. Subjects were instructed to attend to all stimuli and to rate their current desire to consume some of these items (cigarette or food) twice per run, by pressing one of eight buttons covering an eight-point scale ranging from 'not at all' to 'very strongly'. At the end of each run, participants were additionally asked to rate how strongly they desire to smoke a cigarette, using the same rating scale as described above. In total, the task consisted of nine runs, which altogether lasted 38 minutes

gradient-echo echoplanar imaging (TR 869 ms, TE 38 ms, voxel size 2.4 × 2.4 × 2.4 mm) and anatomical images were acquired using a T1-weighted MPRAGE sequence (voxel size 1 × 1 × 1 mm) using a 64-channel head coil. To minimize movement artifacts, participants' heads were positioned on a pillow and fixed using foam pads surrounding the head. Image pre-processing was performed using statistical parametric mapping (SPM12; http://www.fil.ion.ucl.ac.uk/spm/software/spm12) and MATLAB R2020a (Mathworks, Sherborn, MA, USA)-based scripts and comprised slice timing with reference to the middle slice, SPM12 standard re-alignment and unwarping including correction for field deformations based on a previously acquired field map, co-registration, normalization to Montreal Neurological Institute (MNI) stereotactic space using unified segmentation based on the SPM tissue probability map for six tissue classes and spatial smoothing with 8-mm full-width at half-maximum isotropic Gaussian kernel. Following pre-processing, all nine runs were visually inspected, for each subject separately, in order to perform a visual quality control.

## fMRI data analysis pathway

First-level analysis was carried out as described in Supporting information, Text S3. For second-level analysis, group effects were assessed by a 2 × 4 analysis of covariance (ANCOVA) using a full factorial model in SPM12, encompassing the factors 'group' (TUD subjects and never-smokers) and 'stimulus-type' (neutral, alternative reward, drug-related-positive and drug-related-negative). The continuous variables of the STAI-T and ADS-K were included as covariates of no interest because of significant group differences in these measures (see below). In accordance with our hypotheses, we tested the group × stimulus interactions (drug-related-positive versus alternative reward cues and drug-related-negative versus neutral cues). For sensitivity analyses, we investigated the effect of reward processing across both groups and stimulus types (drug-related-positive and alternative reward > neutral) as well as the effect of threat responsivity across both groups (threat > neutral). For the sake of completeness and to replicate findings of previous investigations, we report the effects of drug-related-positive, alternative reward and drug-related-negative cues contrasted to neutral cues within each of the two groups separately. *F*-contrasts were computed followed by post-hoc *t*-contrasts to specify the direction of the effects.

## Region of interest (ROI) and whole brain analysis

As small subcortical brain regions (e.g. VTA) are difficult to investigate using a whole brain approach, an anatomical ROI analysis of a priori-defined subcortical brain areas was conducted. Based on the literature of brain regions involved in reward processing [10,43], the reward system was defined to include NAc, amygdalae, hippocampi, thalamus, pallidum and mid-brain (including VTA) for hypothesis 1. Based on the model proposed by Hayes & Northof [31], the core aversive system was defined to include amygdalae, hippocampi, insulae, ACC and thalamus for hypothesis 2. The a priori-defined anatomical regions of interest were built combining definitions from the Automated Anatomical Labeling Atlas [44], implemented in the Wake Forest

University PickAtlas [45]. The bilateral ROIs were investigated using one single mask. Small volume correction on this single mask was applied using a family-wise error (fwe)-corrected threshold of $P_{fwe} < 0.05$ with a minimum cluster size of $k = 10$ contiguous voxels. ROI analyses were followed by whole brain analyses. To correct for multiple comparisons on a whole brain level, group-level results were thresholded at $P < 0.05$ fwe-corrected.

### Correlation analysis

To evaluate the relationship of altered reward processing and threat responsivity with dimensional measures of nicotine addiction in TUD subjects only, Pearson's correlations were calculated between extracted beta-values of significantly activated brain regions identified in the ROI analysis for the two contrasts of interest [TUD subjects: (drug > alt) and (threat > neutral)] with the FTND (as a dimensional measure of nicotine dependence), cigarettes smoked per day (implying that heavy smokers consume a higher number of cigarettes per day) and pack-years (calculated as the product of smoking amount and time). Moreover, the difference between craving ratings at the end of each run when preceded by other categories in comparison to drug-related-negative cues (mean craving rating after other cues minus mean craving rating after drug-related-negative cues) were correlated with the FTND, cigarettes smoked per day and pack-years. The toolbox marsbar (http://marsbar.sourceforge.net) was used in SPM12 to extract the beta-weights using a sphere of 5 mm around the peak voxel of significant ROIs (see Tables 2 and 3 for MNI coordinates). Age served as covariate in the partial correlation analysis of pack-years (see also Supporting information, Text S2 for more details).

## RESULTS

### Sample characteristics

Demographic data and smoking characteristics are shown in Table 1. TUD subjects were moderately nicotine-dependent, as evidenced by FTND and average cigarettes smoked per day. STAI-T and ADS-K scores were, although subclinical in both groups, significantly higher in TUD subjects, which is in line with results from previous investigations [46].

### Subjective craving ratings

Within-task craving ratings showed the expected main effect of group ($H_{(1/156)} = 33.115$, $P < 0.001$, $\eta^2 = 0.179$), stimulus ($H_{(1/156)} = 128.579$, $P < 0.001$, $\eta^2 = 0.458$) and group × stimulus interaction ($H_{(1/156)} = 27.851$, $P < 0.001$; $\eta^2 = 0.155$); see also Table 1 and Supporting information, Fig. S3. In the group of TUD subjects, final craving ratings were significantly lower when drug-related-negative cues compared to other cues preceded the rating ($t_{(36)} = -6.09$, $P < 0.001$, $d = 1.348$). Within the task, TUD subjects rated their craving for food significantly higher compared to cigarettes ($t_{(35)} = -5.453$, $P < 0.001$, $d = 1.263$), and at the end of each run they rated a medium desire to smoke a cigarette now (for ratings conducted before the fMRI session see Supporting information, Fig. S2).

### fMRI results

#### Altered reward processing

The 2 × 2 interaction of smoking status and stimulus type [TUD subjects (drug > alt) > never-smokers (drug > alt)] revealed stronger activation in the bilateral hippocampi and thalamus as well as in the left mid-brain (including VTA) in TUD subjects regarding drug-related-positive cues in the ROI analysis (Table 2, Figure. 2). On a whole brain level, the OFC was significantly activated.

#### Altered threat responsivity

The 2 × 2 interaction of smoking status and stimulus type [never-smokers (threat > neutral) > TUD subjects (threat > neutral)] reached no significant results, neither on a ROI nor on a whole brain level (Table 2).

### Sensitivity analysis

The processing of reward in general (drug-related-positive and alternative reward) against neutral cues elicited brain activation in the bilateral thalamus, hippocampi, mid-brain (including VTA) and pallidum in the ROI analysis (Supporting information, Table S1) among both groups. On a whole brain level, frontal, parietal, temporal occipital as well as subcortical brain areas (ACC) were activated (Supporting information, Text S4). The effect of threat responsivity reached significant activation in the bilateral insulae, hippocampi, thalamus and in the right ACC in the ROI analysis (Supporting information, Table S1) among both groups. On a whole brain level, frontal, parietal, temporal and occipital brain regions were activated (Supporting information, Text S4 and Supporting information, Table S1).

Investigating the two groups separately regarding threat responsivity, TUD subjects showed significant activation in structures belonging to the aversive brain system (bilateral insulae, right ACC and hippocampus). On a whole brain level, TUD subjects showed significantly activated brain regions in the lingual and occipital gyrus, temporal gyrus, inferior frontal gyrus, superior and inferior parietal gyrus and in the right insula and ACC (Table 3 and Fig. 3). Conversely, never-smokers showed no significant brain activation in the ROI analysis. On a whole brain level, significantly activated brain regions in the occipital and parietal cortex could be observed (Table 3).

**TABLE 1** Socio-demographic and psychometric characteristics of the smoker and never-smoker sample.

| Sample characteristic | TUD subjects $n = 38$ | Never-smokers $n = 42$ | Statistic | P |
|---|---|---|---|---|
| Age (mean, SD) | 35.18 (10.57) | 32.36 (10.97) | $t_{(78)} = -1.17$ | 0.245 |
| Female gender (*n*, %) | 21 (55.26) | 31 (73.81) | $\chi^2_{(1)} = 3.016$ | 0.090 |
| Right-handedness (*n*, %) | 38(100) | 39(92.86) | $\chi^2_{(1)} = 2.747$ | 0.100 |
| Level of education | | | | |
| A-level[a] (*n*, %) | 30 (78.95) | 35 (83.33) | $\chi^2_{(1)} = 0.252$ | 0.616 |
| Monthly income in € (*n*, %) | | | | |
| < 1000 | 7 (18.42) | 16 (38.10) | $\chi^2_{(4)} = 6.401$ | 0.171 |
| 1000–2000 | 12 (31.58) | 12 (28.57) | | |
| 2000–3500 | 16 (42.11) | 11 (26.19) | | |
| 3500–4500 | 2 (5.26) | – | | |
| > 4500 | 1 (2.63) | 1 (2.38) | | |
| MWT (mean, SD) | 28.32 (4.67) | 29.38 (3.26) | $t_{(78)} = 1.192$ | 0.237 |
| Craving ratings (median, IQR) | | | | |
| Alternative reward | 5.66 (1.60) | 5.36 (2.95) | $U_{(78)} = 0.130$ | 0.896 |
| Cigarette cues | 3.56 (2.22) | 1.00 (0.28) | $U_{(78)} = 7.387$ | < 0.001** |
| Final craving rating | 5.00 (3.50) | 1.00 (0.00) | $U_{(80)} = 7.557$ | < 0.001** |
| Drinking days (per week) (mean, SD) | 1.89 (1.13) | 1.26 (0.87) | $t_{(78)} = -2.64$ | 0.012* |
| STAI-T (mean, SD) | 38.90 (7.42) | 31.09 (6.33) | $t_{(78)} = -5.02$ | < 0.001** |
| ADS-K (mean, SD) | 7.97 (5.05) | 4.12 (2.87) | $t_{(78)} = -4.09$ | < 0.001** |
| FTND (mean, SD) | 4.03 (2.27) | | | |
| Pack-years (mean, SD) | 10.75 (9.55) | | | |
| Cigarettes/day (mean, SD) | 14.40 (6.05) | | | |

Abreviations: ADS-K = general depression scale; FTND = Fagerström Test for Nicotine Dependence. Missing values: monthly income: 1; FTND: 2; cigarettes/day: 2; craving ratings for alternative reward and cigarette cues 2; missing values were treated with listwise deletion. For the Mann–Whitney *U*-test we report the standardized test statistic; IQR = interquartile range; MWT = Mehrfachwahl–Wortschatz test (identification test); SD = standard deviation; STAI-T = trait part of the State–Trait Anxiety Inventory.
*$P < 0.05$; **$P < 0.001$.
[a]Abitur.

Regarding the effects of drug-related-positive cues and alternative rewards, separated for the two groups, please refer to Supporting information, Text S4 and Supporting information, Tables S3 and S4.

## Correlation analysis

No significant correlations between activated brain regions and dimensional measures of smoking behavior could be obtained for altered reward processing in TUD subjects (Supporting information, Table S2).

Regarding threat processing, correlation analyses revealed significant negative associations between the number of cigarettes smoked per day and extracted beta weights of the left ($r = -0.351$; $P = 0.042$) and right insula cortex ($r = -0.386$; $P = 0.024$) and ACC ($r = -0.359$; $P = 0.037$) in TUD subjects, showing that brain activations of heavy smokers were less influenced by aversive drug cues. The difference between craving ratings when preceded by other cues minus drug-related-negative cues was significantly and negatively correlated with the number of cigarettes smoked per day ($r = -0.319$; $P = 0.040$) and FTND scores ($r = -0.412$; $P = 0.017$; see also Supporting information, Table S5). Heavy and more dependent smokers exhibited a lower difference between craving ratings, implying a lower impact of drug-related-negative cues on craving. No significant correlations between pack-years and functional activation of brain regions/differences in craving ratings could be observed.

## DISCUSSION

In the present study we found evidence for a 'hijacked' brain reward system in TUD subjects, as they presented an increased functional activation of mesocorticolimbic brain areas elicited by drug-related-positive versus alternative reward cues when compared to never-smokers. We did not observe a reduced activation of the so-called aversive brain network during the processing of drug-related-negative cues in TUD subjects compared to never-smokers. However, within the TUD group, limbic brain structures belonging to the core aversive

**TABLE 2** Locations of significantly activated brain regions during processing of cigarette cues compared to alternative rewards in TUD subjects versus never-smokers (a); results of the interaction contrast of drug-related-negative cues versus neutral cues in TUD subjects versus never-smokers (b).

| Contrast/region | Side | Voxels | x | y | z | *F* or *t* | *P* < 0.05 fwe-corrected |
|---|---|---|---|---|---|---|---|
| a. Altered brain reward processing | | | | | | | |
| F-contrast | | | | | | | |
| Interaction TUD versus NS (drug versus alt) | | | | | | | |
| Region of interest analysis | | | | | | | |
| Hippocampus | L | 37 | −20 | −18 | −16 | 16.93 | 0.020 |
| Mid-brain (incl. VTA) | L | 34 | −2 | −18 | −8 | 16.01 | 0.022 |
| Thalamus | L | 17 | −4 | −16 | −2 | 15.10 | 0.035 |
| Hippocampus | R | 12 | 24 | −38 | −2 | 13.52 | 0.045 |
| Thalamus | R | 10 | 8 | −14 | −2 | 12.53 | 0.043 |
| Whole brain analysis | | | | | | | |
| Orbitofrontal cortex | R | 18 | 0 | 42 | −8 | 25.00 | 0.016 |
| Post-hoc *t*-contrast | | | | | | | |
| TUD (drug > alt) > NS (drug > alt) | | | | | | | |
| Region of interest analysis | | | | | | | |
| Hippocampus | L | 91 | −20 | −18 | −16 | 4.02 | 0.018 |
| Mid-brain (incl. VTA) | L | 42 | −2 | −18 | −8 | 4.00 | 0.027 |
| Thalamus | L | 139 | −4 | −16 | −2 | 3.92 | 0.036 |
| Hippocampus | R | 39 | 24 | −38 | −2 | 3.94 | 0.033 |
| Thalamus | R | 12 | 8 | −14 | −2 | 3.54 | 0.016 |
| Whole brain analysis | | | | | | | |
| Orbitofrontal cortex | R | 69 | 2 | 42 | −8 | 5.00 | 0.006 |
| a. Altered threat responsivity | | | | | | | |
| F-contrast | | | | | | | |
| Interaction TUD versus NS (threat versus neutral) | | | | | | | |
| ROI analysis | | | | No differential activation | | | |
| Whole brain analysis | | | | No differential activation | | | |

*Note:* L: left; R: right; voxels: number of voxels per cluster; x, y, z: MNI coordinates; TUD: tobacco use disorder subjects, NS: never-smokers; drug: drug-related-positive cues; alt: alternative rewards; threat: drug-related-negative cues; fwe = family-wise error; VTA = ventral tegmental area.
*P* < 0.05 fwe-corrected: for region of interest (ROI) analyses a family-wise error-corrected threshold of $P_{fwe} < 0.05$ with k > 10 voxels on a peak level was used. For whole-brain analyses a threshold of $P_{fwe} < 0.05$ was applied.

network were activated during the presentation of drug-related-negative cues, and this activation was negatively correlated with the number of cigarettes smoked per day.

An important component of the mesocorticolimbic brain system is dopaminergic projections from the VTA and related brain stem areas to subcortical (e.g. NAc, thalamus, hippocampus) and pre-frontal brain regions (e.g. OFC), as these pathways appear to be critical in drug-induced reward processing [47–50]. According to the incentive–sensitization theory [8], it can be assumed that brain areas belonging to the reward pathway of TUD subjects gradually became sensitized to tobacco cues and desensitized to alternative reward cues as, in our case, food cues. The main functional activation effects of drug-related-positive cues in TUD subjects are in accordance with results of previous studies [16,19,21]. However, to the best of our knowledge, previous studies did not directly investigate the interaction effect of smoking status and (food) reward processing. For the first time we observed alterations in processing of drug cues versus alternative reward cues related to smoking status and can therefore draw the conclusion of a 'hijacked' brain reward system in TUD. Contrary to our expectations, functional activation of the mesocorticolimbic reward system was not related to the number of cigarettes smoked per day, FTND or pack-years. Further studies need to assess alterations of reward processing over time during the development and maintenance of TUD. Results for subjective craving ratings were inconsistent with the fMRI findings, as TUD subjects rated their craving for food cues significantly higher than for drug-related-positive cues within the task. This phenomenon can be possibly explained by the abstinence motivation of TUD subjects included in

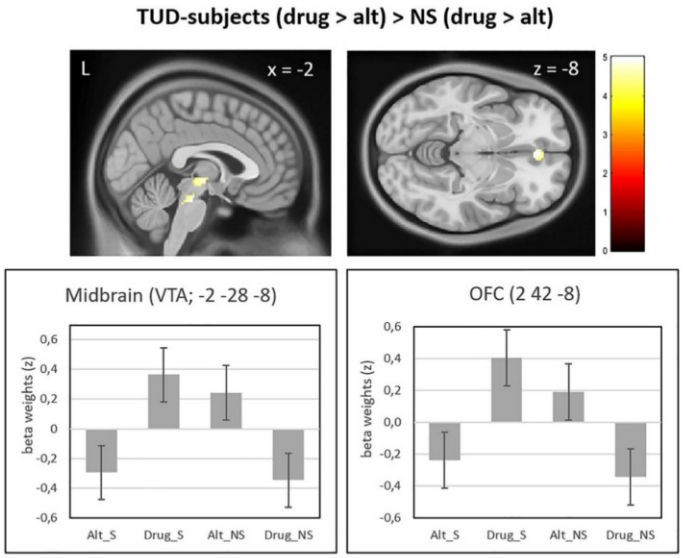


**FIGURE 2** Neural correlates of altered reward processing for drug-related-positive cues in tobacco use disorder (TUD) subjects compared to alternative rewards and never-smokers identified in the region of interest (ROI) (VTA) and whole brain analyses (OFC). VTA = ventral tegmental area; OFC = orbitofrontal cortex; drug = drug-related-positive cues; alt = alternative reward. Significance threshold is $P < 0.05$ family-wise error (fwe)-corrected. Bars represent the estimated, standardized beta values of the corresponding brain region. Error bars represent the standard error of the mean

our study. It may be that the motivation to quit impacts upon the craving ratings given during the task. However, craving ratings are often not predictive of smoking behaviors, which may have a strong habitual component [51].

The second mechanism under examination was functional activation elicited by drug-related-negative cues in a brain network associated with threat processing. We could not confirm our hypothesis of altered processing of drug-related-negative stimuli in TUD subjects compared to never-smokers, pointing to the fact that there may be no general desensitization of the aversive system in TUD. However, investigating the TUD group separately revealed activation of threat-related brain regions in response to drug-related-negative stimuli, consistent with findings of previous studies [32–36], which was not observed in healthy controls. These findings suggest that when quitting-motivated TUD subjects are exposed to the negative value of smoking these cues can, to some degree, be processed as unpleasant and engage structures associated with negative emotions [52,53], even though no significant group difference with never-smokers was found. This result is complemented by our behavioral finding that prior presentation of drug-related-negative cues reduced subjective cigarette craving.

Importantly, the observed activation of the aversive brain network is driven by light smokers, as we observed a negative correlation of medium effect size between brain activation with the number of cigarettes smoked per day. This result suggests that, relative to light smokers, heavy smokers present a desensitization of the aversive brain network; i.e. they are no longer responding strongly to the negative aspects of smoking. Interestingly, this finding was, again, paralleled by behavioral analyses of craving ratings preceded by drug related-negative cues versus other cues. Here, a negative association of medium effect size between the difference in final craving ratings and smoking behavior suggests that heavy and more dependent smokers were less influenced in their craving when drug-related-negative cues were presented beforehand. An alternative explanation for this association could be that TUD subjects who are less sensitive to drug-related-negative cues tend to consume more cigarettes per day. Thus, as our study is limited by its cross-sectional design, longitudinal studies are needed to assess the role of threat processing as a potential marker of vulnerability for smoking onset and maintenance.

When defining the ROIs for the present analysis we chose, based on the literature, to assign some regions to both the brain reward and aversive system (e.g. amygdala and hippocampus). Both reward- and threat-related stimuli are highly emotional, a fact that might be reflected in the common activation of brain regions [31]. In addition, some areas code for multiple, even apparently opponent processes (e.g. aversion and reward). There are numerous cell types with various response characteristics (e.g. throughout the amygdala) which may respond to the presence of rewarding, aversive or both types of stimuli [31]. Conversely, some areas previously also found during processing of drug-related-positive cues (e.g. insula and ACC) were assigned to the aversive system ROI, but not included in the ROI of the reward system. This approach was chosen because previous research linked these brain regions not primarily to reward processes, but rather control, conflict and interoceptive processes during the processing of drug-related-positive cues [53,54].

Pharmacotherapy, such as nicotine substitution and bupropion, have been proved to be effective mainly in reducing withdrawal symptoms experienced during cessation [44]. Additionally, non-pharmacological treatment is important to address the full spectrum of neurobiological mechanisms that underlie TUD. In this context, our results are clinically relevant as they offer important starting-points for interventions. It can be suggested that smoking cessation

**TABLE 3** Significantly activated brain regions contrasting drug-related-negative cues against neutral cues in TUD subjects (a) and never-smokers (b) separately.

| Contrast/region | Side | Voxels | x | y | z | t | *P* < 0.05 fwe-corrected |
|---|---|---|---|---|---|---|---|
| (a) Effect of threat responsivity in TUD subjects | | | | | | | |
| Post-hoc *t*-contrast | | | | | | | |
| TUD (threat > neutral) | | | | | | | |
| ROI analysis | | | | | | | |
| Insula | R | 331 | 40 | 8 | −12 | 4.87 | < 0.001 |
| ACC | R | 542 | 0 | 14 | 26 | 4.77 | 0.002 |
| Hippocampus | R | 46 | 26 | −40 | −2 | 4.49 | 0.005 |
| Insula | L | 276 | −42 | 8 | −10 | 4.47 | 0.006 |
| Whole brain analysis | | | | | | | |
| Lingual gyrus | R | 746 | 16 | −82 | −8 | 8.70 | < 0.001 |
| Lingual gyrus | L | 528 | −24 | −76 | −6 | 7.61 | < 0.001 |
| Middle occipital gyrus | L | 170 | −28 | −78 | 20 | 6.27 | < 0.001 |
| Middle occipital gyrus | R | 116 | 32 | −74 | 22 | 5.78 | < 0.001 |
| Middle temporal gyrus | R | 164 | 50 | −66 | 8 | 5.35 | < 0.001 |
| Inferior frontal gyrus | R | 61 | 46 | 6 | 24 | 5.33 | < 0.001 |
| Middle temporal gyrus | L | 102 | −48 | −80 | 10 | 5.27 | < 0.001 |
| Superior parietal gyrus | L | 100 | −26 | −50 | 56 | 5.26 | < 0.001 |
| Inferior parietal gyrus | R | 41 | 28 | −46 | 48 | 5.23 | 0.012 |
| Supramarginal gyrus | R | 20 | 58 | −22 | 36 | 4.91 | 0.014 |
| Insula | R | 15 | 40 | 8 | −12 | 4.87 | 0.017 |
| ACC | R | 15 | 0 | 14 | 26 | 4.82 | 0.021 |
| (b) Effect of threat responsivity in NS | | | | | | | |
| Post-hoc *t*-contrast | | | | | | | |
| NS (threat > neutral) | | | | | | | |
| Whole brain analysis | | | | | | | |
| Lingual gyrus | R | 825 | 16 | −82 | −6 | 8.99 | < 0.001 |
| Lingual gyrus | L | 435 | −13 | −88 | −6 | 6.53 | < 0.001 |
| Middle occipital gyrus | L | 129 | −48 | −78 | 22 | 6.23 | < 0.001 |
| Middle occipital gyrus | R | 54 | 30 | −76 | 22 | 5.16 | 0.004 |
| Inferior parietal gyrus | L | 26 | −28 | −52 | 52 | 4.83 | 0.016 |

*Note*: L: left; R: right; voxels: number of voxels per cluster; x, y, z: MNI coordinates; TUD: tobacco use disorder subjects, NS: never-smokers; threat: drug-related-negative cues; fwe = family-wise error; ACC = anterior cingulate cortex.
*P* < 0.05 fwe-corrected: for region of interest (ROI) analyses a family-wise error-corrected threshold of $P_{fwe} < 0.05$ with k > 10 voxels on a peak level was used. For whole-brain analyses a threshold of $P_{fwe} < 0.05$ was applied.

treatment should address strategies to enhance the meaning and processing of alternative rewards as, for instance, intended by psycho-education and enjoyment training, in all stages of TUD. Cognitive–behavioral therapy (CBT) approaches could include an individualized training session to identify and activate alternative rewards in the treatment of quitting-motivated TUD subjects. Conversely, the confrontation with long-term consequences of chronic smoking behavior seems to be more efficient for light smokers. Additionally, our results can be used to inform alternative and novel intervention strategies targeting the brain, such as repetitive transcranial magnetic stimulation, deep brain stimulation and real-time fMRI neurofeedback. Such approaches represent potentially useful and clinically meaningful treatment modalities for TUD [55,56], but further research is needed to detect involved brain regions and conditions. Our results can suggest new target regions for such interventions (e.g. OFC and hippocampus) as well as applying new strategies (e.g. enhancing alternative reward processing).

Future studies should investigate the role of aversive processing to inform health advertisement campaigns as well as the use of long-term negative consequences in smoking cessation therapy. Therefore,

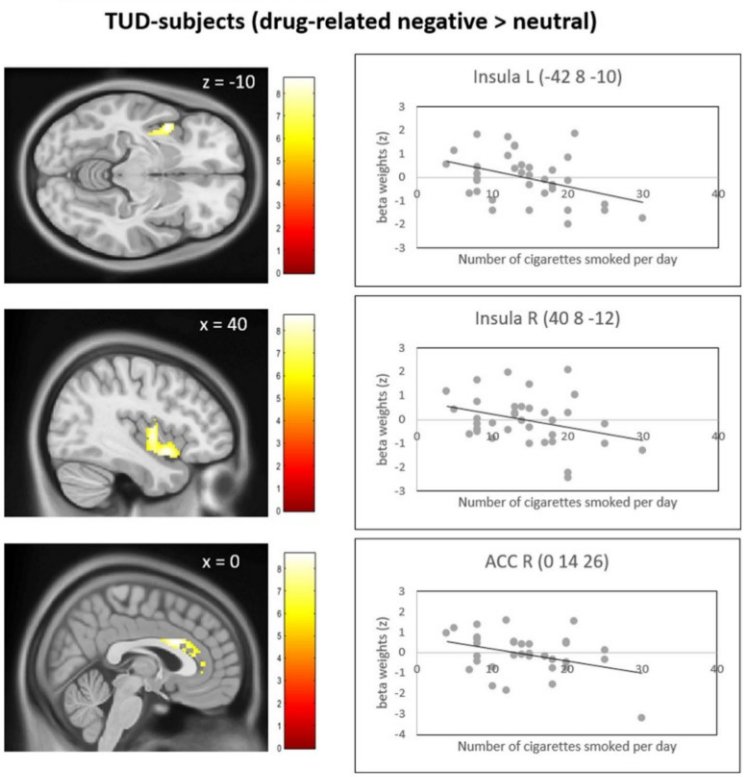


**FIGURE 3** Neural correlates of altered threat processing compared to neutral cues in tobacco use disorder (TUD) subjects only (left side). Scatterplots show the negative relationship between extracted beta weights and the number of cigarettes smoked per day (right side). R = right; L = left; Significance threshold is $P < 0.05$ family-wise error (fwe)-corrected. Dots represent the estimated, Z-standardized beta values of the corresponding brain region

it may be important to specify the effects of threat processing; for example, by including a group of heavy and light smokers who are not motivated to quit and by the use of different categories of drug-related-negative cues (e.g. pictures used on cigarette packets versus unknown pictures). Furthermore, it would be of interest to investigate whether activations in the identified brain regions predict behavioral measures, such as the ability to resist craving or successful smoking cessation.

## Strengths and limitations

Complementing previous investigations on smoking cue-reactivity, we recruited a well-defined sample of TUD subjects who were dependent according to DSM-5-TR criteria and motivated to quit. Participants in our study were free of psychotropic medication and did not suffer from any mental disorder that could influence the processing of the applied cue categories. The groups were well matched for gender, age, handedness, education and income. Cues presented during the task were selected by the participants beforehand to match their individual preferences. Of special note, results are corrected for differences in subclinical trait anxiety and current depression scores, which could represent a potential bias in neurofunctional processes. To the best of our knowledge, this is the first study investigating altered reward and threat processing within one paradigm using individualized pictures. This offers the opportunity to examine interaction effects between TUD subjects and never-smokers to gain a clearer understanding of two promising mechanisms of smoking initiation and preservation which can be used to modify and improve treatment and prevention strategies.

However, several limitations must also be considered. As our study has a cross-sectional design, we cannot infer causal interactions between smoking behavior and neurofunctional findings. It may be possible that TUD subjects already present aberrant reward processes before starting to smoke; this could even present a risk factor for smoking initiation which has to be elaborated in longitudinal designs. Furthermore, it could be possible that heavy smokers were more used to pictures showing drug-related-negative consequences, as they are more confronted with pictures used for health campaigns on cigarette packets. To avoid this potential confound, we explicitly used pictures which are not used by health campaigns and which are not used on cigarette packets. While individualized cues are a strength of the study this requires additional consideration, as any differences in the selected cues between the groups could create a bias. However, we found no systematic preference for a specific food/threat category in one of the two groups and the groups did not differ in their ratings for single cues. Thus, we believe that the risk of bias is somewhat low and does not account for the findings. The inclusion of a younger age

group (e.g. 18 years) may represent a potential bias, as smoking may not yet have been established and the brain is still in its maturation at this age. However, there were only three participants between the ages of 20 and 25 years included in our analysis, assuming that this level of bias is rather low. In addition, TUD subjects were only moderately nicotine-dependent, and stronger dependency could lead to other results which have to be elaborated in future studies. Additionally, it is important to keep in mind that the use of food cues as alternative rewards may not perform identically to other alternative reward categories (e.g. money), as previous studies have shown that nicotine can act as an appetite suppressant [57] or increased appetite can represent a withdrawal symptom [58].

Summarized, our results suggest that altered reward processing is found in moderately dependent TUD subjects and may hence be addressed at all stages of cessation intervention, while the confrontation with long-term consequences might be more promising in light smokers. From a clinical point of view, already existing intervention strategies (e.g. CBT approaches) can be enhanced and new treatment modalities (e.g. real-time fMRI neurofeedback) can be informed by our results. However, these practical consequences still have to be verified in clinical studies.

## ACKNOWLEDGEMENTS

This work is part of the German collaborative Research Center (CRC): 'Losing and regaining control over drug intake'. The study is funded by the German Research Foundation (DFG, Project-ID 402170461, TRR265).

## DECLARATION OF INTERESTS

None.

## AUTHOR CONTRIBUTIONS

**Stefanie L. Kunas:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; visualization; writing-original draft; writing-review & editing. **Heiner Stuke:** Conceptualization; data curation; formal analysis; funding acquisition; methodology; software; supervision; validation. **Andreas Heinz:** Funding acquisition; project administration; resources; supervision; validation; visualization. **Andreas Ströhle:** Conceptualization; funding acquisition; project administration; resources; supervision. **Felix Bermpohl:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization.

## ORCID

*Stefanie L. Kunas* https://orcid.org/0000-0003-3788-229X
*Andreas Heinz* https://orcid.org/0000-0001-5405-9065

## REFERENCES

1. Jha P. Avoidable global cancer deaths and total deaths from smoking. Nat Rev Cancer. 2009;9:655–64.
2. Huxley R, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. Lancet. 2011;378:1297–305.
3. Boyle P, Yasantha Ariyaratne MA, Barrington R, Bartelink H, Bartsch G, Berns A, et al. Tobacco: deadly in any form or disguise. Lancet. 2006;367:1710–2.
4. Samet J. Tobacco smoking: the leading cause of preventable disease worldwide. Thorac Surg Clin. 2013;23:103–12.
5. Nolen-Hoeksema S. Substance Use and Gambling Disorders. Abnormal Psychology. 7th ed. McGraw-Hill; 2017. p. 404–12.
6. Lin X, Deng J, Shi L, et al. Neural substrates of smoking and reward cue reactivity in smokers: a meta-analysis of fMRI studies. Transl Psychiatry. 2020;10:1–9.
7. Versace F, Engelmann JM, Deweese MM, et al. Beyond cue reactivity: non-drug-related motivationally relevant stimuli are necessary to understand reactivity to drug-related cues. Nicotine Tob Res. 2017;19:663–9.
8. Robinson T, Berridge K. The neural basis of drug craving: an incentive-sensitization theory of addiction. Brain Res Rev. 1993;18:247–91.
9. Heinz A. Dopaminergic dysfunction in alcoholism and schizophrenia–psychopathological and behavioral correlates. Eur Psychiatry. 2002;17:9–16.
10. Koob GF, Le Moal M. Drug addiction, dysregulation of reward, and allostasis. Neuropsychopharmacology. 2001;24:97–129.
11. Brody AL, Mandelkern MA, Olmstead RE, et al. Neural substrates of resisting craving during cigarette cue exposure. Biol Psychiatry. 2007;62:642–51.
12. Engelmann JM, Versace F, Robinson JD, et al. Neural substrates of smoking cue reactivity: a meta-analysis of fMRI studies. Neuroimage. 2012;60:252–62.
13. Cortese BM, Uhde TW, Brady KT, et al. The fMRI BOLD response to unisensory and multisensory smoking cues in nicotine-dependent adults. Psychiatry Res. 2015;234:321–7.
14. David SP, Munafò MR, Johansen-Berg H, et al. Ventral striatum/nucleus accumbens activation to smoking-related pictorial cues in smokers and nonsmokers: a functional magnetic resonance imaging study. Biol Psychiatry. 2005;58:488–94.
15. Smolka M, Bühler M, Klein S, et al. Severity of nicotine dependence modulates cue-induced brain activity in regions involved in motor preparation and imagery. Psychopharmacology. 2006;184:577–88.
16. Buhler M, Vollstädt-Klein S, Kobiella A, et al. Nicotine dependence is characterized by disordered reward processing in a network driving motivation. Biol Psychiatry. 2010;67:745–52.
17. Versace F, Lam CY, Engelmann JM, et al. Beyond cue reactivity: blunted brain responses to pleasant stimuli predict long-term smoking abstinence. Addict Biol. 2012;17:991–1000.
18. Oliver JA, Jentink KG, Drobes DJ, Evans DE. Smokers exhibit biased neural processing of smoking and affective images. Health Psychol. 2016;35:866–9.
19. Martin-Soelch C, Missimer J, Leenders KL, Schultz W. Neural activity related to the processing of increasing monetary reward in smokers and nonsmokers. Eur J Neurosci. 2003;18:680–8.
20. Sweitzer MM, Geier CF, Denlinger R, et al. Blunted striatal response to monetary reward anticipation during smoking abstinence predicts lapse during a contingency-managed quit attempt. Psychopharmacology. 2016;233:751–60.
21. Geier CF, Sweitzer MM, Denlinger R, Sparacino G, Donny EC. Abstinent adult daily smokers show reduced anticipatory but elevated saccade-related brain responses during a rewarded antisaccade task. Psychiatry Res. 2014;223:140–7.
22. McClernon FJ, Conklin CA, Kozink RV, et al. Hippocampal and insular response to smoking-related environments: neuroimaging evidence for drug-context effects in nicotine dependence. Neuropsychopharmacology. 2016;41:877–85.

23. Vollstädt-Klein S, Kobiella A, Bühler M, et al. Severity of dependence modulates smokers' neuronal cue reactivity and cigarette craving elicited by tobacco advertisement. Addict Biol. 2011;16:166–75.
24. Versace F, Engelmann JM, Jackson EF, et al. Do brain responses to emotional images and cigarette cues differ? An fMRI study in smokers. Eur J Neurosci. 2011;34:2054–63.
25. Versace F, Engelmann JM, Robinson JD, et al. Prequit fMRI responses to pleasant cues and cigarette-related cues predict smoking cessation outcome. Nicotine Tob Res. 2014;16:697–708.
26. Gray JC, Amlung MT, Acker J, Sweet LH, Brown CL, MacKillop J. Clarifying the neural basis for incentive salience of tobacco cues in smokers. Psychiatry Res. 2014;223:218–25.
27. MacKillop J, Amlung MT, Wier LM, et al. The neuroeconomics of nicotine dependence: a preliminary functional magnetic resonance imaging study of delay discounting of monetary and cigarette rewards in smokers. Psychiatry Res. 2012;202:20–9.
28. Wilson SJ, Delgado MR, McKee SA, et al. Weak ventral striatal responses to monetary outcomes predict an unwillingness to resist cigarette smoking. Cogn Affect Behav Neurosci. 2014;14:1196–207.
29. Campbell WG. Addiction: a disease of volition caused by a cognitive impairment. Can J Psychiatry. 2003;48:669–74.
30. Bechara A. Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. Nat Neurosci. 2005;8:1458–63.
31. Hayes DJ, Northoff G. Identifying a network of brain regions involved in aversion-related processing: a cross-species translational investigation. Front Integr Neurosci. 2011;5:49–58.
32. Dinh-Williams L, Mendrek A, Dumais A, Bourque J, Potvin S. Executive-affective connectivity in smokers viewing anti-smoking images: an fMRI study. Psychiatry Res. 2014;224:262–8.
33. Dinh-Williams L, Mendrek A, Bourque J, Potvin S. Where there's smoke, there's fire: the brain reactivity of chronic smokers when exposed to the negative value of smoking. Prog Neuro-Psychopharmacol Biol Psychiatry. 2014;50:66–73.
34. Langleben DD, Loughead JW, Ruparel K, et al. Reduced prefrontal and temporal processing and recall of high 'sensation value' ads. Neuroimage. 2009;46:219–25.
35. Falk EB, Berkman ET, Whalen D, Lieberman MD. Neural activity during health messaging predicts reductions in smoking above and beyond self-report. Health Psychol. 2011;30:177–85.
36. Chua HF, Ho SS, Jasinska AJ, et al. Self-related neural response to tailored smoking-cessation messages predicts quitting. Nat Neurosci. 2011;14:426–7.
37. Jasinska AJ, Chua HF, Ho SS, Polk TA, Rozek LS, Strecher VJ. Amygdala response to smoking-cessation messages mediates the effects of serotonin transporter gene variation on quitting. Neuroimage. 2012;60:766–73.
38. First MB, Williams JBW, Karg RS, Spitzer RL. Structured Clinical Interview for DSM-5 Disorders, clinician version (SCID-5-CV). Arlington, VA: American Psychiatric Association; 2016.
39. Lehrl S, Triebig G, Fischer B. Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. Acta Neurol Scand. 1995;91:335–45.
40. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the State–Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983. p. 198.
41. Hautzinger M, Bailer M, Hofmeister D, Keller F. *ADS: Allgemeine Depressionsskala [General Depression Scale]*. Göttingen: Hogrefe; 2012.
42. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict. 1991;86:1119–27.
43. Volkow ND, Morales M. The brain on drugs: from reward to addiction. Cell. 2015;162:712–25.
44. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002;1:273–89.
45. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage. 2003;19:1233–9.
46. Morrell H, Cohen LM. Cigarette smoking, anxiety, and depression. J Psychopathol Behav Assess. 2006;28:281–95.
47. Dani JA, De Biasi M. Cellular mechanisms of nicotine addiction. Pharmacol Biochem Behav. 2001;70:439–46.
48. Nestler EJ. Is there a common molecular pathway for addiction? Nat Neurosci. 2005;8:1445–9.
49. Benowitz NL. Neurobiology of nicotine addiction: implications for smoking cessation treatment. Am J Med. 2008;121:S3–10.
50. Markou A. Neurobiology of nicotine dependence. Phil Trans R Soc B Biol Sci. 2008;363:3159–68.
51. Tiffany ST, Carter BL. Is craving the source of compulsive drug use? J Psychopharmacol. 1998;12:23–30.
52. Lane RD, Reiman EM, Bradley MM, et al. Neuroanatomical correlates of pleasant and unpleasant emotion. Neuropsychologia. 1997;35:1437–44.
53. Taylor SF, Liberzon I, Koeppe RA. The effect of graded aversive stimuli on limbic and visual activation. Neuropsychologia. 2000;38:1415–25.
54. Naqvi NH, Gaznick N, Tranel D, Bechara A. The insula: a critical neural substrate for craving and drug seeking under conflict and risk. Ann NY Acad Sci. 2014;1316:53–70.
55. Li X, Hartwell KJ, Borckardt J, et al. Volitional reduction of anterior cingulate cortex activity produces decreased cue craving in smoking cessation: a preliminary real-time fMRI study. Addict Biol. 2013;18:739–48.
56. Wing VC, Barr MS, Wass CE, et al. Brain stimulation methods to treat tobacco addiction. Brain Stimul. 2013;6:221–30.
57. Jo H, Talmage DA, Role LW. Nicotinic receptor-mediated effects on appetite and food intake. J Neurobiol. 2002;53:618–32.
58. al'Absi M, Lemieux A, Nakajima M. Peptide YY and ghrelin predict craving and risk for relapse in abstinent smokers. Psychoneuroendocrinology. 2014;49:253–9.


## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kunas SL, Stuke H, Heinz A, Ströhle A, Bermpohl F. Evidence for a hijacked brain reward system but no desensitized threat system in quitting-motivated smokers: An fMRI study. Addiction. 2022;117:701–12. https://doi.org/10.1111/add.15651

### *2.3 Veränderung von Belohnungsprozessen durch vorhergehende Präsentation negativer Folgen des Rauchens bei Nikotinabhängigkeit*

In der vorangegangenen Studie konnten wir insgesamt keine reduzierte Reaktivität auf die Darstellung von negativen Konsequenzen des Substanzkonsums bei Raucher*innen verglichen mit Nichtraucher*innen nachweisen. Es ist aber dennoch denkbar, dass eine Konfrontation mit den negativen Konsequenzen des Konsums bei Raucher*innen den Belohnungscharakter von anschließend gezeigten positiven substanzbezogenen Reizen reduziert oder die Aktivierung von dem reflektiv-bewussten System zugerechneten Kontrollarealen verstärkt. Sollte dies der Fall sein, wäre es ein Argument, den Einsatz von aversiven substanzbezogenen Reizen als Baustein von Rauchentwöhnungstherapien verstärkt klinisch zu untersuchen.

Um den Einfluss von aversiven substanzbezogenen Reizen auf die Verarbeitung nachfolgender positiver substanzbezogener Reize zu bestimmen, haben wir die Daten der Raucher*innen-Gruppe aus der unter 2.2 beschriebenen Studie re-analysiert.

Wortgetreu und selbstständig übersetztes Abstract des Originalartikels (Kunas, S L, Bermpohl, F, Plank, I S, Strohle, A, & Stuke, H. Aversive drug cues reduce cigarette craving and increase prefrontal cortex activation during processing of cigarette cues in quitting motivated smokers. Addict Biol 2022; 27(1): e13091. doi:10.1111/adb.13091):

*„Aversive Drogenreize können zur Unterstützung der Rauchentwöhnung und zur Sensibilisierung für die negativen gesundheitlichen Folgen des Rauchens eingesetzt werden. Ein besseres Verständnis der Auswirkungen aversiver Drogenreize auf das Suchtverlangen und die Verarbeitung als attraktiv empfundener Drogenreize bei abstinenzmotivierten Raucher*innen ist wichtig, um deren Einsatz in der Entwöhnungstherapie und bei rauchbezogenen gesundheitspolitischen Maßnahmen weiter zu verbessern. In dieser Studie wurden 38 zur Rauchentwöhnung motivierte Personen einer funktionellen Magnetresonanztomographie (fMRT) unterzogen, während sie ein neuartiges erweitertes Reiz-Reaktions-Paradigma durchführten. Bilder von Zigaretten dienten als attraktive Drogenreize, denen entweder aversive Drogenreize (z. B. Raucherbein) oder andere Reize (neutrale oder alternative Belohnungsreize) vorausgingen. Die Teilnehmer*innen wurden angewiesen, ihr Verlangen nach Zigaretten nach der Präsentation der Drogenreize zu bewerten. Wenn aversive Drogenreize vor attraktiven Drogenreizen präsentiert wurden, war das Verlangen nach Zigaretten geringer und die*

*Aktivierungen in präfrontalen (dorsolateraler präfrontaler Kortex) und paralimbischen (dorsaler anteriorer cingulärer Kortex [dACC] und anteriore Insulae) Arealen waren verstärkt. Für den rechten dACC wurde ein positiver Zusammenhang zwischen der Verringerung des Suchtverlanges und den Veränderungen der neurofunktionellen Aktivierung nachgewiesen. Unsere Ergebnisse deuten darauf hin, dass aversive Drogenreize einen Einfluss auf die Verarbeitung attraktiver Drogenreize haben, und zwar sowohl auf neurofunktioneller als auch auf Verhaltensebene. Ein vorgeschlagenes Modell beinhaltet, dass aversive drogenbezogene Hinweise kontrollassoziierte Hirnareale (z. B. dACC) aktivieren, was zu einer verstärkten hemmenden Kontrolle über belohnungsassoziierte Hirnarealen (z. B. Putamen) und zu einer Verringerung des subjektiven Verlangens führt."*

Auch wenn in der unter 2.2 vorgestellten Studie keine unmittelbar gesteigerte Reaktivität bei Raucher*innen bei Konfrontation mit aversiven suchtbezogenen Reizen festgestellt wurde, zeigen die Ergebnisse dieser Studie, dass ein Effekt auf die Verarbeitung nachfolgender positiver suchtbezogener Reize möglich ist. Die Studie ist, wie unter 3.1 ausführlicher diskutiert wird, damit potenziell von Bedeutung für Strategien der Rauchentwöhnung und -prävention.

Received: 4 March 2021 | Revised: 30 May 2021 | Accepted: 29 July 2021
DOI: 10.1111/adb.13091

ORIGINAL ARTICLE

Addiction Biology SSA WILEY

# Aversive drug cues reduce cigarette craving and increase prefrontal cortex activation during processing of cigarette cues in quitting motivated smokers

Stefanie L. Kunas[1] | Felix Bermpohl[1] | Irene S. Plank[1,2,3] | Andreas Ströhle[1] | Heiner Stuke[1]

[1]Department of Psychiatry and Neuroscience, Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Berlin, Germany
[2]Einstein Center for Neurosciences, Charité Campus Mitte, Universitätsmedizin Berlin, Berlin, Germany
[3]Berlin School of Mind and Brain, Institute of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany

**Correspondence**
Stefanie L. Kunas, Department of Psychiatry and Neuroscience, Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Berlin D-10117, Germany.
Email: stefanie.kunas@charite.de

**Funding information**
Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Grant/Award Numbers: 402170461, TRR265; Charité-Universitätsmedizin Berlin

## Abstract

Aversive drug cues can be used to support smoking cessation and create awareness of negative health consequences of smoking. Better understanding of the effects of aversive drug cues on craving and the processing of appetitive drug cues in abstinence motivated smokers is important to further improve their use in cessation therapy and smoking-related public health measures. In this study, 38 quitting motivated smokers underwent functional magnetic resonance imaging (fMRI) scanning while performing a novel extended cue-reactivity paradigm. Pictures of cigarettes served as appetitive drug cues, which were preceded by either aversive drug cues (e.g., smokers' leg) or other cues (neutral or alternative reward cues). Participants were instructed to rate their craving for cigarettes after presentation of drug cues. When aversive drug cues preceded the presentation of appetitive drug cues, behavioural craving was reduced and activations in prefrontal (dorsolateral prefrontal cortex) and paralimbic (dorsal anterior cingulate cortex [dACC] and anterior insulae) areas were enhanced. A positive association between behavioural craving reduction and neurofunctional activation changes was shown for the right dACC. Our results suggest that aversive drug cues have an impact on the processing of appetitive drug cues, both on a neurofunctional and a behavioural level. A proposed model states that aversive drug-related cues activate control-associated brain areas (e.g., dACC), leading to increased inhibitory control on reward-associated brain areas (e.g., putamen) and a reduction in subjective cravings.

**KEYWORDS**
control network, craving, cue reactivity

## 1 | INTRODUCTION

Smoking is a leading cause for cancer, respiratory and cardiovascular diseases and related to an estimated 12% of deaths in the adult population worldwide.[1,2] These approximately 4.8 million cases of premature death each year are preventable, highlighting the importance to identify new and enhance already existing prevention and treatment strategies. One promising approach to target smoking in

public health measures and cessation therapy is the use of aversive drug cues that show negative consequences of smoking in the form of an image or text (e.g., on cigarette packets).

Aversive drug cues have been proven to reduce craving in smokers,[3] to curtail the number of smoking initiators,[4] to augment quitting rates and to raise awareness for health issues related to tobacco consumption.[5] To investigate the neurofunctional mechanisms involved in these beneficial effects, previous functional magnetic resonance imaging (fMRI) studies[6–8] investigated aversive drug cue reactivity. They showed activation of a brain network including regions involved in executive control (e.g., dorsolateral prefrontal cortex [DLPFC]), motor planning regions (e.g., supplementary motor area), limbic regions involved in memory and affect (e.g., hippocampus and thalamus) and visual processing regions (e.g., cuneus and precuneus). These results are complemented by an investigation that found that prefrontal cortex (DLPFC) activation in smokers was associated with increased reward anticipation, poorer learning from errors and decreased attention control.[9] However, while the elucidation of aversive drug cue and punishment processing in smokers is still in its early stages, reactivity towards appetitive drug cues (cigarettes) has already been well investigated.[10,11]

Previous studies, examining cue reactivity in smokers, suggest that the mesolimbic brain reward system (e.g., midbrain, putamen, pallidum, nucleus accumbens [NAc] and ventral striatum) gradually becomes sensitized to drug-related stimuli and desensitized to nondrug-related alternative rewards.[12,13] Increased activation of these reward-associated areas has been directly linked to subjective craving and relapse risk.[14–16]

Neural correlates of resisting craving for tobacco have been linked to prefrontal cortex areas, associated with higher executive functioning and cognitive reward control.[17,18] Importantly, brain regions involved in executive and cognitive reward control processes (e.g., PFC, anterior cingulate cortex [ACC], and anterior insula) possess a rich set of connections to cortical and subcortical areas that are key to emotional and reward processing as well as to craving, and this connectivity is assumed to underlay craving regulation processes.[17,19–21] In line with results of aversive drug cue reactivity and the aforementioned processes, Do and Galván[22] showed negative functional connectivity patterns between prefrontal (DLPFC) and limbic (bilateral amygdala) brain regions in smokers while viewing graphic health warning labels, which was interpreted as improved regulatory control over emotionally responsive brain regions.

However, it is still unclear how aversive drug cues influence the subsequent processing of appetitive drug cues and the related craving. Based on the described prior findings on neurofunctional underpinnings of appetitive and aversive drug cue reactivity as well as craving and its control, an aversive cue model of tobacco use disorder (TUD) can be proposed: aversive drug-related cues change the processing of rewarding drug stimuli by (1) decreasing activation of reward areas (e.g., ventral striatum and putamen), (2) increasing activation of control and self-regulation areas (PFC, ACC and anterior insula) and (3) increasing the down-regulation of reward areas by control areas (Figure 1). In the current study, we aimed to test specific hypotheses derived from the aversive cue model of TUD presented in Figure 1, in quitting-motivated TUD subjects.

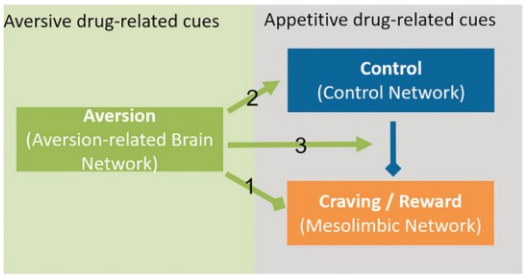


**FIGURE 1** Aversive cue model of tobacco use disorder. The model highlights the effects of aversive drug-related cues on the processing of appetitive drug cues in tobacco use disorder subjects (pointed arrows indicate activation; blunt arrows indicate inhibition). It proposes three pathways through which aversive drug-related cues impact on appetitive drug cue reactivity: drug-related aversive cues (1) directly reduce craving and decrease reactivity of the mesolimbic reward network towards appetitive drug cues, (2) increase activation in control areas during the processing of appetitive drug cues and (3) increase down-regulation of reward areas by control areas

On the basis of the above framework, we hypothesized a reduction of cigarette-cue-induced craving in TUD subjects by prior presentation of aversive drug cues reflected in subjective craving ratings. On a neurofunctional level, we expected reduced activation of the mesolimbic brain circuit (e.g., ventral striatum and putamen) during the processing of appetitive drug cues after presentation of aversive drug cues. On the other hand, we hypothesized greater activations in craving-regulating control areas (e.g., PFC, ACC and anterior insula). Finally, examining functional connectivity patterns, using significantly activated brain regions identified in the group-level analysis as seed regions, we hypothesized negative functional connectivity between control and reward areas after presentation of aversive drug cues.

## 2 | MATERIALS AND METHODS

### 2.1 | Participants

The study was part of the German Collaborative Research Center (TRR 265: losing and regaining control over drug intake), a consortium comprising three German universities funded by the German research foundation (DFG).[23] Here, we present data from one of the main projects of the consortium from Berlin, focusing on understanding specific neural underpinnings underlying human TUD. Whereas previous analyses of the project focused on drug versus alternative reward cue reactivity, the present study investigated the impact of aversive drug stimuli on drug cue reactivity in TUD subjects. Thirty-nine TUD subjects (21 female) were included in the study. One participant had to be excluded due to technical issues with the

autoalignment process during fMRI acquisition, resulting in 38 analysed datasets. Participants were recruited in Berlin through (online and subway) advertising and flyers. Inclusion criteria were current DSM-5 diagnosis of TUD using a structured clinical interview for DSM-5[24] and an age range between 18 and 65 years. Exclusion criteria were comorbid DSM-5 mental disorders within the last 12 months, a lifetime history of any substance use disorder other than TUD, bipolar disorder or psychotic disorder according to DSM-5, current suicidal intent, concurrent psychopharmacological treatment, or psychotherapeutic/psychiatric treatment, a history of brain injury and pregnancy. Additionally, MRI-related exclusion criteria (e.g., ferromagnetic mental implants) were applied. Participants received financial compensation (50 euros) and a 6-week smoking cessation intervention, as all of them were motivated to quit. Additionally, half of the participants were randomized to a sport intervention, of which they were informed before the assessment. The study was approved by the local ethics committee, and all subjects gave written informed consent before participating in the study.

## 2.2 | Clinical assessments

The Fagerstroem Test for Nicotine Dependence (FTND; range 0–10)[25] was used to measure severity of nicotine dependence. To assess participants' global level of intelligence, subjects completed a 35-item multiple choice vocabulary test (MWT; range 0–37).[26] Furthermore, the Alcohol Use Disorder Identification Test (AUDID; range 0–40)[27] was applied to measure everyday alcohol intake and drinking behaviours. Additionally, participants answered five questions assessing their therapy expectancies, evaluating their motivation to take part in the programme and their assessment of its success (range 0–50), and formulated three individualized goal attainments, using the goal attainment scaling.[28]

## 2.3 | fMRI paradigm

A novel fMRI paradigm was established to compare self-reported craving ratings and brain responses to cigarette cues preceded by aversive drug cues with those preceded by other cues (neutral cues or alternative rewarding cues). TUD subjects were instructed to refrain from smoking and eating for 3 h prior to the session. Conventional photographs displaying smoking-related items were used as appetitive drug cues, pictures of attractive food were used as alternative reward cues, pictures showing long-term consequences of smoking (e.g., smokers' leg and lung cancer) were used as aversive drug cues and pictures displaying neutrally valenced items were presented during the neutral control condition. Before the assessment, 140 pictures of each category (appetitive drug cues, alternative reward and aversive drug cues) were rated, with the questions 'how strong is your desire to consume this now?' (appetitive drug cues and alternative reward) and 'how deterrent do you experience this picture?' (aversive drug cues), by each participant using an 8-point Likert scale. The 50% most rewarding/threatening stimuli were automatically selected for the experiment, so that each of the four categories was composed of 70 pictures. In this investigation, we are focusing on the appetitive drug-related and aversive drug-related condition only. Stimuli were presented in the scanner using back-projection. Four pictures of one category were presented per block. Each block lasted 16 s and ended with the presentation of a fixation cross (intertrial interval [ITI]), jittered around 2.5 s. In one run, two blocks of each of the four categories were presented. Within each run, the two blocks with appetitive drug-related cues were preceded once by the aversive drug-related condition and once by one of the other two conditions (either alternative reward or neutral condition). Subjects were instructed to attend to all stimuli and were once per run asked to rate their current desire to consume a cigarette after presentation of the appetitive drug condition and to rate their desire to consume the food after presentation of the alternative reward condition by pressing one of eight buttons covering an 8-point scale ranging from *not at all* to *very strongly*. At the end of each run, participants were additionally asked to rate how strongly they desire to smoke a cigarette, using the same rating scale. In total, the task consisted of nine runs, which altogether lasted maximal 38 min (for an example run, see Figure 2).

## 2.4 | Statistical analysis of behavioural data

Statistical analysis of behavioural data was performed using IBM SPSS statistics 27.0. To quantify the impact of aversive drug cues on subjective desire for cigarettes, we calculated the difference between craving ratings when appetitive drug cues were preceded by aversive drug cues in comparison with other categories (alternative rewards or neutral cues) using a paired-samples *t*-test. In the following, we will refer to this difference as craving reduction induced by aversive drug cues.

## 2.5 | fMRI data acquisition and analysis pathway

Scanning was carried out on a 3T MRI scanner (Siemens Magnetom Prisma) using a 64-channel head coil. Functional images were acquired using a Siemens simultaneous multislice T2*-weighted gradient-echoplanar imaging (EPI) sequence (TR = 869 ms, TE = 38 ms, 60 slices, slice thickness = 2.4 mm, voxel size 2.4 × 2.4 × 2.4 mm, no interslice gap, field of view [FoV] = 210 mm, matrix size 88 × 88, acquisition orientation T > C, interleaved slice order, acceleration factor slice = 6, flip angle = 58°, bandwidth = 1832 Hz/Px, prescan normalize, weak raw data filter, fat sat). Field map images were obtained using a Siemens dual gradient-echo sequence (TR = 698 ms, TE1 = 5.19 ms, TE2 = 7.65 ms, 64 slices, slice thickness = 2.4 mm, no slice gap, voxel size 2.4 × 2.4 × 2.4 mm, FoV = 210 mm, matrix size 88 × 88, acquisition orientation T > C, interleaved slice order, flip angle = 54°, bandwidth = 279 Hz/Px). High-resolution anatomical images were acquired using a T1-weighted MPRAGE sequence

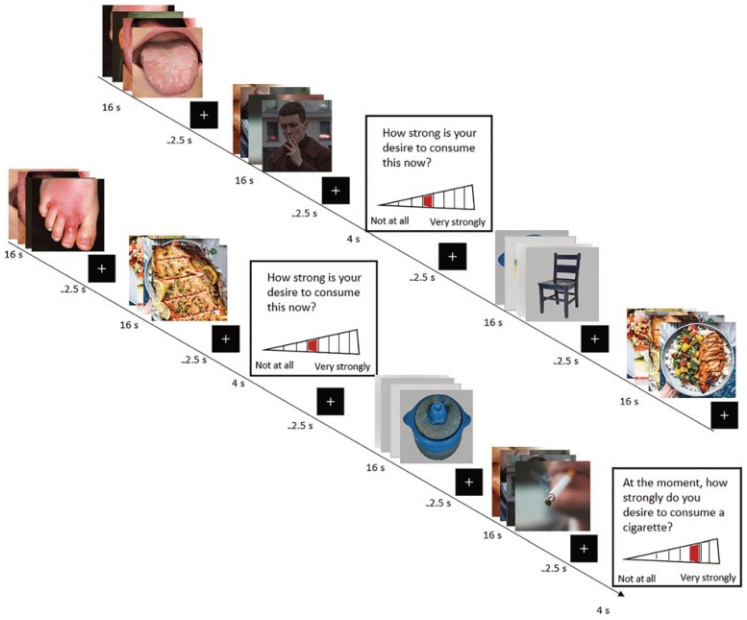


**FIGURE 2** Illustration of an example run of the extended cue-reactivity task. Each condition was presented twice per run (aversive drug related, appetitive drug related, neutral and alternative reward). The appetitive drug-related category was preceded once per run by the aversive drug-related condition (see above) and once by either the neutral or alternative reward condition. Once per run, participants were asked how strong they desire to consume this now (referring to the appetitive drug-related condition). At the end of each run, they were asked how strongly they desire to consume a cigarette now. The task consisted of nine runs with a maximum duration of 38 min

(TR = 2000 ms, TE = 2.01 ms, TI = 880 ms, FoV = 256 mm, 208 sagittal slices, voxel size 1 × 1 × 1 mm, flip angle = 8°, GRAPPA factor 2 [PE], 24 ref. lines, prescan normalize, 23.1% slice oversampling, bandwidth = 240 Hz/Px). To minimize movement artefacts, participants' heads were positioned on a pillow and fixated using foam pads surrounding the head. Image preprocessing was performed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12), implemented in MATLAB R2020a (MathWorks, Sherborn, Massachusetts) and comprised slice timing with reference to the middle slice, SPM12 standard realignment and unwarping including correction for field deformations based on a previously acquired field map, coregistration, normalization to MNI stereotactic space using unified segmentation based on the SPM tissue probability map for six tissue classes, and spatial smoothing with 8-mm full-width at half-maximum isotropic Gaussian kernel (similar to previous studies in our field[29,30]). Following preprocessing, all nine runs were visually inspected, for each subject separately, for a visual quality control.

On the subject level, brain activation differences related to presentation of the different stimuli were analysed using the general linear model (GLM) in SPM12. The blood oxygen level-dependent response was modelled by a canonical haemodynamic response function (HRF) for each of seven conditions: neutral cues, alternative reward cues, aversive drug related cues, nicotine cues preceded by aversive drug-related cues ($Nico^+$), nicotine cues preceded by other cues ($Nico^-$), button presses and ratings, resulting in three regressors of interest ($Nico^+$, $Nico^-$ and neutral) for the current analysis. Model parameter estimates and the resulting *t*-statistic images (condition against baseline) were submitted to the group-level analysis.

Within-group differences were assessed using paired-samples *t*-tests on the second level. To test the effects of presentation of aversive drug-related cues on appetitive cue reactivity, we analysed the contrast $Nico^+ > Nico^-$ and vice versa. Whole-brain analyses as well as an anatomical region of interest (ROI) analysis of a priori defined brain areas were conducted. To investigate the mesolimbic brain reward system, same ROIs as described in previous investigations were included (e.g., Lin et al.[11]): the ventral striatum (NAc), thalamus, pallidum, caudate and midbrain (including ventral tegmental area [VTA]). Furthermore, brain areas responsible for executive and cognitive reward control processes were selected for a second ROI of control areas, including the middle frontal gyrus (DLPFC), orbitofrontal gyrus (ventromedial prefrontal cortex [VMPFC]), superior medial frontal gyrus (dorsomedial prefrontal cortex [DMPFC]), ACC and insula (e.g., Brandl et al.[17] and Morawetz et al.[31]). Combining the definitions from the Automated Anatomical Labeling Atlas,[32] which is implemented in the toolbox 'Wake Forest University PickAtlas'[33] in SPM12, the bilateral ROIs were investigated using one mask. To further specify the regions, we report the corresponding Brodmann areas. Small volume correction was applied using this ROI mask and a family-wise error (FWE) corrected threshold of $p_{fwe} < 0.05$ with a minimum cluster size of $k = 10$ continuous voxels. ROI analyses were followed by whole-brain analyses, thresholded at $p < 0.001$ uncorrected. Furthermore, as sensitivity analysis and to ensure that the experimental manipulation worked, we investigated an activation of the selected ROIs of the reward system in response to appetitive drug cues not preceded by aversive drug cues in comparison with the neutral control condition ($Nico^-$ > neutral).

### 2.6 | Correlation analysis

To test associations between behavioural and neurofunctional effects of aversive drug cues, we computed the Pearson correlations between beta values at significant peaks of the ROIs activated in the group-level analysis and craving reduction induced by aversive drug cues (craving rating after $Nico^-$ minus craving rating after $Nico^+$), implicating that higher values reflect an increased influence of aversive drug cues on subjective cravings. Beta values of the significant ROIs were extracted using the toolbox marsbar[34] with a 5-mm sphere around the peak voxel.

### 2.7 | Generalized psychophysiological interaction analysis (gPPI analysis)

The Functional Connectivity Toolbox (CONN toolbox v18.4)[35] for Matlab and SPM12 was used to perform functional connectivity analyses using the implemented gPPI procedure. This analysis was conducted post hoc to explore the connectivity profile of seed regions identified in the former group-level analysis (bilateral anterior insulae and bilateral dorsal anterior cingulate cortex [dACC] and left DLPFC). A gPPI analysis allows the description of connectivity alterations between brain regions due to an experimental context. The selected seed regions were created as 5-mm spheres around the peak voxel identified in our group-level analysis in the contrast $Nico^+ > Nico^-$ (for MNI coordinates, see Table 2). We used a seed-to-voxel approach to conduct gPPI analyses on the $Nico^+ > Nico^-$ condition. In the first-level analysis, the BOLD time course of all seeds was extracted from each participant and condition, and then, a seed-to-voxel beta map was calculated including the interaction between the seed regions BOLD time series and the $Nico^+ > Nico^-$ contrast condition. Afterwards, the seed regions and beta images were entered into a regression model at the second level. Our goal was to investigate possible stronger negative relationships between control and reward areas. Therefore, we used a one-sided FWE corrected $p < 0.05$ at the cluster level and an uncorrected $p < 0.001$ at the voxel level (as implemented in the CONN toolbox software) to guard against false-positive findings. Cerebrospinal fluid, white matter and six rigid-body parameters were regressed out of the whole-brain grey matter activity.

**TABLE 1** Sociodemographic and psychometric characteristics of the TUD sample

| Sample characteristic | TUD subjects $N = 38$ |
|---|---|
| Age (M [SD]) | 35.18 (10.57) |
| Female gender (*n* [%]) | 21 (55.26) |
| Right-handedness (*n* [%]) | 38 (100) |
| Level of education | |
| A level[a] (*n* [%]) | 30 (78.95) |
| Monthly income in € (*n* [%]) | |
| <1000 | 7 (18.42) |
| 1000–2000 | 12 (31.58) |
| 2000–3500 | 16 (42.11) |
| 3500–4500 | 2 (5.26) |
| >4500 | 1 (2.63) |
| MWT (M [SD]) | 28.32 (4.67) |
| TE (M [SD]) | 40.92 (6.91) |
| FTND (M [SD]) | 4.03 (2.27) |
| Pack years (M [SD]) | 10.75 (9.55) |
| Cigarettes/day (M [SD]) | 14.40 (6.05) |
| AUDID | 5.92 (4.69) |

*Note:* Missing values = monthly income: 1.
Abbreviations: AUDID, Alcohol Use Disorder Identification Test; FTND, Fagerstroem Test of Nicotine Dependence; MWT, Mehrfachwahl-Wortschatz Test; TE, therapy expectancies; TUD, tobacco use disorder.
[a]Abitur.

## 3 | RESULTS

### 3.1 | Sample characteristics and subjective craving ratings

Demographic and smoking characteristics of the TUD sample are shown in Table 1. Craving ratings within the task were significantly lower in the $Nico^+$ condition compared with the $Nico^-$ condition, $t(37) = -4.03$, $p < 0.001$, $d = 0.922$ (see also Table 1 and Figure 3). Participants showed a high motivation to quit smoking and expected the therapy to be helpful to reach this goal (see also Table 1 and Text S1).

### 3.2 | fMRI results

Contrasting the $Nico^+ > Nico^-$ condition, we found greater activations in the bilateral anterior insulae and dACC and in the left DLPFC in the ROI analysis. No significant activations in the VMPFC and DMPFC were observed. On the whole-brain level, significant activations were found in the left middle frontal gyrus (DLPFC), insula, superior frontal gyrus (pre-supplementary motor area, SMA), precentral gyrus, SMA, angular gyrus and caudate as well as in the right calcarine sulcus, cerebellum, SMA, angular gyrus, middle occipital gyrus, insula, middle fontal gyrus (DLPFC), ACC and thalamus (Table 2 and Figure 3). The contrast $Nico^- > Nico^+$ revealed no significant results, neither in the ROI analysis nor in the whole-brain approach.

### 3.3 | Correlation results

We found a positive correlation between craving reduction induced by aversive drug cues (Figure 3B) and right dACC activation

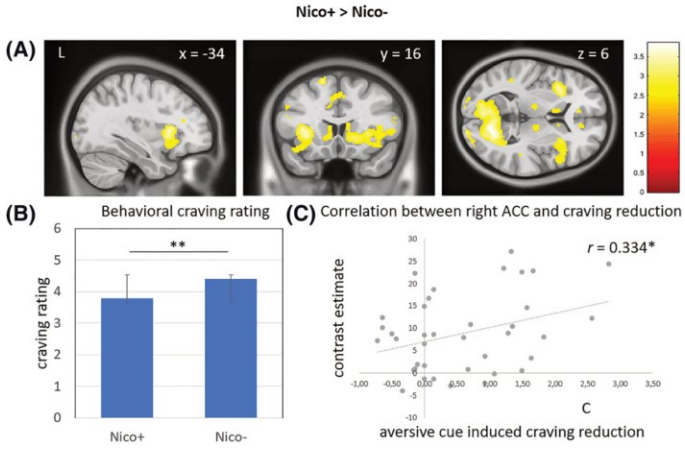


**FIGURE 3** (A) Significantly activated brain regions during the processing of appetitive drug cues, preceded by aversive drug cues in the whole-brain analysis ($p < 0.001$ unc.). (B) Behavioural craving ratings after Nico$^+$ and Nico$^-$. (C) Positive correlation between significantly activated right anterior cingulate cortex (ACC) in the contrast Nico$^+$ > Nico$^-$ and behavioural craving reduction induced by aversive cues. $^*p < 0.05$, $^{**}p < 0.001$

($r = 0.386$, $p = 0.040$) and a trend for the left DLPFC ($r = 0.334$, $p = 0.056$) for the contrast Nico$^+$ > Nico$^-$ (Figure 3C). No other correlations reached significance (left dACC and bilateral insulae).

### 3.4 | Sensitivity results

To ensure that the experimental manipulation worked, we investigated the contrast Nico$^-$ > neutral as sensitivity analysis. We could show an activation of the brain reward system to appetitive drug cues not preceded by aversive drug cues in the ROI analysis (left ventral striatum [NAc] and caudate as well as bilateral pallidum, midbrain and thalamus) and that these activations could be positively associated with behavioural craving ratings (see Tables S1 and S2).

### 3.5 | gPPI results

To exploratively examine functional connectivity patterns of significantly activated brain regions identified in the group-level analysis (left DLPFC as well as bilateral dACC and bilateral anterior insulae), a gPPI analysis was conducted for the contrast Nico$^+$ > Nico$^-$. We found a stronger negative functional connectivity between the left DLPFC and the right supramarginal gyrus, fusiform gyrus, superior occipital gyrus and the left cerebellum. The right anterior insula showed a significant negative functional connectivity to the right nucleus caudatus and the left anterior insula to the right superior occipital gyrus. The right dACC showed a significant negative functional connectivity to the left putamen and the left dACC to the right brain stem (see Table 3). These stronger inverse couplings point towards an aversive cue-induced down-regulation process on mesolimbic brain reward areas (putamen and caudate) by prefrontal and paralimbic control areas (dACC and anterior insula) in TUD subjects.

## 4 | DISCUSSION

The present study proposed an aversive cue model of TUD (Figure 1), describing different pathways through which drug-related aversive cues impact on the processing of appetitive drug cues, based on the literature. According to this aversive cue model of TUD, aversive drug-related cues modulate subsequent drug cue reactivity by (1) reducing subjective craving and neural responsivity in mesolimbic reward areas, (2) enhancing activation in prefrontal control areas and (3) increasing prefrontal top-down-regulation of mesolimbic reward areas. To test our hypotheses, derived from this model in quitting motivated TUD subjects, we employed a novel extended cue-reactivity paradigm, where aversive drug cues (displaying negative consequences of tobacco consumption) preceded the presentation of appetitive drug cues. When appetitive drug cues were preceded by aversive drug cues, we found (1) reduced cigarette craving, but not reduced reactivity in mesolimbic reward areas towards appetitive drug cues; (2) enhanced activation of prefrontal and paralimbic control areas (DLPFC, dACC and anterior insulae), with a positive association between aversion-related reduction of craving and prefrontal activation; and (3) down-regulation of mesolimbic reward areas (putamen and caudate) by prefrontal and paralimbic control areas (dACC and anterior insula). Overall, these findings support our hypotheses referring to all three pathways proposed by the model.

### 4.1 | Impact on craving and mesolimbic reward areas

The craving reduction induced by aversive drug cues is in accordance with our hypothesis (Pathway 1 in Figure 1), consistent with findings from a previous investigation that used graphic health warning labels with different emotional contents[5] and was expected a priori. However, contrary to our hypothesis, we found no corresponding activation reduction of reward-associated brain areas (e.g., NAc and

**TABLE 2** Significant activated brain regions during the processing of appetitive drug cues preceded by aversive drug cues ($Nico^+$) or other cues ($Nico^-$)

| Contrast/region | Side | Voxels | *x* | *y* | *z* | *t* | BA | *p* < 0.001 |
|---|---|---|---|---|---|---|---|---|
| **$Nico^+$ > $Nico^-$** | | | | | | | | |
| *Region of interest analysis* | | | | | | | | |
| Insula | L | 460 | −34 | 16 | 6 | 4.92 | 13 | 0.013* |
| Insula | R | 241 | 32 | 16 | 4 | 4.02 | 13 | 0.049* |
| ACC | L | 59 | −10 | 42 | 10 | 4.14 | 32 | 0.026* |
| ACC | R | 92 | 8 | 34 | 22 | 4.32 | 32 | 0.017* |
| DLPFC | L | 59 | −40 | 36 | 18 | 4.77 | 46 | 0.030* |
| *Whole-brain analysis* | | | | | | | | |
| Calcarine sulcus | R | 3503 | 14 | −70 | 8 | 5.44 | 17 | <0.001 |
| Insula | R | 938 | 31 | 18 | 4 | 4.65 | 13 | <0.001 |
| Cerebellum | R | 678 | 6 | −44 | −16 | 4.80 | - | <0.001 |
| Insula | L | 650 | −34 | 16 | 6 | 4.92 | 13 | <0.001 |
| Cerebellum | R | 311 | 2 | −46 | −16 | 4.59 | - | <0.001 |
| Supplementary motor area | R | 286 | 16 | 4 | 62 | 4.56 | 6 | <0.001 |
| Angular gyrus | L | 230 | −58 | −42 | 26 | 4.30 | 39 | <0.001 |
| Angular gyrus | R | 211 | 54 | −44 | 22 | 4.40 | 39 | <0.001 |
| Supplementary motor area | L | 191 | −18 | 10 | 66 | 4.83 | 6 | <0.001 |
| ACC | R | 192 | 10 | 34 | 20 | 4.32 | 32 | <0.001 |
| Middle occipital gyrus | R | 177 | 20 | −100 | 0 | 4.68 | 18 | <0.001 |
| Caudate | L | 174 | −10 | 16 | 0 | 4.03 | - | <0.001 |
| Precentral gyrus | L | 143 | −16 | −32 | 66 | 4.70 | 4 | <0.001 |
| Middle frontal gyrus (DLPFC) | L | 134 | −40 | 36 | 18 | 5.36 | 46 | <0.001 |
| Thalamus | R | 129 | 12 | −20 | 14 | 4.07 | - | <0.001 |
| Superior frontal gyrus (pre-SMA) | L | 127 | −14 | 14 | 70 | 4.74 | 6 | <0.001 |
| Middle frontal gyrus (DLPFC) | R | 105 | 30 | 42 | 38 | 4.45 | 9 | <0.001 |
| **$Nico^-$ > $Nico^+$** | | | | | | | | |
| *Region of interest analysis* | | | | | No differential activation | | | |
| *Whole-brain analysis* | | | | | No differential activation | | | |

Abbreviations: ACC, anterior cingulate cortex; BA, Brodmann area; DLPFC, dorsolateral prefrontal cortex; L, left; $Nico^+$, appetitive drug-related cues preceded by aversive drug-related cues; $Nico^-$, appetitive drug-related cues preceded by other cues; R, right; SMA, supplementary motor area; voxels, number of voxels per cluster; *x*, *y*, *z*, MNI coordinates.
*$p < 0.05$ family-wise error (FWE) corrected: For ROI analyses, an FWE corrected threshold of $p_{fwe} < 0.05$ with $k > 10$ voxels on the peak level was applied. For whole-brain analyses, an uncorrected threshold of $p < 0.001$ was applied.

**TABLE 3** Results of the seed-based generalized psychophysiological interaction analysis for the contrast $Nico^+$ > $Nico^-$

| Seed | Region | Side | Voxels | *x* | *y* | *z* | *t* | *p* < 0.05 FWE* |
|---|---|---|---|---|---|---|---|---|
| DLPFC (L) | Supramarginal gyrus | R | 227 | 56 | −30 | 56 | −4.52 | 0.023 |
| | Cerebellum | L | 137 | −26 | −82 | −22 | −4.58 | 0.031 |
| | Fusiform gyrus | R | 87 | 22 | −82 | −12 | −4.16 | 0.049 |
| | Superior occipital gyrus | R | 82 | 16 | −66 | 64 | −4.10 | 0.010 |
| Insula (R) | Caudate | R | 76 | 10 | 6 | 16 | −5.17 | <0.001 |
| Insula (L) | Superior occipital gyrus | R | 57 | 16 | −66 | 64 | −4.11 | 0.045 |
| ACC (R) | Putamen | L | 66 | −26 | 0 | −8 | −4.99 | 0.027 |
| ACC (L) | Brainstem | R | 30 | 22 | −28 | −32 | −4.80 | 0.041 |

Abbreviations: ACC, anterior cingulate cortex; DLPFC, dorsolateral prefrontal cortex; FWE, family-wise error; L, left; R, right.
*One-sided, multiple testing correction of $p_{fwe\ (cluster\ level)} < 0.05$.

caudate). This result suggests that activation of the brain reward system through appetitive drug cues is not directly weakened by previous presentation of the negative consequences of smoking, at least not in quitting motivated TUD subjects. Former investigations[36,37] found that a quit interest of smokers modulates smoking cue-reactivity responses in brain reward areas. Thus, it can be assumed that an effect in the mesolimbic reward circuit is more difficult to detect in quitting motivated smokers and may require a lager sample of TUD subjects.

### 4.2 | Impact on prefrontal control areas

Our finding of increased activation of prefrontal control areas related to aversive drug cues complements previous studies on drug cue reactivity, which found appetitive drug cues to activate prefrontal control areas in addition to mesolimbic reward areas.[18] The present findings also add to studies that have linked cognitive control of craving for hedonic stimuli with activations in lateral fronto-parietal cortices (DLPFC and anterior insulae).[17] In accordance with Pathway 2 of the aversive cue model of TUD (Figure 1), we here demonstrate that aversive drug cues enhance the activity of prefrontal control areas during subsequent processing of appetitive drug cues in quitting motivated TUD subjects. The areas identified here (DLPFC, dACC and anterior insulae) are known to play a role in executive functions and different strategies and goals of emotion regulation and cognitive reward control.[17,31,38–40]

Previous studies found the dACC to be involved in cognitive reappraisal and cognitive modulation of emotion as well as in quitting motivated smokers when instructed to actively suppress their urge for cigarettes.[41,42] These results suggest that dACC activation represents an important substrate of inhibition of cue-induced craving in smokers. On the other hand, the DLPFC was found to be involved in different aspects of (cognitive) emotion regulation[31] such as the down-regulation of different kinds of appetitive desires.[43,44] Furthermore, Kober et al.[21] showed DLPFC activation during cognitive down-regulation of craving for cigarettes in smokers when explicitly applying cognitive strategies to regulate craving. These findings suggest that the DLPFC is involved in deliberate regulation of automatic responses to various kinds of affective cues, including drug cues. In terms of emotion regulation, the anterior part of the insula has been suggested to control activity in other brain regions, to initiate and adjust cognitive control mechanisms.[31,45]

Summarizing the above and integrating our own results, two ways of activating the control network can be distinguished. First, previous studies found that the usage of explicit instructions to exert deliberate control over different kinds of stimuli (positive, negative emotions or drug cues) activates the prefrontal control network. Second, we could demonstrate an indirect, implicit activation of prefrontal control areas through aversive drug-related cues, without the instruction to actively apply any strategies, suggesting a rather subsidiary increase of the control network. This second way, which is consistent with our aversive cue model of TUD, may be relevant for smoking prevention programmes and cessation therapy. Based on the knowledge that explicitly targeting the control system in smokers through cognitive interventions (e.g., cognitive behavioural therapy) is limited in its success,[40] alternative strategies are clearly needed. Our results suggest an indirect and automatic activation of control processes through the presentation of (unknown) aversive drug cues, preceding appetitive drug cues. Such strategies could complement explicit cognitive approaches through different forms of application. As a novel part of smoking cessation therapy, (unknown) aversive drug cues could be paired with (individualized) appetitive drug cues of quitting motivated smokers in a conditioning paradigm, maybe inducing decreased craving for these favourite drug cues through enhanced cognitive control. Furthermore, it could be beneficial to make aversive drug cues more visible in different places where smokers are used to consume cigarettes (e.g., smoking areas in public places). By applying such strategies, prevention efforts and cessation success may be enhanced, at least in quitting motivated TUD subjects.

### 4.3 | Impact on top-down control processes

Confirming our hypothesis, we found stronger negative functional connectivity of activated prefrontal control areas to parts of the mesolimbic reward system. Together with the observation of a positive association between right dACC activation and craving reduction induced by aversive drug cues, these findings suggest that aversive drug cues may induce down-regulation processes,[21,22] as proposed by our model (Pathway 3 in Figure 1). A reduction of the overall motivational appeal of smoking may be achieved through balancing the value of cigarettes with the value of the anticipated reward. The value of anticipated reward in mesolimbic brain regions may be down-regulated by prefrontal/paralimbic control areas when aversive drug cues were presented before. However, we only found increased down-regulation of reward areas to appetitive drug cues immediately preceded by aversive drug cues, which might suggest a short-lasting effect of aversive drug cues. This underlines that prevention strategies or cessation interventions, which use aversive drug cues, may benefit from the immediate and contingent presence of aversive drug cues during drug consumption (e.g., on cigarette packets or in novel conditioning paradigms).

### 4.4 | Impact on extended visual system and (pre-) SMA

In addition to prefrontal control areas, we found activation of the extended visual system (e.g., calcarine sulcus and occipital gyrus) as well as in the SMA and pre-SMA in our whole-brain analysis. While the SMA and pre-SMA have been associated with cognitive reward control across a wide range of rewarding stimuli,[17] the extended visual system has consistently been more responsive to smoking cues than neutral cues in previous investigations (e.g., Engelmann et al.[10]). Former fMRI studies, comparing emotionally arousing stimuli with

neutral stimuli, have found that emotionally arousing stimuli consistently evoke larger responses than neutral stimuli in these brain regions, a finding that has been interpreted as increased allocation of attentional resources to the processing of the arousing stimuli.[46,47] In our study, stronger activation of the (extended) visual system during appetitive cue reactivity, when aversive drug cues preceded the presentation, may suggest that those cues are processed as emotionally arousing, particularly in quitting motivated TUD subjects.

## 5 | LIMITATIONS AND CONCLUSION

Our findings should be interpreted within the limitations of this study. Our sample consists of quitting motivated TUD subjects who are medium nicotine dependent according to the FTND scores. Including strong, nonquitting motivated smokers may have changed the results and led to other implications. To specify and extend the effects of this investigation, it would be desirable to study a sample of strong smokers who are not intended to quit smoking. Furthermore, we recruited participants through online or subway advertising, which could possibly have led to a selection bias (e.g., recruiting those who are actually working) and therefore probably limit the external validity of the study.

In conclusion, we assume that cues displaying the negative consequences of smoking have an impact on cigarette cue reactivity and craving in TUD subjects who are motivated to quit. On the basis of previous studies, we proposed an aversive cue model of TUD including three different pathways of the impact of aversive drug-related cues on the processing of appetitive drug-related cues through a reduction of subjective craving and neural responsivity in mesolimbic reward areas, enhanced activation in prefrontal control areas and increased prefrontal top-down-regulation of mesolimbic reward areas. Derived from this model, specific hypotheses were tested. We found a reduction of craving for cigarettes in TUD subjects on a behavioural level. The pattern of brain areas activated when aversive drug cues preceded the presentation of appetite drug cues suggests increased cognitive control (of reward), as well as down-regulation of brain reward areas. Thus, from a neurofunctional perspective, TUD subjects automatically and implicitly applied self-regulation and control strategies. Implications for prevention programmes and smoking cessation interventions include the application of aversive drug cues in different ways (e.g., as conditioning paradigm in cessation interventions). Further research is clearly needed to specify the effect and to investigate the applicability of negative drug-associated stimuli in cessation therapy.

### ACKNOWLEDGEMENTS

H.S. is a participant in the Charité Clinical Scientist Program funded by the Charité-Universitätsmedizin Berlin and the Berlin Institute of Health. We want to thank Ms. Ebba Laing and Ms. Saskia Baumgardt for their support in data acquisition and participant recruitment. Furthermore, we want to express our gratitude to Mr. Michael Marxen for his technical support in fMRI data acquisition. The work is part of the German collaborative Research Center (CRC): Loosing and regaining control over drug intake. The study is funded by the German Research Foundation (DFG Project-ID 402170461, TRR265).

Open Access funding enabled and organized by Projekt DEAL.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### AUTHOR CONTRIBUTIONS

The following authors have approved the final article and have participated in the research. Stefanie L. Kunas, Heiner Stuke, Andreas Ströhle and Felix Bermpohl designed the study. Stefanie L. Kunas collected the data. Stefanie L. Kunas and Heiner Stuke analysed the data. Stefanie L. Kunas, Heiner Stuke, Irene S. Plank, Andreas Ströhle and Felix Bermpohl interpreted the results. Stefanie L. Kunas wrote the first draft. All authors revised the article critically.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on reasonable request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

### ORCID

*Stefanie L. Kunas* https://orcid.org/0000-0003-3788-229X

### REFERENCES

1. Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet*. 2003;362(9387):847-852. https://doi.org/10.1016/S0140-6736(03)14338-3
2. Ezzati M, Lopez AD. Regional, disease specific patterns of smoking-attributable mortality in 2000. *Tob Control*. 2004;13(4):388-395.
3. Partos T, Borland R, Yong H, Thrasher J, Hammond D. Cigarette packet warning labels can prevent relapse: findings from the international tobacco control 4-country policy evaluation cohort study. *Tob Control*. 2013;22(E1):43-50. https://doi.org/10.1136/tobaccocontrol-2011-050254
4. Villanti AC, Cantrell J, Pearson JL, Vallone DM, Rath JM. Perceptions and perceived impact of graphic cigarette health warning labels on smoking behavior among US young adults nicotine & tobacco research. *Nicotine Tob Res*. 2014;16(4):469-477. https://doi.org/10.1093/ntr/ntt176
5. Wang AL, Romer D, Elman I, Turetsky BI, Gur RC, Langleben DD. Emotional graphic cigarette warning labels reduce the electrophysiological brain response to smoking cues. *Addict Biol*. 2015;20(2):368-376. https://doi.org/10.1111/adb.12117
6. Dinh-Williams L, Mendrek A, Bourque J, Potvin S. Where there's smoke, there's fire: the brain reactivity of chronic smokers when exposed to the negative value of smoking. *Prog Neuropsychopharmacol Biol Psychiatry*. 2014a;50:66-73. https://doi.org/10.1016/j.pnpbp.2013.12.009
7. Dinh-Williams L, Mendrek A, Dumais A, Bourque J, Potvin S. Executive-affective connectivity in smokers viewing anti-smoking images: an fMRI study. *Psychiatry Res Neuroimaging*. 2014b;224(3):262-268. https://doi.org/10.1016/j.pscychresns.2014.10.018
8. Chua HF, Ho SS, Jasinska AJ, et al. Self-related neural response to tailored smoking-cessation messages predicts quitting. *Nat Neurosci*. 2011;14(4):426-427. https://doi.org/10.1038/nn.2761
9. Duehlmeyer L, Hester R. Impaired learning from punishment of errors in smokers: differences in dorsolateral prefrontal cortex and

sensorimotor cortex blood-oxygen-level dependent responses. *NeuroImage Clin*. 2019;23:101819. https://doi.org/10.1016/j.nicl.2019.101819
10. Engelmann JM, Versace F, Robinson JD, et al. Neural substrates of smoking cue reactivity: a meta-analysis of fMRI studies. *Neuroimage*. 2012;60(1):252-262. https://doi.org/10.1016/j.neuroimage.2011.12.024
11. Lin X, Deng J, Shi L, et al. Neural substrates of smoking and reward cue reactivity in smokers: a meta-analysis of fMRI studies. *Transl Psychiatry*. 2020;10(1):1-9. https://doi.org/10.1038/s41398-020-0775-0
12. Nolen-Hoeksema S. *Abnormal Psychology*. Substance Use and Gambling Disorders. seventhed. New York: McGraw-Hill; 2017:404.
13. Robinson T, Berridge K. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res Rev*. 1993;18(3):247-291. https://doi.org/10.1016/01650173(93)90013-P
14. Owens MM, MacKillop J, Gray JC, et al. Neural correlates of tobacco cue reactivity predict duration to lapse and continuous abstinence in smoking cessation treatment. *Addict Biol*. 2018;23(5):1189-1199. https://doi.org/10.1111/adb.12549
15. Due DL, Huettel SA, Hall WG, Rubin DC. Activation in mesolimbic and visuospatial neural circuits elicited by smoking cues: evidence from functional magnetic resonance imaging. *Am J Psychiatry*. 2002;159(6):954-960. https://doi.org/10.1176/appi.ajp.159.6.954
16. David SP, Munafò MR, Johansen-Berg H, et al. Ventral striatum/nucleus accumbens activation to smoking-related pictorial cues in smokers and nonsmokers: a functional magnetic resonance imaging study. *Biol Psychiatry*. 2005;58(6):488-494. https://doi.org/10.1016/j.biopsych.2005.04.028
17. Brandl F, Corbi ZLH, Bratec SM, Sorg C. Cognitive reward control recruits medial and lateral frontal cortices, which are also involved in cognitive emotion regulation: a coordinate-based meta-analysis of fMRI studies. *Neuroimage*. 2019;200:659-673. https://doi.org/10.1016/j.neuroimage.2019.07.008
18. Hartwell KJ, Johnson KA, Li X, et al. Neural correlates of craving and resisting craving for tobacco in nicotine dependent smokers. *Addict Biol*. 2011;16(4):654-666. https://doi.org/10.1111/j.1369-1600.2011.00340.x
19. Eippert F, Veit R, Weiskopf N, Erb M, Birbaumer N, Anders S. Regulation of emotional responses elicited by threat-related stimuli. *Hum Brain Mapp*. 2007;28(5):409-423. https://doi.org/10.1002/hbm.20291
20. Wager TD, Davidson ML, Hughes BL, Lindquist MA, Ochsner KN. Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*. 2008;59(6):1037-1050. https://doi.org/10.1016/j.neuron.2008.09.006
21. Kober H, Mende-Siedlecki P, Kross EF, et al. Prefrontal–striatal pathway underlies cognitive regulation of craving. *Proc Natl Acad Sci*. 2010;107(33):14811-14816. https://doi.org/10.1073/pnas.1007779107
22. Do KT, Galván A. FDA cigarette warning labels lower craving and elicit frontoinsular activation in adolescent smokers. *Scan*. 2015;10(11):1484-1496.
23. Heinz A, Kiefer F, Smolka MN, et al. Addiction Research Consortium: losing and regaining control over drug intake (ReCoDe)—from trajectories to mechanisms and interventions. *Addict Biol*. 2020;25(2):e12866. https://doi.org/10.1111/adb.12866
24. First MB, Williams JB, Karg RS, Spitzer RL. *User's Guide for the SCID-5-CV Structured Clinical Interview for DSM-5® Disorders: Clinical Version*. Arlington, VA: American Psychiatric Association; 2016.
25. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerström test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict*. 1991;86(9):1119-1127. https://doi.org/10.1111/j.1360-0443.1991.tb01879.x
26. Lehrl S, Triebig G, Fischer BANS. Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurol Scand*. 1995;91(5):335-345. https://doi.org/10.1111/j.1600-0404.1995.tb07018.x
27. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. *The Alcohol Use Disorders Identification Test, Guidelines for Use in Primary Care*. 2nd ed. Geneva, Switzerland: Department of Mental Health and Substance Dependence, World Health Organization; 2001.
28. Kiresuk TJ, Sherman RE. Goal attainment scaling: a general method for evaluating comprehensive community mental health programs. *Community Ment Health J*. 1968;4(6):443-453.
29. Karoly HC, Schacht JP, Meredith LR, et al. Investigating a novel fMRI cannabis cue reactivity task in youth. *Addict Behav*. 2019;89:20-28. https://doi.org/10.1016/j.addbeh.2018.09.015
30. Wiers CE, Stelzel C, Gladwin TE, et al. Effects of cognitive bias modification training on neural alcohol cue reactivity in alcohol dependence. *AJP*. 2015;172(4):335-343. https://doi.org/10.1176/appi.ajp.2014.13111495
31. Morawetz C, Bode S, Derntl B, Heekeren HR. The effect of strategies, goals and stimulus material on the neural mechanisms of emotion regulation: a meta-analysis of fMRI studies. *Neurosci Biobehav Rev*. 2017;72:111-128. https://doi.org/10.1016/j.neubiorev.2016.11.014
32. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002;15(1):273-289. https://doi.org/10.1006/nimg.2001.0978
33. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*. 2003;19(3):1233-1239. https://doi.org/10.1016/S1053-8119(03)00169-1
34. Brett M, Anton JL, Valabregue R, Poline JB. Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-ROM in. *Neuroimage*. 16(2).
35. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect*. 2012;2(3):125-141. https://doi.org/10.1089/brain.2012.0073
36. Veilleux JC, Skinner KD, Pollert GA. Quit interest influences smoking cue-reactivity. *Addict Behav*. 2016;63:137-140. https://doi.org/10.1016/j.addbeh.2016.07.017
37. Wilson SJ, Sayette MA, Fiez JA. Quitting-unmotivated and quitting-motivated cigarette smokers exhibit different patterns of cue-elicited brain activation when anticipating an opportunity to smoke. *J Abnorm Psychol*. 2012;121(1):198-211. https://doi.org/10.1037/a0025112
38. Lubman DI, Yücel M, Pantelis C. Addiction, a condition of compulsive behaviour? Neuroimaging and neuropsychological evidence of inhibitory dysregulation. *Addiction*. 2004;99(12):1491-1502. https://doi.org/10.1111/j.1360-0443.2004.00808.x
39. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*. 2001;24(1):167-202. https://doi.org/10.1146/annurev.neuro.24.1.167
40. Langner R, Leiberg S, Hoffstaedter F, Eickhoff SB. Towards a human self-regulation system: common and distinct neural signatures of emotional and behavioural control. *Neurosci Biobehav Rev*. 2018;90:400-410. https://doi.org/10.1016/j.neubiorev.2018.04.022
41. Kalisch R, Wiech K, Critchley HD, Dolan RJ. Levels of appraisal: a medial prefrontal role in high-level appraisal of emotional material. *Neuroimage*. 2006;30(4):1458-1466. https://doi.org/10.1016/j.neuroimage.2005.11.011
42. Brody AL, Mandelkern MA, Olmstead RE, et al. Neural substrates of resisting craving during cigarette cue exposure. *Biol Psychiatry*. 2007;62(6):642-651. https://doi.org/10.1016/j.biopsych.2006.10.026

43. Ochsner KN, Gross JJ. Cognitive emotion regulation: insights from social cognitive and affective neuroscience. *Curr Dir Psychol Sci*. 2008; 17(2):153-158. https://doi.org/10.1111/j.1467-8721.2008.00566.x
44. Kim SH, Hamann S. Neural correlates of positive and negative emotion regulation. *J Cogn Neurosci*. 2007;19(5):776-798. https://doi.org/10.1162/jocn.2007.19.5.776
45. Dosenbach NU, Visscher KM, Palmer ED, et al. A core system for the implementation of task sets. *Neuron*. 2006;50(5):799-812. https://doi.org/10.1016/j.neuron.2006.04.031
46. Bradley MM, Sabatinelli D, Lang PJ, Fitzsimmons JR, King W, Desai P. Activation of the visual cortex in motivated attention. *Behav Neurosci*. 2003;117(2):369-375. https://doi.org/10.1037/0735-7044.117.2.369
47. Sabatinelli D, Bradley MM, Lang PJ, Costa VD, Versace F. Pleasure rather than salience activates human nucleus accumbens and medial prefrontal cortex. *J Neurophysiol*. 2007;98(3):1374-1379. https://doi.org/10.1152/jn.00230.2007

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kunas SL, Bermpohl F, Plank IS, Ströhle A, Stuke H. Aversive drug cues reduce cigarette craving and increase prefrontal cortex activation during processing of cigarette cues in quitting motivated smokers. *Addiction Biology*. 2022;27(1):e13091. doi:10.1111/adb.13091

## *2.4 Reduzierte Nutzung von Vorinformationen bei visueller Entscheidungsfindung bei Psychoseneigung*

Wie in der Einleitung dargestellt, postulieren Bayesianische Theorien der Schizophrenie eine veränderte Gewichtung von Vorinformationen und aktueller Information als einen grundsätzlichen computationalen Mechanismus bei der Entstehung von Psychosen. Insbesondere wurde hypothetisiert, dass ein reduzierter Einfluss von Vorinformationen die Korrektur von unwahrscheinlichen sensorischen Informationen verhindert, was zu einer Übergewichtung dieser potenziell irreführenden Informationen führt (s. Abschnitt 1.4). In dieser Studie haben wir untersucht, ob diese Veränderung in Gewichtung von Vorinformation und aktueller Information spezifisch für Wahrnehmungsprozesse ist oder ob es sich um ein umfassenderes Defizit handelt, das auch kognitive Prozesse betrifft.

Wortgetreu und selbstständig übersetztes Abstract des Originalartikels (Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86. doi:10.1093/schbul/sbx189):

*„Predictive coding Theorien postulieren eine veränderte Gewichtung von Vorüberzeugungen und aktueller sensorischer Informationen als eine zentrale Pathologie bei Psychosen. Insbesondere wurde vorgeschlagen, dass der Einfluss von Vorüberzeugungen, die unwahrscheinliche aktuelle sensorische Informationen korrigieren, geschwächt wird, was zu einer Übergewichtung dieser potenziell irreführenden aktuellen Informationen führt. Derzeit ist jedoch unklar, ob diese Veränderung spezifisch für Wahrnehmungsprozesse ist oder ob sie ein umfassenderes Defizit darstellt, das sich auch auf kognitive Prozesse erstreckt. Wir haben zwei Verhaltensexperimente durchgeführt, in denen wir die Nutzung von Vorüberzeugungen bei Wahrnehmungs- bzw. kognitiven Prozessen bei 123 gesunden Personen mit unterschiedlichem Grad an Wahnvorstellungen untersucht haben. In einer audiovisuellen Wahrnehmungsaufgabe mussten die Teilnehmer*innen die globale Bewegungsrichtung von Punktkinematogrammen beurteilen. Die Vorüberzeugungen wurden durch akustische Hinweise induziert, die die globale Bewegungsrichtung der Punktkinematogramme probabilistisch vorhersagten. Ein Kontrollexperiment entsprach dem Design der Wahrnehmungsentscheidungsaufgabe im Bereich der kognitiven Entscheidungsfindung. Durch die Anpassung eines probabilistischen*

*Entscheidungsmodells an die Antworten der Teilnehmer*innen konnten wir den Einfluss von Vorüberzeugungen auf die Entscheidungen der Teilnehmer*innen in beiden Aufgaben quantifizieren. Mit zunehmenden Wahnvorstellungen fanden wir einen geringeren Einfluss von Vorüberzeugungen auf die wahrnehmungsbezogene, nicht aber auf die kognitive Entscheidungsfindung. Unsere Ergebnisse zeigen, dass der Grad an Wahnvorstellungen mit einer spezifisch verringerten Verwendung von Vorüberzeugungen bei Entscheidungen zur Wahrnehmung verbunden ist, wodurch wir predictive coding Theorien von Psychosen empirisch untermauern."*

Diese Ergebnisse legen nahe, dass die Neigung zu Wahnvorstellungen mit einem reduzierten Einfluss von Vorannahmen auf perzeptuelle Entscheidungen einhergeht und stehen damit im Einklang mit der weak prior Hypothese Bayesianischer Psychosetheorien (s. Abschnitt 3.2 für eine übergreifende Diskussion).

Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86.

https://doi.org/10.1093/schbul/sbx189

Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86.

https://doi.org/10.1093/schbul/sbx189

Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86.

https://doi.org/10.1093/schbul/sbx189

Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86.

https://doi.org/10.1093/schbul/sbx189

Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86.

https://doi.org/10.1093/schbul/sbx189

Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86.

https://doi.org/10.1093/schbul/sbx189

Stuke, H, Weilnhammer, V A, Sterzer, P, & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull 2019; 45(1): 80-86.

https://doi.org/10.1093/schbul/sbx189

## *2.5 Verstärkte Detektion bedeutungsvoller Strukturen in visuellem Rauschen bei Psychoseneigung*

In der vorhergehenden Studie fanden wir einen reduzierten Einfluss von Vorannahmen mit zunehmender Psychoseneigung bei perzeptuellen, aber nicht bei kognitiv-probabilistischen Entscheidungen. Andere Arbeiten fanden aber, zunächst konträr zu diesem Befund, einen verstärkten Einfluss von Vorannahmen bei Patient*innen mit psychotischen Erkrankungen (Alderson-Day et al., 2017; Powers, Mathys, & Corlett, 2017; Teufel et al., 2015). Eine denkbare Erklärung für diese Diskrepanz könnte sein (Schmack et al., 2013), dass, während der Einfluss von zeitlich instabilen Vorannahmen bei Psychose reduziert ist, kompensatorisch eine verstärkte Nutzung von früh angeeigneten und zeitlich stabilen Vorannahmen erfolgt (eine Kategorie von Vorannahmen die in späteren theoretischen Arbeiten als β prior bezeichnet wurde (Zeki & Chen, 2020)). Gesichter und gerichteter Blick werden als sozial bedeutsame Reize generell präferentiell verarbeitet (Palermo & Rhodes, 2007), was in einem Bayesianischem Rahmen als Ausdruck einer starken, früh angeeigneten und zeitlich stabilen Vorannahme für Gesichter und gerichteten Blick interpretiert wird. Um diese Hypothese zu untersuchen, setzten wir die Psychoseneigung gesunder Proband*innen ins Verhältnis zu ihrer Neigung, Gesichter wahrzunehmen und den Blick von Gesichtern als auf sie gerichtet zu interpretieren.

Wortgetreu und selbstständig übersetztes Abstract des Originalartikels (Stuke, H, Kress, E, Weilnhammer, V, Sterzer, P, & Schmack, K. Overly Strong Priors for Socially Meaningful Visual Signals Are Linked to Psychosis Proneness in Healthy Individuals. Front. Psychol. 2021; 12(1083). doi:10.3389/fpsyg.2021.583637):

*„Nach der predictive coding Theorie der Psychose werden Halluzinationen und Wahnvorstellungen durch eine Übergewichtung von Vorüberzeugungen auf hohen Verarbeitungsebenen im Verhältnis zu sensorischen Informationen erklärt, die zu fälschlichen Wahrnehmungen bedeutungsvoller Signale führt. Es ist jedoch derzeit unklar, ob die angenommene Übergewichtung von Vorüberzeugungen (1) eine durchgängige Veränderung darstellt, die sich auch auf die visuelle Modalität erstreckt, und (2) bereits in frühen, automatischen Phasen der Reizverarbeitung wirksam wird. Um diese Fragen zu klären, untersuchten wir die visuelle Wahrnehmung sozial bedeutsamer Stimuli bei gesunden Personen*

*mit unterschiedlichem Grad an Neigung zu psychotischen Symptomen (n = 39). In einer ersten Aufgabe quantifizierten wir die Vorüberzeugung der Teilnehmer*innen bezüglich des Vorliegens von Gesichtern im visuellen Rauschen mithilfe eines Bayes'schen Entscheidungsmodells. In einer zweiten Aufgabe maßen wir die Vorüberzeugung der Teilnehmer*innen bezüglich des Vorliegens von direktem Blickkontakt in Gesichtern, die durch continuous flash suppression unsichtbar gemacht wurden. Wir fanden, dass die Vorüberzeugung bezüglich des Vorliegens von Gesichtern im Rauschen mit der Neigung zu Halluzinationen (r = 0,50, p = 0,001, Bayes-Faktor 1/20,1) sowie mit der Neigung zu Wahnvorstellungen (r = 0,46, p = 0,003, BF 1/9,4) korrelierte. Die Vorüberzeugung bezüglich des Vorliegens von direktem Blickkontakt war signifikant mit der Neigung zu Halluzinationen assoziiert (r = 0,43, p = 0,009, BF 1/3,8), aber nicht eindeutig mit der Neigung zu Wahnvorstellungen (r = 0,30, p = 0,079, BF 1,7). Unsere Ergebnisse liefern Belege für die Idee, dass übermäßig starke Vorüberzeugungen auf hohen Verarbeitungsebenen für die automatische Erkennung sozial bedeutsamer Reize eine Störung der Informationsverabreitung bei Psychosen darstellen könnten."*

Diese Ergebnisse sprechen für die strong prior Hypothese in dem Sinne, dass wir eine stärkere Vorannahme für sozial bedeutsame Reize bei Proband*innen mit stärkerer Neigung zu subklinischen Psychosesymptomen fanden. Eine Diskussion im Gesamtkontext der Bayesianischen Psychosemodelle findet sich in Abschnitt 3.2.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 08 April 2021
doi: 10.3389/fpsyg.2021.583637

# Overly Strong Priors for Socially Meaningful Visual Signals Are Linked to Psychosis Proneness in Healthy Individuals

*Heiner Stuke[1]*, Elisabeth Kress[2], Veith Andreas Weilnhammer[1], Philipp Sterzer[1,2] and Katharina Schmack[1]*

[1]Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Berlin, Germany, [2]Bernstein Center of Computational Neuroscience, Berlin, Germany

According to the predictive coding theory of psychosis, hallucinations and delusions are explained by an overweighing of high-level prior expectations relative to sensory information that leads to false perceptions of meaningful signals. However, it is currently unclear whether the hypothesized overweighing of priors (1) represents a pervasive alteration that extends to the visual modality and (2) takes already effect at early automatic processing stages. Here, we addressed these questions by studying visual perception of socially meaningful stimuli in healthy individuals with varying degrees of psychosis proneness ($n$ = 39). In a first task, we quantified participants' prior for detecting faces in visual noise using a Bayesian decision model. In a second task, we measured participants' prior for detecting direct gaze stimuli that were rendered invisible by continuous flash suppression. We found that the prior for detecting faces in noise correlated with hallucination proneness ($r$ = 0.50, $p$ = 0.001, Bayes factor 1/20.1) as well as delusion proneness ($r$ = 0.46, $p$ = 0.003, BF 1/9.4). The prior for detecting invisible direct gaze was significantly associated with hallucination proneness ($r$ = 0.43, $p$ = 0.009, BF 1/3.8) but not conclusively with delusion proneness ($r$ = 0.30, $p$ = 0.079, BF 1.7). Our results provide evidence for the idea that overly strong high-level priors for automatically detecting socially meaningful stimuli might constitute a processing alteration in psychosis.

**Keywords: face processing, perceptual bias, predictive coding, psychosis proneness, hallucination, gaze detection**

OPEN ACCESS

***Edited by:***
*Lars Muckli,*
*University of Glasgow, United Kingdom*

***Reviewed by:***
*Vincenzo Romei,*
*University of Bologna, Italy*
*Taiyong Bi,*
*Zunyi Medical University, China*

****Correspondence:***
*Heiner Stuke*
*heiner.stuke@charite.de*

***Specialty section:***
*This article was submitted to Perception Science, a section of the journal Frontiers in Psychology*

***Received:*** *21 October 2020*
***Accepted:*** *11 March 2021*
***Published:*** *08 April 2021*

***Citation:***
*Stuke H, Kress E, Weilnhammer VA, Sterzer P and Schmack K (2021) Overly Strong Priors for Socially Meaningful Visual Signals Are Linked to Psychosis Proneness in Healthy Individuals. Front. Psychol. 12:583637. doi: 10.3389/fpsyg.2021.583637*

## INTRODUCTION

Schizophrenia is characterized by psychotic symptoms such as delusions and hallucinations. Neurocognitive theories that draw on predictive coding and Bayesian theories of brain function have proposed an imbalance between prior expectations and current sensory information as a central disturbance underlying psychotic experiences (Fletcher and Frith, 2009; Adams et al., 2013; Sterzer et al., 2018). In this context, an overly strong prior for socially meaningful signals can account for hallucinatory experiences, such as hearing voices in the absence of causative stimulus, or delusional experiences, such as the feeling of being looked at by strangers (Corlett et al., 2009, 2019).

Consistent with this theoretical framework, an increased tendency to perceive voices in auditory noise has been observed in psychosis and related conditions (Bentall and Slade, 1985; Hoffman et al., 2007; Vercammen et al., 2008; Galdos et al., 2011; Alderson-Day et al., 2017), in line with the idea of overly strong prior for socially meaningful signals in the auditory domain. A similar shift toward perceiving abstract signals, such as pure tones, in auditory noise (Powers et al., 2017) points to the possibility that overly strong priors might affect auditory perception in general.

Hence, while there is evidence to support the idea of overly strong priors for meaningful auditory signals in psychosis, it is currently unclear whether this reflects a generic processing deficits that reliably extends to the visual modality. A few studies have related an increased tendency to perceive faces in visual noise (Partos et al., 2016), and an increased tendency to perceive visual gaze as direct (Rosse et al., 1994; Hooker and Park, 2005; Tso et al., 2012) to psychosis and related conditions, but results have been mixed (see Franck et al., 2002 for a negative report). Assessing relationships between psychotic experiences and the use of priors toward meaningful visual signals is crucial for probing the generalizability of strong prior accounts of psychosis. Here, we therefore related psychosis proneness in individuals from the general population to behavior in a visual detection-in-noise task. We hypothesized that psychosis proneness would positively correlate with the tendency to detect faces in visual noise, and hence a prior toward detecting meaningful stimuli.

Moreover, it is currently unclear that which stage of information processing is affected by overly strong priors underlying psychotic experiences. It is conceivable that overly strong priors might only affect the late, conscious processing stage of cognitive interpretation. Alternatively, the effects of overly strong priors might extend to early, automatic sensory processing stages that determine the access of stimuli to awareness. In the visual domain, the potency of visual stimuli to gain access to awareness can be assessed with interocular masking techniques such as continuous flash suppression (CFS; Tsuchiya and Koch, 2005). In CFS, one eye is presented with a target stimulus, while the other eye is presented with a dynamic mask that initially suppresses the target stimulus from conscious perception. The time that the suppressed stimulus takes to overcome interocular suppression has been proposed as a measure for the potency of a specific stimulus to gain access to awareness (Jiang et al., 2007; Stein and Sterzer, 2014). For example, this "breaking CFS" paradigm (b-CFS; Stein et al., 2011a) has been used to show that suppression times are decreased for stimuli with direct gaze as compared to stimuli with averted gaze (Stein et al., 2011b). Inter-individual variability in breakthrough time depends on individual factors related to the stimuli that compete for perceptual dominance. For example, the advantage for faces with direct gaze in gaining access to awareness is reduced in individuals with autistic traits (Akechi et al., 2014; Madipakkam et al., 2019). Similarly, suppression times are reduced for sad faces in patients with major depression (Sterzer et al., 2011) and for spider stimuli in individuals with spider phobia (Schmack et al., 2016). Here, we asked whether a strong prior for direct gaze may affect those processing stages that determine access of face stimuli to awareness and therefore tested whether suppression times for direct compared to averted gaze may be shorter in individuals with high psychosis proneness.

The "Psychosis Continuum" view postulates that the clinical manifestations of psychosis represent the most extreme form of psychosis proneness, which is continuously distributed in the general population (Barrantes-Vidal et al., 2015; DeRosse and Karlsgodt, 2015). Indeed, psychotic experiences are not confined to clinical populations, but can be found to varying degrees in the general population (Peters et al., 2004; Bell et al., 2006). Interestingly, subclinical psychosis proneness and clinical psychosis are associated with similar risk factors (van Os et al., 2009; Linscott and van Os, 2013) and exhibit a shared factor structure of symptoms (Shevlin et al., 2017). Furthermore, the relatives of patients with psychotic disorders show increased levels of subclinical psychosis proneness, suggesting common genetic underpinnings (Kendler et al., 1993; Fanous et al., 2001; Tienari et al., 2003). Importantly, high levels of subclinical psychosis proneness increase the risk for later clinical psychosis (Chapman et al., 1994; Hanssen et al., 2005; Welham et al., 2009). Taken together, these findings suggest that subclinical and clinical psychotic experiences are mediated by shared processes. Hence, the investigation of subclinical psychosis proneness in non-patient populations can provide insights into the processes underlying psychotic experiences in general, while not being confounded by psychotropic medications or other concomitants of clinical psychotic disorders.

Here, we tested whether delusion and hallucination proneness relate to overly strong priors for detecting socially meaningful stimuli, as quantified in two visual detection tasks. Specifically, we hypothesized that psychosis proneness would correlate to an enhanced prior for detecting faces in noisy sensory information and an enhanced prior for detecting direct gaze in stimuli rendered invisible with continuous flash suppression.

## MATERIALS AND METHODS

### Participants and Psychometry

Thirty-nine participants were recruited from the general population through advertising. The study was approved by the Ethical Committee of the Charité, Universitätsmedizin Berlin. After complete description of the study to the participants, written informed consent was obtained in accordance with the Declaration of Helsinki of 1975 before participation.

Psychosis proneness was assessed with questionnaires previously validated in non-clinical populations. Here, proneness to delusional ideation was quantified using the Peters Delusion Inventory, 21-item version (PDI-21; Peters et al., 2004). The 21 items of this self-rating questionnaire cover a wide range of delusional convictions including beliefs in the paranormal, grandiosity ideas, or suspicious thoughts. For every endorsed belief, the questionnaire asks for dimensional ratings of belief-related distress, preoccupation, and conviction.

Additionally, proneness to hallucinatory experiences was assessed with the Cardiff anomalous perception scale (CAPS; Bell et al., 2006). This 32-item self-rating scale assesses anomalous perceptual experiences in different sensory domains including proprioception, time perception, somatosensory perception, and visual and auditory perception. The intensity of every anomalous perception is quantified on subscales for intrusiveness, frequency, and distress. As in our previous work (Stuke et al., 2017, 2018), we used total PDI and CAPS scores obtained by adding up their three subscales.

## Face Task

To quantify priors toward socially meaningful stimuli in visual perception, we measured psychosis-like mispercepts of illusory faces in noise. To this end, we devised a face detection task that required the participants to detect faces embedded in noise. One-hundred stimuli (40 target and 60 noise stimuli) were created. Participants were instructed that a sequence of noisy stimuli will be presented to them and that some of stimuli will contain a human face. Each stimulus was presented for 3,000 ms followed by a forced-choice decision of whether a face was present or not. After a response had been made and a subsequent inter trial interval of 800 ms, the next stimulus was presented (**Figure 1**).

Stimuli were designed to resemble those that have proven to induce psychosis-like percepts of illusory faces in previous work (Partos et al., 2016). Noise stimuli consisted of a noise pattern only (without embedded face) and were created in three steps using Matlab and the Image Processing Toolbox. Firstly, basic noise patterns were generated by randomly placing a total of 1,000 black circles with diameters varying randomly from 1 to 15 pixels (0.04°–0.64° of visual angle) on a white image of 450 × 450 pixels (19.45° of visual angle). Secondly, the basic noise patterns were degraded by adding multiplicative noise (as implemented in the "speckle" command of the Matlab imnoise routine with a distribution variance of 2). Finally, the resulting noise stimuli were blurred with a Gaussian filter ("gaussian" command of the Matlab imnoise routine with a distribution variance of 10) and image contrast was reduced with the "imadjust" routine (resetting gray scale intensities to values between 0.1 and 0.9). For the target stimuli, 20 adult faces with neutral expression were taken from the Productive Aging Laboratory Face Database (Minear and Park, 2004) and placed at random positions in the noise stimuli before the third step of noise image generation (i.e., before the Gaussian filter and contrast reduction). All faces were oriented upright. The specific image generation parameters were chosen to ensure that participants were imperfectly able to distinguish the faces from the noise stimuli in a pilot study with five participants (discriminability mean = 0.81, SD = 0.03; bias = 1.52, SD = 0.86; computed using signal detection theory equations; Stanislaw and Todorov, 1999).

## Face Task Analysis

Face task behavior was analyzed with a Bayesian model combining an individual prior for detecting faces with a sensory likelihood of a face depending on whether a face was embedded in the stimulus or not.

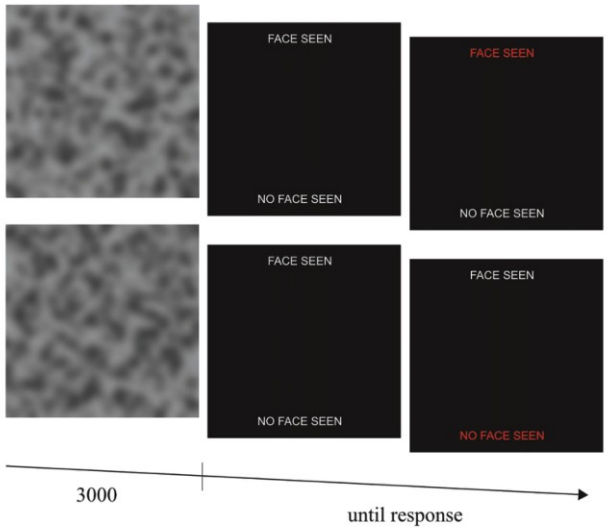


**FIGURE 1** | Experimental sequence of the face task for two exemplary trials with (top row) and without (lower row) embedded face. Face images were shown for 3,000 ms followed by a binary forced choice indication of whether a face had been detected by the participants. By fitting a Bayesian model to the individual participants' behavior, we obtained a measure of the prior for detecting faces.

Hence, the probability of detecting a face in each trial was:
Equation 1:

$$P(face\ \ detected) = \frac{Prior \times Likelihood}{Prior \times Likelihood + (1 - Prior) \times (1 - Likelihood)}$$

where *Prior* is an estimated free parameter and *Likelihood* was computed as follows.

The sensory likelihood of a face *Likelihood* depended on whether a face was embedded in the stimulus:
Equation 2:

$$Likelihood = Sensitivity^{face} \times (1 - Sensitivity)^{1-face}$$

where sensitivity is an estimated free parameter and *face* is a binary vector, indicating whether a face was embedded in each trial.

The objective function to be maximized for each participant was hence:
Equation 3:

$$L = \sum_{i=1} \log\left(P(face\ \ detected)_i^{face\ \ detected_i} \times \left(1 - P(face\ \ detected)_i\right)^{(1-face\ \ detected_i)}\right)$$

where $i$ is an index denoting the trial number and *face detected* is a binary vector, indicating whether a face was detected in each trial by the participant.

For each participant, this model estimates a prior probability for detecting a face as well as a sensitivity parameter capturing how much the likelihood of detecting a face depended on whether the stimulus contained a face or not. Estimation of individual face prior and sensitivity values by maximizing the objective function given by the equations above was carried out using Powell's optimization (Powell, 1964) as implemented in SciPy for Python with prior bounds between 0 and 1.

## Gaze Task

To quantify the effect of individual priors for socially meaningful information on the access of visual stimuli to awareness, we used an established interocular suppression task with face stimuli that displayed either direct or averted gaze (Stein et al., 2011b; Seymour et al., 2016; Madipakkam et al., 2019). In this task, stimuli were photographs of three different female faces, each in a version with direct and averted gaze. The impression of eye gaze being either directed at or away from the observer was achieved by a shift of the pupil to the left or the right. For example, a head rotated to the right together with the pupil shifted to the left resulted in the impression of a face looking at the observer (see **Figure 2**, lower left). All faces were cut into oval shapes comprising a size of 3.8° × 4.5° and equalized for global contrast and luminance. Participants viewed the screen through a mirror stereoscope, which provided separate visual input to the two eyes. The participant's head was stabilized by a chin rest at a viewing distance of 50 cm and stimuli were displayed on a 19-inch CRT monitor (resolution: 1024 × 768 Px; refresh rate: 60 Hz).

The effect of eye gaze on access of face stimuli to awareness was assessed using bCFS. Each trial began with a 2-s presentation of white frames (12.0° × 12.0°) with a gray background and a red fixation cross (**Figure 2**). Thereafter, high-contrast, gray scale, dynamic masks were flashed to a randomly selected eye at a frequency of 10 Hz, while simultaneously a face stimulus with either a direct or averted gaze was gradually introduced to the other eye. The contrast of the face stimulus was gradually increased from 0 to 100% within the first second from the beginning of the trial and the stimulus remained at maximum contrast until a response was made or for a maximum of 15 s. The stimuli could be presented in one of the four quadrants of the white frame (3.4° horizontal displacement from the fixation cross and 3° vertical displacement). Participants had to indicate the location of the face (i.e., the quadrant) by button press as soon as they detected a face. Importantly, participants' task (i.e., location discrimination) was orthogonal to the condition of interest (i.e., gaze direction of the presented faces). Participants were therefore unaware of the existence of two different gaze directions. Participants completed a total of 48 trials (12 trials with direct gaze shown on the left, 12 trials with direct gaze shown on the right, 12 trials with averted gaze shown on the left, and 12 trials with averted gaze shown on the right) in a randomized order. Target variables were the response (breakthrough) times for correctly localized faces.

## Gaze Task Analysis

Three of the 39 participants were not included in the gaze task analysis, one because of technical problems and two because of the task did not work due to excessive stimulus suppression by the mask (more than 65% missed trials).

Analogously to previous studies using the same task, we compared mean breakthrough times separately for faces with direct and averted gaze. By subtracting breakthrough times for direct gaze from breakthrough times for averted gaze, we obtained a measure of the tendency toward faster access to awareness of direct gaze. In the following, we denote this measure as "direct gaze bias," where a positive value indicates shorter breakthrough times for direct gaze relatively relative to averted gaze. As a sanity check, we first tested if this measure was significantly above zero (Bayesian one sample $t$ test), e.g., whether we could replicate previous findings of a generally faster breakthrough of direct gaze. Secondly, we tested whether the degree of this direct gaze bias depended on the individual's psychosis proneness by correlating it with CAPS and PDI scores.

## Relationships Between Psychosis Proneness, Face Bias, and Direct Gaze Bias

Statistical analyses were carried out in SPSS 27 and SciPy for Python. Psychosis proneness (CAPS and PDI scores) has been found to describe non-normal, skewed distributions in the general population samples (i.e., Peters et al., 2004; Bell et al., 2006; Stuke et al., 2017, 2018), and the distribution of the direct gaze bias was better described by uniform than by a normal distribution (Akaike information criterion for fitted uniform or normal distribution, as implemented in the scipy.stats library). Therefore, we could not assume normality of

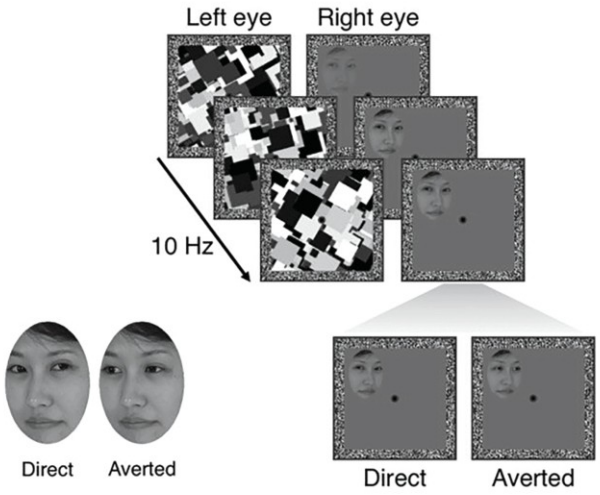


**FIGURE 2 |** Experimental sequence of the unconscious gaze task. Both eyes of the participants received separate stimulation through a mirror stereoscope. One eye was shown the target stimulus (faces with either direct or averted gaze), whose conscious perception was suppressed by a dynamical mask shown to the other eye. Participants were instructed to indicate the localization of a face as soon as it broke through the mask, while not being told that faces differed with respect to gaze. By subtracting mean response times for averted gaze from mean response times for direct gaze, we obtained a measure of the preferential unconscious processing or prior for direct gaze. Figure adapted from Stein et al. (2011b).

our data and analyzed the relationship between psychosis proneness and behavioral measures using rank correlations. In order to obtain Bayes factors for hypothesis testing, we first performed a rank transformation of the data and then used Bayesian correlations (default implementation in SPSS 27) to investigate relationships between delusion and hallucination proneness, face bias, and direct gaze bias.

We report correlation coefficients both with frequentist values of $p$ as well as Bayes factors with the likelihood ratio between the hypothesis of no correlation, and the hypothesis of existing correlation between the tested variables [P(D|H0)/P(D|H1)]. Here, BF > 1 indicates evidence against a correlation, while BF < 1 indicate evidence for a correlation. Moreover, BF > 10 or BF < 1/10 are considered as "strong" evidence, while 3.2 < BF < 10 or 1/10 < BF < 1/3.2 indicate "substantial" evidence and BF < 3.2 or BF > 1/3.2 are viewed as evidence "barely worth mentioning" (Jeffreys, 1998).

## RESULTS

### Participants and Psychometry

**Table 1** summarizes basic demographics as well as delusion proneness (PDI scores) and hallucination proneness (CAPS scores) of the sample. In our non-clinical sample, the mean PDI-21 score was comparably high with 77.0 (42.9) as compared to 58.9 (48.0) in the non-clinical sample of the original publication of the questionnaire (Peters et al., 2004). Moreover, 12.8% (five individuals) had PDI total scores above 130, which was the mean score for the clinical sample of schizophrenia patients in the original publication. Thus, we observed a range of delusional symptoms that overlapped with the range found in samples with clinical disease.

**TABLE 1 |** Participants' characteristics.

| Characteristic | Mean (SD) |
|---|---|
| Age | 30.31 (10.06) |
| PDI score | 76.95 (42.86) |
| CAPS score | 106.90 (48.81) |
| **Characteristic** | **Absolute numbers** |
| Sex | Female: 19 and male: 20 |
| Smoking | Yes: 11; no: 26; and missing information: 2 |
| Vocation | None: 8; apprenticeship: 1; bachelor: 16; and master: 14 |

Similarly, the mean CAPS score we observed was comparably high with 106.9 as compared to 44.4 in the non-clinical sample in the original publication of the questionnaire (Bell et al., 2011). Here, 7.7% (three individuals) had a total score higher than 172, which was the mean of the clinical patient sample in the original study by Bell et al. (2011). Hence, our sample showed comparably high psychosis proneness with a considerable number of individuals with a degree of symptoms previously observed in clinical populations.

### Face and Gaze Task Results

In the face task, the mean value (SD) for the estimated face prior was 0.427 (0.143) and the sensitivity parameter 0.737 (0.080). The prior for face detection correlated with both

hallucination proneness ($r$ = 0.496, $p$ = 0.001, $n$ = 39, BF 1/20.83) and delusion proneness ($r$ = 0.461, $p$ = 0.003, $n$ = 39, BF 1/9.43). These results suggest that psychosis proneness is associated with an increased prior for faces in a detection-in-noise task (**Figure 3B**).

In contrast, the sensitivity measure was not significantly related to hallucination proneness ($r$ = −0.138, $p$ = 0.401, $n$ = 39, BF 5.65) or delusion proneness ($r$ = −0.218, $p$ = 0.183, $n$ = 39, BF 3.33). Thus, there was no evidence for a significant association of psychosis proneness with the ability to discriminate between face and noise stimuli.

In the gaze task, breakthrough times were on average 3.385 s (1.310) for direct gaze and 4.055 s (1.465) for averted gaze. Hence, breakthrough times were significantly shorter for direct compared to averted gaze (paired $t$ test, $T$ = −4.362, $p$ < 0.001, $n$ = 36, BF = 1/20). Consistent with previous work (Stein et al., 2011b; Seymour et al., 2016), this result indicates a general direct-gaze bias for access to awareness in the whole sample (**Figure 3A**).

Breakthrough times were not directly correlated with hallucination or delusion proneness (all values of $p$ > 0.105, all BF < 2.1). However, when we calculated the difference between breakthrough times for direct and averted gaze as a measure of the direct gaze bias, this direct gaze bias correlated significantly with hallucination proneness ($r$ = 0.429, $p$ = 0.008, $n$ = 36, BF 1/3.83), but not significantly with delusion proneness (PDI scores, $r$ = 0.297, $p$ = 0.079, $n$ = 36, BF 1.67). These results indicate that hallucination proneness is associated with enhanced access of direct gaze to awareness, suggesting a stronger a prior for socially relevant visual information (**Figure 3C**).

## DISCUSSION

In the present work, we showed that an increased prior for faces in noisy visual stimuli was related to higher hallucination and delusion proneness and an enhanced processing of direct compared to averted invisible gaze was related to higher hallucination proneness, but not to delusion proneness. These results are largely compatible with strong prior accounts of hallucinations (Corlett et al., 2019; Horga and Abi-Dargham, 2019) that state that overly strong priors during naturally noisy perception lead to the false perception of meaning in noise, which in turn is the substrate of psychotic experiences such as hallucinations and delusions.

A Bayesian approach to experimental evidence can take into consideration not only the single study, but also integrate it with evidence from prior studies when estimating posterior probabilities of hypotheses. In our study, the Bayes factors yielded strong evidence (BF > 10; Jeffreys, 1998) for correlations between hallucination proneness and an increased face prior as well as substantial evidence (BF > 3.2) for correlations between delusion proneness and an increased face prior. Moreover, there is prior work pointing into the direction of an increased face perception bias in psychosis proneness (Partos et al., 2016). Hence, the combined evidence renders it very likely that psychosis (proneness) is related to an increased prior for face detection in noisy but visible stimuli.

The combined evidence for an increased unconscious bias for direct gaze is less clear. In the present study, we found substantial evidence (BF > 3.2) *for* a correlation between hallucination proneness and direct gaze bias. We also found evidence *against* the hypothesis of a correlation between delusion proneness and direct gaze bias, whose significance, however, was "barely worth mentioning" (BF < 3.2; Jeffreys, 1998). Prior work investigating an unconscious direct gaze bias in psychotic patients compared to healthy controls with a very similar task yielded a result that was numerically in the direction of an increased direct gaze bias in patients but fell of significance (Seymour et al., 2016). One explanation for these conflicting findings might be that a direct gaze bias was less present in

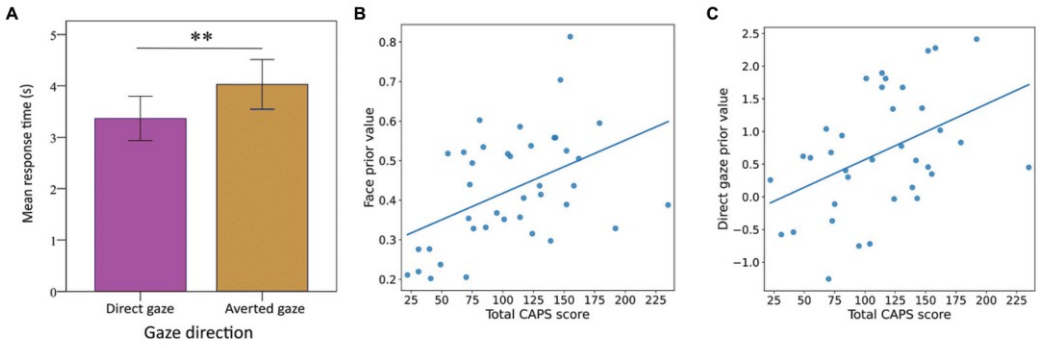


**FIGURE 3 |** Mean response times for direct and averted gaze in the unconscious gaze task **(A)**. Consistent with previous studies, response times were significantly faster for direct gaze ($T$ = −4.362, $p$ < 0.001). Relationships between participants' hallucination proneness (CAPS scores) and signal detection bias in the face task **(B)** and direct gaze bias in the unconscious gaze task **(C)**. With growing hallucination proneness, participants show an increased readiness to detect faces in noise and actual face stimuli (rho = 0.500, $p$ = 0.001) as well as to unconsciously process direct gaze faster than averted gaze (rho = 0.424, $p$ = 0.010). $^{**}p$ < 0.01.

the specific sample of patients investigated by Seymour et al. (2016) which was a high functioning, stable medicated chronic sample with mean illness duration of more 23 years and mild symptomatology. In this context it is noteworthy, that in our study, the direct gaze prior was specific to hallucinations (contrary to delusions), which are absent in around one third of untreated schizophrenia patients (Sartorius et al., 1986) and are significantly further reduced by antipsychotic medication (Sommer et al., 2012). Another reason might lie in slightly different properties of the CFS tasks used in this study and in the study by Seymour et al. For instance, in the task used by Seymour et al., the mask intensity was gradually decreased after target stimulus fade-in, whereas in our task version, the mask intensity did not change. Gradual fade-out of the mask may render bCFS less sensitive for the detection of individual differences (Munkler et al., 2015). Nevertheless, the mean breakthrough times as well as gaze-dependent breakthrough time differences lie in a similar range in both studies, rendering major differences in task-elicited effects unlikely. Finally, it should be noted that Seymour et al. also report increased response time differences in patients, which however did not reach significance. While it is difficult to combine these results with our present findings in a formal Bayesian analysis due to different analyses (group analyses vs. correlational analyses), interpreting these results together suggests the possibility of increased processing of direct gaze related to hallucinations, while the evidence speaks against a specific association with delusions. However, for further clarification, a follow-up experiment would be necessary to investigate the unconscious direct gaze bias in acutely psychotic patients with hallucinations.

Our finding of a relationship between hallucination proneness and an increased prior for direct gaze in a masking task is of relevance for the ongoing debate about the processing stage, at which psychosis-typical perceptual alterations take effect (Berkovitch et al., 2017). Here, our finding speaks for an involvement of unconscious processing stages. In this context, the neural correlates of these psychosis-associated unconscious processing alterations remain subject to speculation. In a first study on the neural correlates of unconscious direct gaze bias using EEG and CFS, Yokoyama et al. (2013) found increased activity on the fronto-parietal, but not on the occipital electrodes, for invisible direct gaze compared to averted gaze. Using fMRI, Madipakkam et al. (2015) found decreased activation of the fusiform face area, superior temporal sulcus, amygdala, and intraparietal sulcus for invisible direct gaze and concluded that in these regions, lower levels of neural activity are sufficient to give rise to awareness for direct than for averted gaze. Both findings speak for a differential involvement of higher-level processing stages in the processing of invisible direct compared to averted gaze. Studies investigating psychosis-associated changes in the neural underpinnings of unconscious processing of gaze remain to be conducted.

It should be noted that, in our current work "prior" does not refer to experimentally manipulated information, but to an implicit expectation of socially meaningful signals (faces and direct gaze) in noisy and ambiguous stimuli. This is in contrast to a body of previous work, where prior information was experimentally varied and had to be balanced against (potentially contradictory) sensory information. Here, relationships between psychotic experiences and prior usage were less consistent. In some experiments, there was an increased prior usage with growing psychosis proneness (Corlett et al., 2019 for a review), which is in line with our current results, while other studies showed even porting *decreased* prior usage with growing psychosis proneness (Jardri et al., 2017; Stuke et al., 2018). This inconsistency goes well with the emerging understanding that different kinds of priors may be differentially affected in psychosis, and that alterations in perceptual inference go beyond a simple over- or underweighting of priors. In short, it is proposed, that strong "high-level" belief priors might compensate for weak "low-level" sensory priors (for detailed discussions, see Schmack et al., 2013; Sterzer et al., 2018; Heinz et al., 2019). In this framework, our results are consistent with the hypothesis of stronger high-level priors (i.e., an increased prior probability for the presence of faces and direct gaze in hallucination-prone individuals).

The present results raise the question of what processes might underlie the bias toward detecting socially relevant information we observed in hallucination proneness. In theory, three possibilities are conceivable here: first, the increased tendency to perceive faces and direct gaze could be based on a *general information processing* bias such as overhasty decision-making. However, this possibility has not been confirmed by previous work that failed to show a connection between jumping to conclusions in a non-perceptual task on the one hand, and hallucinations or jumping to erroneous perceptions in a perceptual task on the other (Bristow et al., 2014). An increased tendency to perceive faces and direct gaze could, second, represent specific changes in the *processing of sensory information*, or, third, even more specific changes in the *processing of socially relevant sensory information*. In our view, these latter two possibilities cannot be distinguished with certainty at the moment. Partos et al. (2016) reported a bias toward perceiving signals in visual noise using analyses that combined non-socially relevant stimuli (natural scences) and potentially socially relevant stimuli (cartoons with mostly anthropomorphised animals). In a second experiment, the content participants perceived in visual noise could be entered freely by the participants and participants reported "36% contained human faces or facial features, 25% animals or mythical creatures, 20% humanoid figures, 15% natural objects or scenes, and 4% other" (Partos et al., 2016). These figures suggest a predominance of socially relevant stimuli in the erroneous percepts, but do not rule out that this might be due to changes in the processing in sensory information in general. Bristow et al. (2014) expanded a classic auditory experiment (Bentall and Slade, 1985) into the visual domain. Here, the target stimulus was the word "who," which had been denoted as socially relevant in the original paper ("The word 'Who' was chosen because it is short, common and because it was considered that the word used should make some reference to the subject. Hallucinating individuals commonly hear their voices speaking to themselves or commenting about their own actions"; Bentall and Slade, 1985). Participants with current hallucinations showed an increased bias for perceiving this written or spoken word in visual and

auditory noise, respectively, in line with an altered processing of socially relevant sensory information in both the visual and auditory domain. However, because only socially relevant sensory stimuli were used, these findings do not rule out a more generic alteration in the processing of sensory information. Hence, future work using a detection task similar to ours but with socially irrelevant stimuli as an additional control condition will be necessary to pinpoint whether the visual processing alterations related to hallucinations represent generic sensory processing alterations, or are specific to socially relevant stimuli.

False alarms in detection tasks (perception of meaning from noise stimuli) have an intuitive "face validity" as experimental hallucination markers. Similarly, the preferential unconscious processing of direct gaze directly relates to the psychotic feeling of being stared at in public. Hence, as opposed to other common markers for psychosis proneness (e.g., a reduced EEG mismatch negativity; Naatanen et al., 2015; Erickson et al., 2016) and cognitive biases, such as jumping-to-conclusions (Dudley et al., 2016), the two tasks used here have an immediate connection to the phenomenology of psychosis and might serve as symptom-related markers for the severity of psychotic experiences. It might be a worthwhile endeavor to investigate the predictive power of these markers in further research. In clinical settings, an early response of psychosis-related markers after initiation of antipsychotic treatment might help to predict following treatment response. In preclinical research, similar detection-in-noise tasks might help to assess effects of pro- or anti-psychotic interventions in animal models. In any case, the development of suited experimental markers to monitor and predict the effect of psychosis-targeting interventions remains an important cornerstone for progressing our still limited understanding and treatment options for psychotic disorders.

A limitation of the present results is that we investigated correlates of psychosis proneness in healthy individuals only. While the psychosis continuum framework described in the introduction suggests that the relationships found are meaningful for clinical manifestations of psychosis, a follow-up study involving psychotic patients would be required for confirmation.

In summary, our results speak to an overly strong prior for socially meaningful information in people with psychotic experiences that extends beyond the domain of auditory perception and might also affect early unconscious stages of sensory processing.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee of the Charité, Universitätsmedizin Berlin. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

HS, KS, PS, and VW developed the experimental design. EK carried out the experiment. HS, EK, and KS performed the data analysis and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This project was supported by the German Federal Ministry of Education and Research within the framework of the e:Med research and funding concept (01ZX1404A to KS). HS and VW are participants in the Charité Clinical Scientist Program funded by the Charité – Universitätsmedizin, Berlin and the Berlin Institute of Health. These sources had no further role in study design, the collection analysis and interpretation of data, the writing of the report, and the decision to submit the paper for publication.

## ACKNOWLEDGMENTS

This manuscript has been released as a preprint at bioRxiv (Stuke et al., 2018).

## REFERENCES

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., and Friston, K. J. (2013). The computational anatomy of psychosis. *Front. Psychol.* 4:47. doi: 10.3389/fpsyt.2013.00047

Akechi, H., Stein, T., Senju, A., Kikuchi, Y., Tojo, Y., Osanai, H., et al. (2014). Absence of preferential unconscious processing of eye contact in adolescents with autism spectrum disorder. *Autism Res.* 7, 590–597. doi: 10.1002/aur.1397

Alderson-Day, B., Lima, C. F., Evans, S., Krishnan, S., Shanmugalingam, P., Fernyhough, C., et al. (2017). Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. *Brain* 140, 2475–2489. doi: 10.1093/brain/awx206

Barrantes-Vidal, N., Grant, P., and Kwapil, T. R. (2015). The role of schizotypy in the study of the etiology of schizophrenia spectrum disorders. *Schizophr. Bull.* 41(Suppl 2), S408–S416. doi: 10.1093/schbul/sbu191

Bell, V., Halligan, P. W., and Ellis, H. D. (2006). The Cardiff anomalous perceptions scale (CAPS): a new validated measure of anomalous perceptual experience. *Schizophr. Bull.* 32, 366–377. doi: 10.1093/schbul/sbj014

Bell, V., Halligan, P. W., Pugh, K., and Freeman, D. (2011). Correlates of perceptual distortions in clinical and non-clinical populations using the Cardiff anomalous perceptions scale (CAPS): associations with anxiety and depression and a re-validation using a representative population sample. *Psychiatry Res.* 189, 451–457. doi: 10.1016/j.psychres.2011.05.025

Bentall, R. P., and Slade, P. D. (1985). Reality testing and auditory hallucinations: a signal detection analysis. *Br. J. Clin. Psychol.* 24, 159–169. doi: 10.1111/j.2044-8260.1985.tb01331.x

Berkovitch, L., Dehaene, S., and Gaillard, R. (2017). Disruption of conscious access in schizophrenia. *Trends Cogn. Sci.* 21, 878–892. doi: 10.1016/j.tics.2017.08.006

Bristow, E., Tabraham, P., Smedley, N., Ward, T., and Peters, E. (2014). Jumping to perceptions and to conclusions: specificity to hallucinations and delusions. *Schizophr. Res.* 154, 68–72. doi: 10.1016/j.schres.2014.02.004

Chapman, L. J., Chapman, J. P., Kwapil, T. R., Eckblad, M., and Zinser, M. C. (1994). Putatively psychosis-prone subjects 10 years later. *J. Abnorm. Psychol.* 103, 171–183. doi: 10.1037/0021-843X.103.2.171

Corlett, P. R., Frith, C. D., and Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology* 206, 515–530. doi: 10.1007/s00213-009-1561-0

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., and Powers, A. R. 3rd (2019). Hallucinations and strong priors. *Trends Cogn. Sci.* 23, 114–127. doi: 10.1016/j.tics.2018.12.001

DeRosse, P., and Karlsgodt, K. H. (2015). Examining the psychosis continuum. *Curr. Behav. Neurosci. Rep.* 2, 80–89. doi: 10.1007/s40473-015-0040-7

Dudley, R., Taylor, P., Wickham, S., and Hutton, P. (2016). Psychosis, delusions and the "jumping to conclusions" reasoning bias: a systematic review and meta-analysis. *Schizophr. Bull.* 42, 652–665. doi: 10.1093/schbul/sbv150

Erickson, M. A., Ruffle, A., and Gold, J. M. (2016). A meta-analysis of mismatch negativity in schizophrenia: from clinical risk to disease specificity and progression. *Biol. Psychiatry* 79, 980–987. doi: 10.1016/j.biopsych.2015.08.025

Fanous, A., Gardner, C., Walsh, D., and Kendler, K. S. (2001). Relationship between positive and negative symptoms of schizophrenia and schizotypal symptoms in nonpsychotic relatives. *Arch. Gen. Psychiatry* 58, 669–673. doi: 10.1001/archpsyc.58.7.669

Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

Franck, N., Montoute, T., Labruyere, N., Tiberghien, G., Marie-Cardine, M., Dalery, J., et al. (2002). Gaze direction determination in schizophrenia. *Schizophr. Res.* 56, 225–234. doi: 10.1016/S0920-9964(01)00263-8

Galdos, M., Simons, C., Fernandez-Rivas, A., Wichers, M., Peralta, C., Lataster, T., et al. (2011). Affectively salient meaning in random noise: a task sensitive to psychosis liability. *Schizophr. Bull.* 37, 1179–1186. doi: 10.1093/schbul/sbq029

Hanssen, M., Bak, M., Bijl, R., Vollebergh, W., and van Os, J. (2005). The incidence and outcome of subclinical psychotic experiences in the general population. *Br. J. Clin. Psychol.* 44, 181–191. doi: 10.1348/014466505X29611

Heinz, A., Murray, G. K., Schlagenhauf, F., Sterzer, P., Grace, A. A., and Waltz, J. A. (2019). Towards a unifying cognitive, neurophysiological, and computational neuroscience account of schizophrenia. *Schizophr. Bull.* 45, 1092–1100. doi: 10.1093/schbul/sby154

Hoffman, R. E., Woods, S. W., Hawkins, K. A., Pittman, B., Tohen, M., Preda, A., et al. (2007). Extracting spurious messages from noise and risk of schizophrenia-spectrum disorders in a prodromal population. *Br. J. Psychiatry* 191, 355–356. doi: 10.1192/bjp.bp.106.031195

Hooker, C., and Park, S. (2005). You must be looking at me: the nature of gaze perception in schizophrenia patients. *Cogn. Neuropsychol.* 10, 327–345. doi: 10.1080/13546800444000083

Horga, G., and Abi-Dargham, A. (2019). An integrative framework for perceptual disturbances in psychosis. *Nat. Rev. Neurosci.* 20, 763–778. doi: 10.1038/s41583-019-0234-1

Jardri, R., Duverne, S., Litvinova, A. S., and Deneve, S. (2017). Experimental evidence for circular inference in schizophrenia. *Nat. Commun.* 8:14218. doi: 10.1038/ncomms14218

Jeffreys, H. (1998). *The theory of probability*. Oxford: Oxford University Press.

Jiang, Y., Costello, P., and He, S. (2007). Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychol. Sci.* 18, 349–355. doi: 10.1111/j.1467-9280.2007.01902.x

Kendler, K. S., McGuire, M., Gruenberg, A. M., O'Hare, A., Spellman, M., and Walsh, D. (1993). The roscommon family study. III. Schizophrenia-related personality disorders in relatives. *Arch. Gen. Psychiatry* 50, 781–788. doi: 10.1001/archpsyc.1993.01820220033004

Linscott, R. J., and van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychol. Med.* 43, 1133–1149. doi: 10.1017/S0033291712001626

Madipakkam, A. R., Rothkirch, M., Dziobek, I., and Sterzer, P. (2019). Access to awareness of direct gaze is related to autistic traits. *Psychol. Med.* 49, 980–986. doi: 10.1017/S0033291718001630

Madipakkam, A. R., Rothkirch, M., Guggenmos, M., Heinz, A., and Sterzer, P. (2015). Gaze direction modulates the relation between neural responses to faces and visual awareness. *J. Neurosci.* 35, 13287–13299. doi: 10.1523/JNEUROSCI.0815-15.2015

Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods. Instrum. Comput.* 36, 630–633. doi: 10.3758/bf03206543

Munkler, P., Rothkirch, M., Dalati, Y., Schmack, K., and Sterzer, P. (2015). Biased recognition of facial affect in patients with major depressive disorder reflects clinical state. *PLoS One* 10:e0129863. doi: 10.1371/journal.pone.0129863

Naatanen, R., Shiga, T., Asano, S., and Yabe, H. (2015). Mismatch negativity (MMN) deficiency: a break-through biomarker in predicting psychosis onset. *Int. J. Psychophysiol.* 95, 338–344. doi: 10.1016/j.ijpsycho.2014.12.012

Partos, T. R., Cropper, S. J., and Rawlings, D. (2016). You don't see what I see: individual differences in the perception of meaning from visual stimuli. *PLoS One* 11:e0150615. doi: 10.1371/journal.pone.0150615

Peters, E., Joseph, S., Day, S., and Garety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. delusions inventory (PDI). *Schizophr. Bull.* 30, 1005–1022. doi: 10.1093/oxfordjournals.schbul.a007116

Powell, M. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* 7, 155–162. doi: 10.1093/comjnl/7.2.155

Powers, A. R., Mathys, C., and Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* 357, 596–600. doi: 10.1126/science.aan3458

Rosse, R. B., Kendrick, K., Wyatt, R. J., Isaac, A., and Deutsch, S. I. (1994). Gaze discrimination in patients with schizophrenia: preliminary report. *Am. J. Psychiatry* 151, 919–921. doi: 10.1176/ajp.151.6.919

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., et al. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures. A preliminary report on the initial evaluation phase of the WHO collaborative study on determinants of outcome of severe mental disorders. *Psychol. Med.* 16, 909–928. doi: 10.1017/S0033291700011910

Schmack, K., Burk, J., Haynes, J. D., and Sterzer, P. (2016). Predicting subjective affective salience from cortical responses to invisible object stimuli. *Cereb. Cortex* 26, 3453–3460. doi: 10.1093/cercor/bhv174

Schmack, K., Gomez-Carrillo de Castro, A., Rothkirch, M., Sekutowicz, M., Rossler, H., Haynes, J. D., et al. (2013). Delusions and the role of beliefs in perceptual inference. *J. Neurosci.* 33, 13701–13712. doi: 10.1523/JNEUROSCI.1778-13.2013

Seymour, K., Rhodes, G., Stein, T., and Langdon, R. (2016). Intact unconscious processing of eye contact in schizophrenia. *Schizophr. Res. Cogn.* 3, 15–19. doi: 10.1016/j.scog.2015.11.001

Shevlin, M., McElroy, E., Bentall, R. P., Reininghaus, U., and Murphy, J. (2017). The psychosis continuum: testing a bifactor model of psychosis in a general population sample. *Schizophr. Bull.* 43, 133–141. doi: 10.1093/schbul/sbw067

Sommer, I. E., Slotema, C. W., Daskalakis, Z. J., Derks, E. M., Blom, J. D., and van der Gaag, M. (2012). The treatment of hallucinations in schizophrenia spectrum disorders. *Schizophr. Bull.* 38, 704–714. doi: 10.1093/schbul/sbs034

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/BF03207704

Stein, T., Hebart, M. N., and Sterzer, P. (2011a). Breaking continuous flash suppression: a new measure of unconscious processing during interocular suppression? *Front. Hum. Neurosci.* 5:167. doi: 10.3389/fnhum.2011.00167

Stein, T., Senju, A., Peelen, M. V., and Sterzer, P. (2011b). Eye contact facilitates awareness of faces during interocular suppression. *Cognition* 119, 307–311. doi: 10.1016/j.cognition.2011.01.008

Stein, T., and Sterzer, P. (2014). Unconscious processing under interocular suppression: getting the right measure. *Front. Psychol.* 5:387. doi: 10.3389/fpsyg.2014.00387

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., et al. (2018). The predictive coding account of psychosis. *Biol. Psychiatry* 84, 634–643. doi: 10.1016/j.biopsych.2018.05.015

Sterzer, P., Hilgenfeldt, T., Freudenberg, P., Bermpohl, F., and Adli, M. (2011). Access of emotional information to visual awareness in patients with major depressive disorder. *Psychol. Med.* 41, 1615–1624. doi: 10.1017/S0033291710002540

Stuke, H., Kress, E., Weilnhammer, V. A., Sterzer, P., and Schmack, K. (2018). Overly strong priors for socially meaningful visual signals in psychosis proneness. bioRxiv, 473421 [Preprint]. doi: 10.1101/473421

Stuke, H., Stuke, H., Weilnhammer, V. A., and Schmack, K. (2017). Psychotic experiences and overhasty inferences are related to maladaptive learning. *PLoS Comput. Biol.* 13:e1005328. doi: 10.1371/journal.pcbi.1005328

Stuke, H., Weilnhammer, V. A., Sterzer, P., and Schmack, K. (2018). Delusion proneness is linked to a reduced usage of prior beliefs in perceptual decisions. *Schizophr. Bull.* 45, 80–86. doi: 10.1093/schbul/sbx189

Tienari, P., Wynne, L. C., Laksy, K., Moring, J., Nieminen, P., Sorri, A., et al. (2003). Genetic boundaries of the schizophrenia spectrum: evidence from the Finnish adoptive family study of schizophrenia. *Am. J. Psychiatry* 160, 1587–1594. doi: 10.1176/appi.ajp.160.9.1587

Tso, I. F., Mui, M. L., Taylor, S. F., and Deldin, P. J. (2012). Eye-contact perception in schizophrenia: relationship with symptoms and socioemotional functioning. *J. Abnorm. Psychol.* 121, 616–627. doi: 10.1037/a0026596

Tsuchiya, N., and Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat. Neurosci.* 8, 1096–1101. doi: 10.1038/nn1500

van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., and Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol. Med.* 39, 179–195. doi: 10.1017/S0033291708003814

Vercammen, A., de Haan, E. H., and Aleman, A. (2008). Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychol. Med.* 38, 1177–1184. doi: 10.1017/S0033291707002437

Welham, J., Scott, J., Williams, G., Najman, J., Bor, W., O'Callaghan, M., et al. (2009). Emotional and behavioural antecedents of young adults who screen positive for non-affective psychosis: a 21-year birth cohort study. *Psychol. Med.* 39, 625–634. doi: 10.1017/S0033291708003760

Yokoyama, T., Noguchi, Y., and Kita, S. (2013). Unconscious processing of direct gaze: evidence from an ERP study. *Neuropsychologia* 51, 1161–1168. doi: 10.1016/j.neuropsychologia.2013.04.002

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## *2.6 Maladaptives Lernen und reduzierte Suppression unwahrscheinlicher Informationen bei Psychoseneigung*

Wie in Abschnitt 1.3 dargelegt, werden psychotische Symptome in theoretischen Modellen als Resultat einer veränderten Anpassung von Vorannahmen durch neue Informationen (d.h., einer pathologischen Veränderung von Lernprozessen) konzipiert. Insbesondere kann demzufolge eine Übergewichtung von eigentlich unwahrscheinlichen (und damit potenziell irrelevanten) Informationen zu einer fehlerhaften Anpassung von Vorannahmen führen und damit die Neigung zu einer psychotischen Verarbeitung sensorischer Informationen verstärkt werden. In dieser Studie versuchten wir, diesen vermuteten Zusammenhang zwischen maladaptivem Lernen und Psychose empirisch zu testen. Hierfür verwendeten wir ein neu entwickeltes computationales Lernmodell für den unter 1.4 geschilderten beads task.

Wortgetreu und selbstständig übersetztes Abstract des Originalartikels (Stuke, H, Stuke, H, Weilnhammer, V A, & Schmack, K. Psychotic Experiences and Overhasty Inferences Are Related to Maladaptive Learning. PLoS Comput Biol 2017; 13(1): e1005328. doi:10.1371/journal.pcbi.1005328):

*„Theoretische Überlegungen legen nahe, dass eine Veränderung von Lernmechanismen im Gehirn zu übereilten Schlussfolgerungen und damit zu psychotischen Symptomen führen könnte. In dieser Studie wurde versucht, den vermuteten Zusammenhang zwischen maladaptivem Lernen und Psychosen zu ergründen. Achtundneunzig gesunde Personen mit unterschiedlichem Grad von Wahnvorstellungen und halluzinatorischen Erfahrungen führten eine probabilistische Schlussfolgerungsaufgabe durch, mit der wir übereilte Schlussfolgerungen quantifizieren konnten. Passend zu früheren Studienergebnissen fanden wir einen Zusammenhang zwischen psychotischen Erfahrungen und übereilten Schlussfolgerungen. Die computergestützte Modellierung ergab, dass die Verhaltensdaten am besten durch ein neuartiges Lernmodell erklärt werden können, das die Adaptivität des Lernens durch eine nichtlineare Verzerrung der Verarbeitung von Vorhersagefehlern formalisiert, wobei eine zunehmende Nichtlinearität eine wachsende Widerstandsfähigkeit gegen das Lernen aus überraschenden und daher unzuverlässigen Informationen (große Vorhersagefehler) impliziert. Eine verminderte Adaptivität des Lernens sagte wahnhafte Vorstellungen und halluzinatorische Erfahrungen vorher. Unsere aktuellen Ergebnisse liefern eine formale Beschreibung der computationalen*

*Mechanismen, die übereilten Schlussfolgerungen zugrunde liegen und untermauern damit empirisch Theorien, die Psychosen mit maladaptivem Lernen in Verbindung bringen.“*

Im Kontext der Bayesianischen Psychosemodelle lässt sich dies sowohl als eine potenziell dysfunktionale Anpassung von Vorannahmen, als auch als eine reduzierte „Korrektur“ unwahrscheinlicher neuer Information durch Vorannahmen (im Sinne der weak prior Hypothese) interpretieren.

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

# Psychotic Experiences and Overhasty Inferences Are Related to Maladaptive Learning

**Heiner Stuke[1]*, Hannes Stuke[2], Veith Andreas Weilnhammer[1], Katharina Schmack[1]**

**1** Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin Berlin Berlin, Germany,
**2** Department of Mathematics, Freie Universität Berlin, Berlin, Germany

* heiner.stuke@charite.de



OPEN ACCESS

**Citation:** Stuke H, Stuke H, Weilnhammer VA, Schmack K (2017) Psychotic Experiences and Overhasty Inferences Are Related to Maladaptive Learning. PLoS Comput Biol 13(1): e1005328. doi:10.1371/journal.pcbi.1005328

**Editor:** Wolfgang Einhäuser, Technische Universitat Chemnitz, GERMANY

**Received:** August 8, 2016

**Accepted:** December 20, 2016

**Published:** January 20, 2017



**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This project was supported by the German Federal Ministry of Education and Research within the framework of the e:Med research and funding concept (01ZX1404A to KS). KS is participant in the Charité Clinical Scientist Program funded by the Charité Universitätsmedizin Berlin and the Berlin Institute of Health. The funders had no role in study design, data collection

## Abstract

Theoretical accounts suggest that an alteration in the brain's learning mechanisms might lead to overhasty inferences, resulting in psychotic symptoms. Here, we sought to elucidate the suggested link between maladaptive learning and psychosis. Ninety-eight healthy individuals with varying degrees of delusional ideation and hallucinatory experiences performed a probabilistic reasoning task that allowed us to quantify overhasty inferences. Replicating previous results, we found a relationship between psychotic experiences and overhasty inferences during probabilistic reasoning. Computational modelling revealed that the behavioral data was best explained by a novel computational learning model that formalizes the adaptiveness of learning by a non-linear distortion of prediction error processing, where an increased non-linearity implies a growing resilience against learning from surprising and thus unreliable information (large prediction errors). Most importantly, a decreased adaptiveness of learning predicted delusional ideation and hallucinatory experiences. Our current findings provide a formal description of the computational mechanisms underlying overhasty inferences, thereby empirically substantiating theories that link psychosis to maladaptive learning.

## Author Summary

Predictive coding theories represent a unifying account of psychosis, stating that the central psychosis-related alteration affects the interplay between prior predictions and incoming information. Since every incoming information is imprecise and potentially allows for different interpretations, prior expectations achieve the enforcement of interpretations with a higher prior probability. Disturbances in this basic framework might let unlikely interpretations come into effect, resulting in proneness for delusions and hallucinations. Here, we contribute to these theories by devising a novel computational model for behavior in a reasoning task that quantifies the participants' readiness to draw inferences from very surprising information. We thereby demonstrate that precisely this increased learning from surprising and thus potentially spurious information, as opposed to non-specific alterations in the general learning speed, predispose healthy individuals for delusions and

and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

hallucinations. The present results hence speak for the hypothesis that hallucinations and delusions arise when noisy information is considered as precise and is thus not suppressed by opposing prior beliefs. In this sense, our findings also tie with recent neurophysiological models of psychosis that posit aberrations in modulatory neurotransmitters such as dopamine (or its interactions with GABAergic interneurons) as a correlate of perturbed computations of information precision in the cortex.

## Introduction

Psychotic symptoms are a core symptom of devastating psychiatric disorders such as schizophrenia. They comprise many different kinds of experiences, among others beliefs that are unfounded in the external reality (delusions), and percepts in the absence of a causative stimulus (hallucinations). Accordingly, it poses a key challenge to theoretically and empirically establish models that can capture the multifariousness of psychotic experiences by a few (or even one) core alterations.

Influential theories [1–3] explain psychotic symptoms in the framework of predictive coding [4–6]. According to predictive coding, one central challenge for the brain is to draw inferences about the state of the external world from incoming information of relatively poor quality. It is stated that the brain deals with this challenge by recurring to predictive beliefs about the world. Such predictive beliefs are proposed to shape incoming information via top-down signals, thereby enabling stable and unitary inferences from imprecise and ambiguous information and constituting a protection against an over-interpretation of sporadically occurring irrelevant information. Importantly, predictive beliefs are assumed to be continuously updated by prediction errors. Such prediction errors are thought to drive learning via bottom-up signals, and to arise when predictive beliefs do not precisely match incoming information. Hence, ongoing learning in response to surprising information is thought to ensure the flexible adaptation of belief-dependent inferences.

Along these lines, psychotic symptoms can be framed as maladaptive learning that occurs if irrelevant information is considered as surprising and relevant due to altered prediction error signaling [1,7,8]. As a result, no stable and valid predictive beliefs would be built up and the brain would become susceptible to overhasty and erroneous inferences yielding delusions and hallucinations. In line with the idea that overhasty and erroneous perceptual inferences from irrelevant noise information are implicated in hallucinations and hallucination-proneness, hallucinatory experiences have been repeatedly associated with a greater tendency to perceive illusory contents in auditory noise [9,10]. Moreover, delusional ideation has been consistently linked to "jumping to conclusions" (JTC, see [11–13] for detailed meta-analyses), a cognitive reasoning bias that leads to a rash acceptance of hypotheses based on little evidence. However, it is a matter of ongoing debate, which particular cognitive alteration predisposes delusional and delusion-prone individuals for an overhasty acceptance of possible hypotheses [14–17]. With regard to the predictive coding account of psychosis outlined above, we suggest that JTC might reflect a pivotal alteration underlying psychotic symptoms, namely maladaptive learning from irrelevant information, leading to overhasty inferences.

To empirically test the claim that maladaptive learning contributes to psychotic symptoms, one will necessarily have to tackle the question of what constitutes adaptive learning, or, in other words, how *non-psychotic* individuals can generate and adapt beliefs sufficiently quickly in response to relevant information, and, nevertheless, resist inadequate belief revision due to irrelevant noise. Common computational learning models (e.g., [18]) formalize learning in terms of prediction errors and learning rates. Here, the current belief is obtained as a function

of the prediction error that denotes the difference between the expectation (i.e., the belief before the actual observation) and the actual observation. The magnitude of this prediction error multiplied with a subject-specific learning rate determine the degree to which the belief is updated (i.e., the learning). An alternative formulation of evidence accumulation (and state estimation) calls on Bayesian filtering schemes as metaphors for neuronal computations. These schemes accumulate evidence for hidden states of the world in proportion to their estimated precision or reliability. The most celebrated Bayesian filter is called the Kalman filter, where the Kalman gain corresponds to the relative precision (inverse variance) of sensory evidence in relation to prior beliefs. Biologically plausible implementations of Kalman filtering include predictive coding, where Bayesian belief updating (i.e., evidence accumulation) is mediated by precision weighted prediction errors. In short, Rescorla Wagner models, Bayesian filtering and predictive coding are all equivalent formulations of evidence accumulation (see [19]). They all speak to the importance of precision as learning rates in modulating the impact of prediction errors on belief updating, which we will refer to as adaptive learning.

Thus, these common computational learning models capture adaptive learning, as opposed to maladaptive learning from irrelevant information that lead to overhasty and erroneous inferences, by small learning rates. Hence, resilience against irrelevant information would be formalized by smaller learning rates and thus comes at the expense of a generally decreased speed of learning (see Fig 1). Here, we propose a novel computational learning model that is able to capture the resilience against irrelevant information without substantially impairing the general speed of learning. The central and very simple idea of our model is that prediction errors are processed in a non-linear fashion. Concretely, we introduce a saturating non-linear function of prediction error that attenuates the effect of very large prediction errors on belief updating, relative to smaller prediction errors. Effectively, this means that very surprising or large prediction errors are treated as imprecise information; very much in the same way that we discard outliers in statistical analyses of data. In the technical literature this is known as Winsorizing and represents one of the simplest and most fundamental modifications of linear predictive coding. Formally, this compressive non-linearity can be considered a hyperprior that certain prediction errors are generated by a class of outliers that can be construed as "irrelevant". In other words, the non-linearity enables the accumulation of evidence in a way that is resistant to the effect of spurious (i.e., very surprising) events. Importantly, learning from small prediction errors is preserved, leading to adaptive inferences in response to moderately surprising and hence relevant information. Thus, our model captures the resilience against irrelevant information, and hence overhasty and erroneous inferences, by the non-linearity of prediction error processing (see Fig 1). Conversely, we would predict that a weaker resilience against irrelevant information that leads to overhasty and erroneous inferences in psychotic and psychosis-prone individuals is paralleled by a more linear processing of prediction errors.

In this work, we sought to devise a formal approach to assess and quantify the maladaptive learning mechanisms underlying overhasty and erroneous inferences related to psychotic symptoms. To this end, we devised an adapted probabilistic reasoning task that allowed us to continuously track participants' belief trajectories. We then used this task to quantify overhasty inferences in a sample of healthy individuals with varying degrees of delusional ideation and hallucinatory experiences, based on the view that clinically relevant psychotic symptoms represent an extreme of a trait continuously distributed in the general population [20,21]. In order to investigate the computational mechanisms underlying psychosis-related biases in learning and inference, we fitted the behavioral data with our novel learning model that quantifies the adaptiveness of learning by a non-linear prediction error processing. We hypothesized that psychosis-related experiences would inversely relate to the resilience against irrelevant information quantified by the non-linearity of prediction error processing.

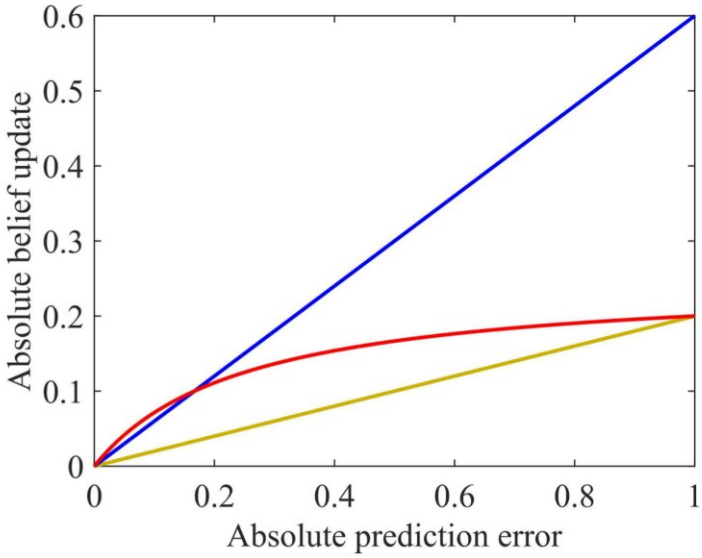


**Fig 1. Relationship between prediction error (x-axis) and belief update (y-axis).** Linear relationships with a high learning rate in blue and with low learning rate in yellow, non-linear relationship in red. We can see that although achieving a resilience against irrelevant information (attenuation of high prediction errors) comparable to the slow-learning agent in yellow, the non-linear red agent learns from small prediction errors similarly to the fast learning blue agent. Two hypotheses regarding the learning alterations that lead to hasty inferences and psychotic experiences may be suggested: Firstly, increased psychosis-proneness might be linked to a generally increased learning speed that predisposes for unfounded cognitive and perceptual inferences. According to this hypothesis, a psychosis-prone individual would behave like the blue as compared to the yellow agent (i.e., show an increased learning rate). Secondly, psychosis-proneness might be linked to a specifically decreased attenuation of large prediction errors (that can be interpreted as a reduced resilience against irrelevant and strongly surprising noise information). According to this hypothesis, a psychosis-prone individual would behave like the blue as compared to the red agent (i.e., show a decreased non-linearity of the relationship between prediction error and learning).

doi:10.1371/journal.pcbi.1005328.g001

## Methods

### Participants and psychometric assessments

Ninety-eight healthy individuals from the general population were recruited for study participation through advertising. The study was approved by the Ethical Committee of the Charité, Universitätsmedizin Berlin. Participants who received treatment due to psychiatric diseases were excluded. After complete description of the study to the participants, written informed consent was obtained in accordance with the Declaration of Helsinki of 1975 before participation.

The participants' tendency towards delusional ideation was quantified using the Peters Delusion Inventory (PDI, [22]). The 40 items of this self-rating questionnaire cover a wide range of delusional convictions, including beliefs in the paranormal, grandiosity ideas or suspicious thoughts. For every endorsed belief, the questionnaire asks for dimensional ratings on the degree of belief-related distress, preoccupation and conviction. The total score obtained by adding up these three dimensional ratings was used for analyses.

Additionally, proneness to hallucinatory experiences was assessed with the Cardiff anomalous perception scale (CAPS, [23]). This 32-item self-rating scale assesses anomalous

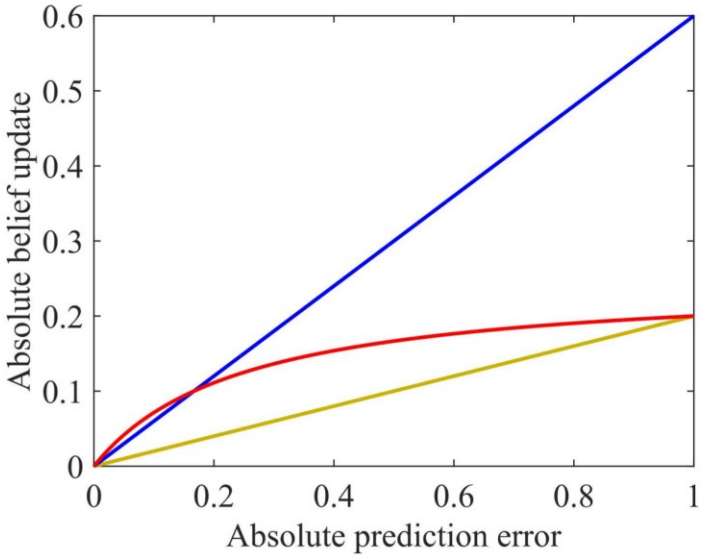


**Fig 1. Relationship between prediction error (x-axis) and belief update (y-axis).** Linear relationships with a high learning rate in blue and with low learning rate in yellow, non-linear relationship in red. We can see that although achieving a resilience against irrelevant information (attenuation of high prediction errors) comparable to the slow-learning agent in yellow, the non-linear red agent learns from small prediction errors similarly to the fast learning blue agent. Two hypotheses regarding the learning alterations that lead to hasty inferences and psychotic experiences may be suggested: Firstly, increased psychosis-proneness might be linked to a generally increased learning speed that predisposes for unfounded cognitive and perceptual inferences. According to this hypothesis, a psychosis-prone individual would behave like the blue as compared to the yellow agent (i.e., show an increased learning rate). Secondly, psychosis-proneness might be linked to a specifically decreased attenuation of large prediction errors (that can be interpreted as a reduced resilience against irrelevant and strongly surprising noise information). According to this hypothesis, a psychosis-prone individual would behave like the blue as compared to the red agent (i.e., show a decreased non-linearity of the relationship between prediction error and learning).

doi:10.1371/journal.pcbi.1005328.g001

## Methods

### Participants and psychometric assessments

Ninety-eight healthy individuals from the general population were recruited for study participation through advertising. The study was approved by the Ethical Committee of the Charité, Universitätsmedizin Berlin. Participants who received treatment due to psychiatric diseases were excluded. After complete description of the study to the participants, written informed consent was obtained in accordance with the Declaration of Helsinki of 1975 before participation.

The participants' tendency towards delusional ideation was quantified using the Peters Delusion Inventory (PDI, [22]). The 40 items of this self-rating questionnaire cover a wide range of delusional convictions, including beliefs in the paranormal, grandiosity ideas or suspicious thoughts. For every endorsed belief, the questionnaire asks for dimensional ratings on the degree of belief-related distress, preoccupation and conviction. The total score obtained by adding up these three dimensional ratings was used for analyses.

Additionally, proneness to hallucinatory experiences was assessed with the Cardiff anomalous perception scale (CAPS, [23]). This 32-item self-rating scale assesses anomalous

perceptual experiences in different sensory domains like proprioception, time perception, somatosensation and visual and auditory perception. The intensity of every anomalous perception is quantified from one to five on subscales for intrusiveness, frequency and distress. Again, the total score was calculated by adding up all subscore ratings and used for analyses.

## Probabilistic reasoning task

An adapted version of the "beads task" [24] was used to assess psychosis-related alterations in probabilistic reasoning, especially overhasty inferences such as the JTC bias. In the beads task, beads are continuously drawn from one of two different urns that contain different numerical proportions of different kinds of beads. The participants have to infer from which urn beads are currently being drawn based on their knowledge about the numerical proportions of different kinds of beads in the two urns and the number of already drawn beads of each kind. The task thus implies a continuous update of the belief about the correct urn with every new draw, which can be either consistent with the current belief about the correct urn (relevant information) or inconsistent with it (irrelevant information).

In our version of this task, the participants were shown pictures of two different lakes (a "mountain lake" and a "flatland lake") and told that these lakes are home to a different proportion of carps and trouts with the mountain lake containing 70% carps and 30% trouts and the flatland lake 30% carps and 70% trouts. For reasons of simplicity, we will refer to the mountain lake as the "carp lake" and to the flatland lake as the "trout lake" in the following.

The task was structured in 30 rounds with a varying number of draws. On each round, fishes were sequentially angled from one of the two lakes and the participants were instructed to evaluate from which of the lakes the fishes were more likely angled in this round using the number of so far angled carps and trouts and their knowledge about the numerical proportion of fishes in the two lakes (thus with every angled carp making the carp lake more probable and every angled trout making the trout lake more probable). Moreover, participants were told that both lakes contained so many carps and trouts that the numerical proportions did not change due to the fishing.

Each round started with only one angled fish and, accordingly, with a rather imprecise information about which of the two lakes being correct in this round. To gain further information, participants were allowed to make new draws until they felt confident enough to make a final decision about the correct lake in this round (Fig 2). With every new draw, one new fish was angled and the number of so far angled trouts and carps was updated and presented. After each draw, the participants indicated their new belief about from which lake the fishes were probably angled in this round. For this purpose, they entered their guess and its certainty using the mouse on a visual scale (ranging from absolute certainty of the carp lake being correct at the very left to absolute certainty of the trout lake being correct at the very right with positions close to the center indicating uncertainty). In this way, we obtained a continuous assessment on the participants' current belief for each draw. After having placed their guess, the participants were asked if they wanted to commit themselves to the given response on the correct lake (by pressing either the up or the down arrow key). If they did not commit to their response, a new fish was angled (new draw). If they committed to their response, a final decision on the lake was made and a new round started with once again only one angled fish and accumulating evidence with every further draw.

To induce prediction errors even in rounds with few draws, we added a prior information about the lakes' probabilities in the form of a high- or low-pitched tone that was played shortly before every newly angled fish. In one round, always the same tone pitch was played. If the fishes were angled from the carp lake in the round, the high-pitched tone was played more

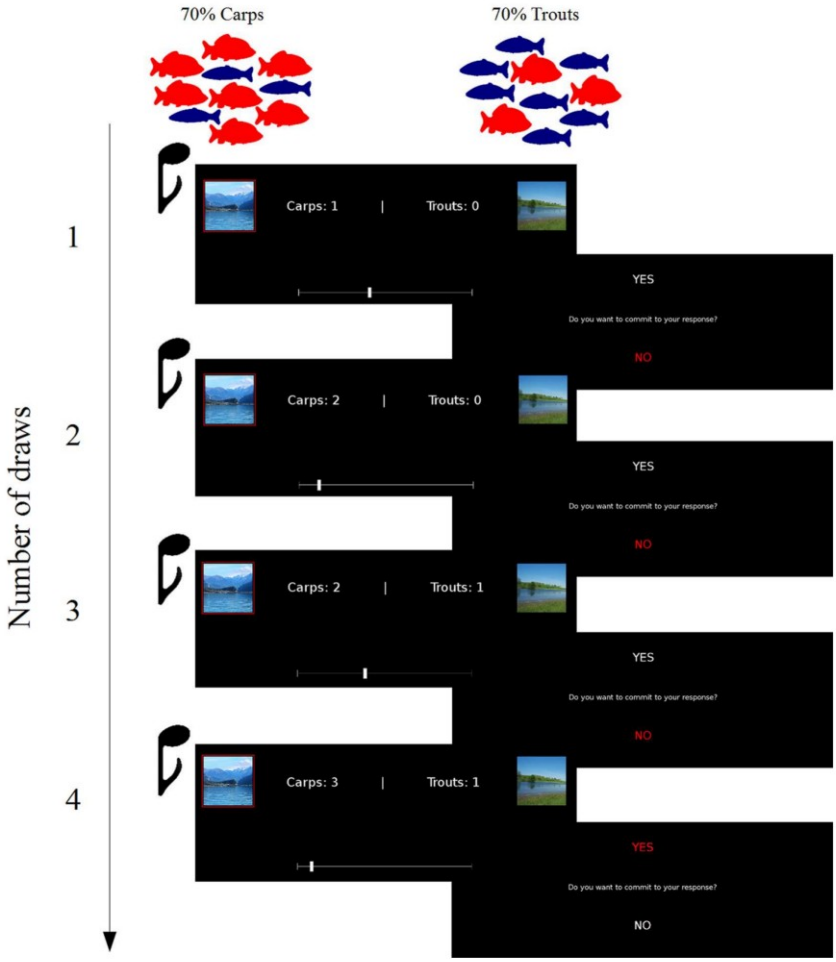


**Fig 2. Experimental sequence of the probabilistic reasoning task.** To determine, from which of the two possible lakes fishes were being angled, the participants used their knowledge about the numerical proportions of fishes in the lakes (70% carps and 30% trouts in the carp lake and vice versa in the trout lake) and the number of so far angled carps and trouts. Prior information about the lakes' probabilities was given by a tone that was–dependent on the pitch–in 80% of the cases associated with the one or the other lake. After each draw, the current belief about the correct lake was indicated on a continuous response bar. Subsequently, the participants decided if they want to make a final decision on the lake (commit to their response) or if they want to gain further information in the form of newly angled fishes (i.e., making a new draw).

doi:10.1371/journal.pcbi.1005328.g002

frequently (80%) and if the fishes were angled from the trout lake, the low-pitched tone was played more frequently (80%). Thus, the tone pitch constituted a probabilistic initial information about the lake probabilities for each round. The associations between tone pitch and lake probabilities were learned in a preceding learning run of 15 rounds and did not change throughout the experiment. By these means, we could assess prediction errors already in the

first draw (one angled fish) and increase the variance of prediction error values occurring throughout the course of the experiment.

## Relationship between jumping-to-conclusions, delusional convictions and anomalous perceptions

To quantify the tendency towards overhasty inferences in each participant, we calculated the mean number of draws a participant needed on each round before committing to a final decision. This measure ("draws to decision") is an accepted measure for the JTC bias found to be associated with psychotic symptoms [14].

To replicate prior findings that participants with growing psychosis proneness tend to exert jumping-to-conclusions (see introduction), we tested associations between the participants' draws to decision and the tendency towards delusional convictions (PDI scores) as well as the proneness to hallucinatory experiences (CAPS scores) in two different ways. Firstly, as suggested by [25], we performed a binary analysis with our sample separated into two groups (with and without JTC). There were only six of 94 participants showing JTC according to the commonly applied threshold of two draws to decision, probably due to the differing set-up of our adapted version of the lake task (introduction of prior knowledge associated with the tone, usage of continuous response bar). Thus, we used a slightly higher threshold and considered participants in the lowest quartile of draws to decision (i.e., with an average of 3.2 or less draws to decision) as exhibiting JTC and compared their PDI and CAPS scores with the remaining (non-JTC) sample. Secondly, we investigated continuous relationships by correlating PDI and CAPS scores with the mean number of draws to decision.

Since the distribution of PDI and CAPS scores differed significantly from a normal distribution ($Z = 1.374$, $p = 0.046$ for PDI scores and $Z = 1.941$, $p = 0.001$ for CAPS scores, one-sample Kolmogorov-Smirnov tests), we used non-parametric Mann-Whitney tests for the first (categorical) and Spearman rank correlations for the second (correlational) analysis.

To our knowledge, there are no previous studies reporting associations between hallucinations and JTC, giving our analysis on relationships between CAPS scores and JTC a rather exploratory character. Nevertheless, because we tested associations between JTC and both PDI and CAPS scores, we report among uncorrected p values also p values with adjustment for multiple testing (tests for PDI and CAPS scores, e.g., two tests with correlated outcomes). To this end, p values were adjusted according to the approach proposed by [26] and outlined in [27] for multiple comparisons with correlated outcomes.

## Computational modeling

By fitting the behavioral data with computational learning models, we aimed at quantifying the resilience against irrelevant information and thereby assessing the adaptiveness of learning, which we expected to be inversely related to psychotic symptoms.

Two computational models were designed to track the participants' trajectories of belief in the probabilistic reasoning task. Firstly, we applied a conventional linear prediction-error-based learning model (e.g., [28]). Secondly, we developed a novel model that enabled the quantification of the participants' resilience against irrelevant information through a non-linear relationship between prediction error and learning, which we expected to provide a more precise description of adaptive learning in probabilistic reasoning.

In both models, the participants' beliefs about the correct lake were captured on a trial-by-trial basis as a continuous value between 0 (certainty that the "carp lake" is correct) and 1 (certainty that the "trout lake" is correct). Thus, the high-pitched tone as well as newly angled carps brought the belief nearer to the 0 and the low-pitched tone as well as newly angled trouts

nearer to the 1. Eq 1 shows accordingly, that the neutral belief of 0.5 was initially shifted towards 1 in case of the "trout-lake"-associated low-pitched tone and towards 0 in case of the "carp-lake"-associated high-pitched tone and that the magnitude of the tone-dependent belief shift depended on the subject-specific parameter θ. Since the neutral belief of 0.5 could be shifted by maximally 0.5 by the tone, we used a uniform distribution between 0 and 0.5 as a prior distribution for the estimation of θ values based upon choice behavior.

$$\text{Initial tone} - \text{dependent belief } (\pm: \ + \text{ if trout is angled}, \ - \text{ if carp is angled}).$$
$$b_1 = 0.5 + \theta \qquad \text{Eq1}$$

Whereas this initial tone-dependent belief was calculated in the same way in both models, the effect of newly angled fishes differed between the conventional linear and our novel non-linear model. In the linear model, the prediction error determined the learning linearly. Eq 2 shows that the belief update here depended on the non-modified prediction error $b_{i-1} - o_i$ (difference between the former belief $b_{i-1}$ and the current observation $o_i$) that was multiplied with a subject-specific constant learning rate α that captures the general rapidity of belief generation regardless of the typicality of the new information. Since the learning rate is naturally bounded between 0 and 1, we used a uniform distribution between 0 and 1 as a prior distribution for estimation of α values based upon choice behavior.

$$\text{Linear belief update } (\text{bi}: \text{ current belief}, \ \text{oi}: \text{ current (binary) observation } (1 \text{ if trout is angled},$$
$$0 \text{ if carp is angled}), \ \pm: \ + \text{ if trout is angled}, \ - \text{ if carp is angled}).$$
$$b_i = b_{i-1} \pm \alpha * (b_{i-1} - o_i) \qquad \text{Eq2}$$

In the non-linear model on the other hand, the learning depended on the prediction error with a varying degree of non-linearity expressed by the non-linearity parameter ζ. Please note that high values of ζ imply a marked non-linearity / flattening of the relationship between prediction error and learning, whereas this relationship is linear for ζ = 0. Thus, high values of ζ imply a strong resilience against irrelevant information, since high prediction errors have a reduced impact on learning in this case: Hence, this modulation can be thought of as a dynamic learning rate that adaptively decreases if information is unreliable and potentially irrelevant. As in common behavioral learning models, the resulting non-linear learning term was multiplied with a subject-specific constant learning rate α that captures the general rapidity of belief generation regardless of the typicality of the new information. Eq 3 shows how the current belief $b_i$ is updated depending on the learning rate α and the prediction error $b_{i-1} - o_i$, whose impact on the learning decreases with increasing values of ζ. Compared to other possible implementations of a non-linear prediction error, the definition outlined above has the advantage of yielding one simple parameter that determines the degree of non-linearity and is zero for an entirely linear relationship between prediction error and learning. Furthermore and importantly, it cannot generate overshooting beliefs below zero or above one without having to assume an additional softmax transformation (see proof in the Supplementary Material). Fig 1 shows exemplary relationships between prediction error and learning, with linear relationships (ζ = 0) in blue (α = 0.6) and yellow (α = 0.2) and a non-linear relationship in red with ζ = 4 and α = 1. Since such a non-linear prediction error processing has to our knowledge not been implemented so far, we used a uniform distribution between 0 and 5 (thus allowing for a wide range of non-linearity) as a prior distribution for estimation of ζ values based upon

choice behavior.

Non − linear belief update (bi : current belief, oi : current (binary) observation (1 if trout is angled, 0 if carp is angled), ± : + if trout is angled, − if carp is angled).

$$b_i = b_{i-1} \pm \alpha * \frac{1}{\zeta + \frac{1}{(b_{i-1} - o_i)}} \quad \text{Eq3}$$

Both models were applied to explain the trajectory of the participants' beliefs about the correct lake throughout the course of the experiment. For this purpose, the trial-by-trial belief indicated on the continuous response bar was scaled between 0 and 1, yielding the trajectory of belief vector g. Subsequently, each participants' trajectory of belief g was fitted with both models using the VBA Toolbox for Matlab [29]. This approach uses Variational Bayesian methods to estimate the parameter values of our two models for which the trajectory of the belief b predicted by the model optimally traces the real belief g indicated by the participants (Fig 3). Furthermore, the (lower bound on the) model's evidence (marginal likelihood), i.e., the likelihood that the real trajectory of belief g could have been generated by the respective model, was computed and used for model comparison (see below).

Summing up, the following parameters were estimated to optimally model the participants' behavior:

θ: Tone-dependent initial (i.e., prior) belief

α: General learning rate

ζ: Non-linear prediction error processing (resilience against irrelevant information, $\zeta = 0$ in case of the linear model)

## Model comparison

To test if the non-linear model that allowed for a non-linear relationship between prediction error and learning explained the participants' behavior better than the conventional linear model, we performed a formal Bayesian model comparison between the two models. Therefore, both models were used to fit the participants' continuous belief trajectories and the resulting model evidences were compared using the approach outlined in [30] and implemented in Statistical Parametric Mapping 12 (SPM 12). In this approach, the ability of a model to accurately predict the participants behavior is balanced against its complexity, where growing model complexity is punished. In our case, this means that if the non-linear model proves to be superior in the model comparison, the growing complexity which results from the inclusion of the additional non-linearity parameter is overcompensated by the gain in accuracy afforded by it. In addition to formal model comparison, we calculated the explained variance $R^2$ of each participants' belief trajectory by the model in order to obtain a clear measure of how well the models were able to capture the participants' behavior. Please note that in contrast to Bayesian model comparison as described above, this assessment of model fit does not take into account the model complexity and is therefore not an appropriate measure for formal model comparison.

## Learning and usage of the tone parameter

The introduction of the tone as prior information about the lakes' probabilities allowed us to assess prediction errors even in participants with few draws to decision (see above). However, it has to be ensured, that the tone meanings has been learned by the participants during the

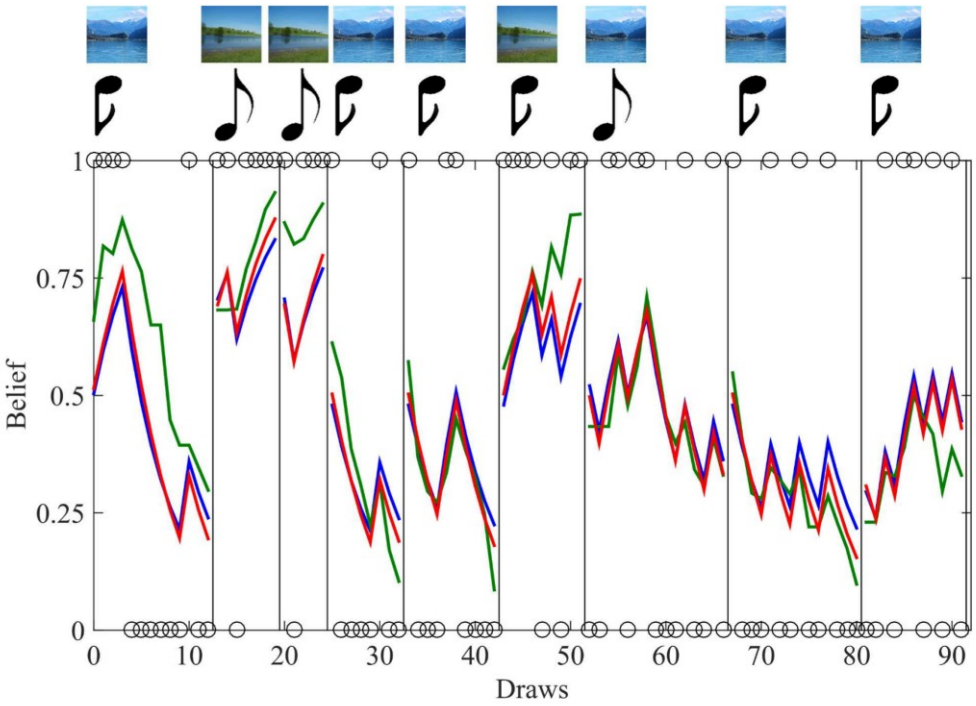


**Fig 3. Sample trajectory of belief for nine rounds in one exemplary participant.** The participant's belief, i.e., the probability with which the participant considered the one or the other lake as correct (indicated on the response bar), is shown in green, the predicted belief by the non-linear model in red and by the linear model in blue. The upper bound (belief = 1) indicates absolute certainty that the trout lake is correct in this round, whereas the lower bound (belief = 0) indicates absolute certainty that the carp lake is correct. Circles on the upper bound mean that a trout and on the lower bound that a carp has been angled in this draw. Vertical lines indicate that a decision on the lake has been made and a new round started. If a decision was made at a current belief of below 0.5, the participant had decided in favor of the carp lake (and accordingly above 0.5 in favor of the trout lake). The top row shows which of the two lakes has been correct in each round and we can see that the participant gave no incorrect answer in these nine rounds. The row below shows if a high- or a low-pitched tone has been played. Thus, we can see that the sixth and seventh round had unrepresentative tone meanings.

doi:10.1371/journal.pcbi.1005328.g003

learning run. Moreover, it has to be ensured, that differences in learning the tone meaning associated with varying psychosis-proneness constitute no alternative explanation for the relationships between psychosis-proneness and altered information processing assessed in this study.

For the former purpose, we performed a "proof of concept" model comparison between the full model with the tone parameter estimated as a free parameter and a model in which the tone parameter was fixed to 0 (i.e., that assumes that the tone has not been used by the participants). A superiority of the full model over the model with fixed tone parameter would prove that the tone was indeed used by the participants. Formal model comparison was carried out as described in the "model comparison" paragraph above.

For the latter purpose (excluding psychosis-related differences in learning the tone), we correlated PDI and CAPS scores with the value of the tone parameter $\theta$. A lack of such a relationship would demonstrate that there is no evidence that learning and usage of the tone depended on the participants' psychosis proneness.

### Distribution of the non-linearity parameter

Contrary to most computational learning models, our model included a non-linear relationship between prediction error and learning that captures a reduced impact of high prediction errors resulting in adaptive reduced learning from irrelevant information, and, thus, in a resilience against overhasty and erroneous inferences. The degree of the adaptiveness of learning is quantified by the non-linearity parameter $\zeta$, with higher values of $\zeta$ indicating a stronger adaptiveness of learning. Because there are to our knowledge no prior studies that implemented this kind of non-linear prediction error processing, we computed the subject-specific values of $\zeta$ without prior assumptions on its distribution, i.e., with a uniform prior distribution that made every value between 0 and 5 equally likely. To generally assess the form in which the degree of resilience against irrelevant information was distributed in our sample of participants, we tested the hypotheses that the estimated values of $\zeta$ were uniformly distributed (like the naive prior distribution) or normally distributed (like many psychological and biological variables). To this end, one-sample Kolmogorov-Smirnov tests were applied.

### Relationship between model parameters and jumping-to-conclusions

To test if a low resilience against irrelevant information was related to overhasty inferences (i.e., to jumping-to-conclusions), we correlated the values of $\zeta$ with the participants' mean number of draws to decision in the probabilistic reasoning task. This analysis primarily served as a proof-of-concept since the parameter values of an appropriate behavioral model for the probabilistic reasoning task should be able to explain a large part of the interindividual variance in jumping-to-conclusions.

To additionally demonstrate that JTC can be more accurately explained by a reduced resilience against irrelevant information compared to a generally increased learning speed, we subsequently calculated correlations between the number of draws to decision and the participants' learning rates (values of $\alpha$ in the linear model).

### Relationships between model parameters, delusional convictions and anomalous perceptions

To test our hypothesis that psychosis-related experiences would inversely relate to the adaptiveness of learning, we calculated correlations between the participants' values of $\zeta$ and the proneness for delusional convictions (PDI scores) as well as hallucinatory experiences (CAPS scores). Since the distribution of PDI and CAPS scores differed significantly from a normal distribution (as reported above), we again used non-parametric Spearman rank correlations for this purpose.

To additionally demonstrate that this relationship is specific for a reduced resilience against irrelevant information and does not only reflect a generally increased learning speed, we repeated the analysis including the participants' learning rates (values of $\alpha$ in the linear model) as a covariate in a Spearman partial correlation between PDI and CAPS scores and $\zeta$ values.

Finally, we repeated these analyses correcting for multiple comparisons with correlated outcomes (according to [26], see above).

## Results

### Sample characteristics

Four participants were excluded from analyses because they showed excessively high error rates at the final decision of the probabilistic reasoning task (more than two standard deviations above sample mean, corresponding to more than 26.2% wrong decisions), suggesting

that they did not perform properly in the task. The characteristics of the remaining sample of 94 participants are summarized in Table 1.

## Relationship between jumping-to-conclusions, delusional convictions and anomalous perceptions

A large body of evidence has linked psychosis and psychosis-proneness to a reduced number of draws to decisions in the beads task (JTC, [11,31]). To test whether we could replicate this relationship in our current sample, we related the mean number of draws to decision to PDI scores (proneness for delusional convictions) and CAPS scores (proneness for anomalous sensory experiences). Indeed, our categorical assessment revealed a significantly increased PDI in the JTC group (n = 23, median of PDI scores = 79) as compared with the no-JTC group (n = 71, median of PDI scores = 45), Mann-Whitney test with U = 585, p = 0.042, two-sided. No significant difference was found for CAPS scores (median CAPS score in JTC group = 30, median score in no-JTC group = 15, Mann-Whitney test with U = 681,5, p = 0.233).

Correlational analyses reproduced the effect for PDI scores on a trend level (fewer draws to decision with rising PDI scores, rho = -0.177, p = 0.089, two-sided Spearman rank correlation), whereas CAPS scores again showed no significant relationship (rho = -0.154, p = 0.139, two-sided Spearman rank correlation).

When adjusted for multiple comparisons with correlated outcomes, the relationship between JTC and PDI scores was still present, but failed to reach statistical significance (p adjusted = 0.051 in the categorical analysis and p adjusted = 0.108 for the correlational analysis).

## Model comparison

To test if our model that allowed for a non-linear relationship between prediction error and learning explained the participants' behavior better than a standard linear model, we performed a Bayesian Model Comparison between the two models. Here, the non-linear model proved to be superior to the standard linear model in explaining participants' behavior with a protected exceedance probability of 100%. The mean value of explained variance ($R^2$) of the participants trajectory of belief was 0.696 for the non-linear and 0.661 for the linear model. Since this measure of accuracy does not take into account the differing model complexity, it is not suitable for directly comparing the quality of the models. It however shows that especially the non-linear model appropriately tracked the course of the participants' belief.

## Learning and usage of the tone parameter

With two proof-of-concept analyses, we aimed at ensuring, that the tone meaning was learned by the participants during the learning run, but that psychosis-proneness was not associated with differences in learning the tone meaning.

**Table 1. Sample characteristics. PDI = Peters Delusions Inventory; CAPS = Cardiff Anomalous Perceptions Scale.**

| **Characteristic** | **Mean (SD)** |
|---|---|
| Age | 30.52 (10.08) |
| PDI score | 64.54 (57.11) |
| CAPS score | 30.39 (35.61) |
| **Characteristic** | **Absolute numbers** |
| Sex | female: 56; male: 38 |
| Smoking | yes: 27; no: 67 |
| Graduation | lower secondary school: 6; higher secondary school: 24; high school: 63; missing information: 1 |

doi:10.1371/journal.pcbi.1005328.t001

To ensure the significance of the tone for the participants' belief updating, we performed a formal model comparison between the full model, in which the tone parameter was estimated as an individual free parameter and a model without tone parameter (i.e., with θ fixed to 0, assuming that the tone was not used by the participants). Here, the model, in which θ was freely estimated, proved to be superior with a protected exceedance probability of 99.5%. Thus, taking into account the usage of the tone significantly improved the tracking of the participants' belief trajectory.

No significant or trend-wise relationships were found between the value of the tone parameter and delusion-proneness (PDI scores, rho = 0.030, p = 0.771, two-sided Spearman correlation) or hallucination-proneness (CAPS scores, rho = 0.038, p = 0.714), not providing any evidence that individual differences in the learning of the tone might provide an alternative explanation for the found relationships between psychosis proneness and the lowered resilience against irrelevant information.

## Distribution of the non-linearity parameter

The non-linearity parameter ζ quantifies the degree to which the impact of the prediction error on learning is attenuated if new information is very surprising. Accordingly, low values indicate maladaptive learning with a low resilience against irrelevant information. In common learning models, a linear relationship, i.e., a parameter value of ζ = 0, is assumed. We estimated the values of ζ that optimally explained our participants' behavior with an uninformative prior distribution uniformly distributed between 0 (linear relationship) and 5 (strongly non-linear relationship). It turned out that none of our participants showed a linear processing of the prediction error. Instead, more than 95% of the participants showed values of ζ between 2.5 and 4 with a marked peak around 3.5 (Fig 4). Interestingly, the distribution of the estimated values of ζ differed significantly from the prior uniform distribution ($Z = 4.663$, $p < 0.001$, Kolmogorov-Smirnov test), but not from a normal distribution ($Z = 1.051$, $p = 0.219$, Kolmogorov-Smirnov test). These results suggest that in our probabilistic reasoning task, learning depended on the prediction error in a non-linear fashion and all participants showed resilience against irrelevant information, although the degree of this resilience varied across participants.

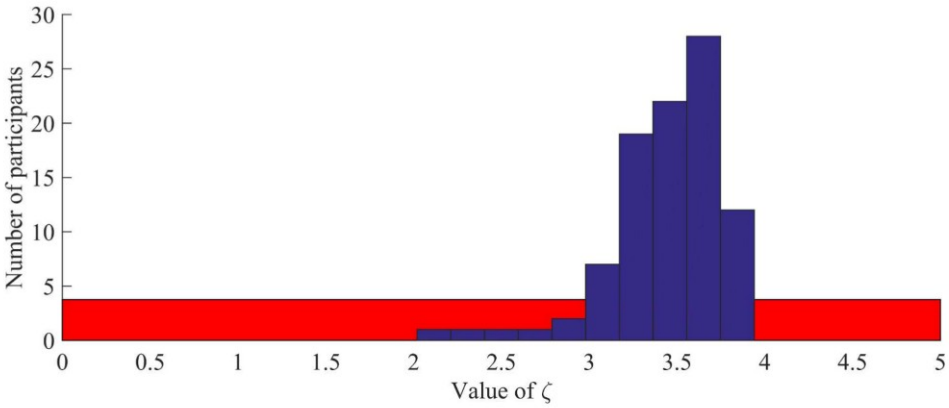


**Fig 4. Histogram of the distribution of ζ values in our sample.** Prior distribution (uniform between 0 and 5) in red, real distribution in blue. It can be seen that, contrary to the uniform prior distribution, estimated parameter values show a normal-like distribution with a mean value of 3.44. This is confirmed by Kolmogorov-Smirnov tests on the form of the underlying distribution.

doi:10.1371/journal.pcbi.1005328.g004

## Relationships between model parameters and jumping-to-conclusions

To test if the participants' resilience against atypical information was associated with jumping-to-conclusions behavior, we correlated parameter values of $\zeta$ with the participants mean number of draws to decision in the probabilistic reasoning task. This analysis yielded a strong positive correlation ($r = 0.705$, $p < 0.001$, Pearson Correlation), indicating that participants with a lower resilience against irrelevant information took decisions based on less evidence (JTC).

As predicted, the learning rate $\alpha$ of the linear model also showed a (negative) correlation with draws to decision, albeit weaker than $\zeta$ in the non-linear model ($r = -0.460$, $p < 0.001$, Pearson Correlation).

## Relationships between model parameters, delusional ideation and hallucinatory experiences

According to predictive coding models of psychosis, maladaptive learning with a reduced resilience against irrelevant information would result in overhasty and erroneous inferences, and should therefore be related to an increased proneness towards delusional ideation and hallucinatory experiences.

Notably and in line with this hypothesis, estimated parameter values of $\zeta$ were negatively correlated with PDI scores (rho = -0.235, p = 0.022, two-sided Spearman rank correlation, Fig 5A) and trend-wise with CAPS scores (rho = -0.198, p = 0.056, two-sided Spearman rank correlation, Fig 5B), indicating that individuals with a low resilience against irrelevant information showed an increased proneness for delusional ideation and (as a tendency) hallucinatory experiences. To exclude significant correlations due to four outliers with $\zeta$ values two standard deviations below mean (i.e., below 2.794, outliers are marked with white squares in Fig 5), we repeated these analyses excluding the outliers (thus with n = 90). This resulted in rather

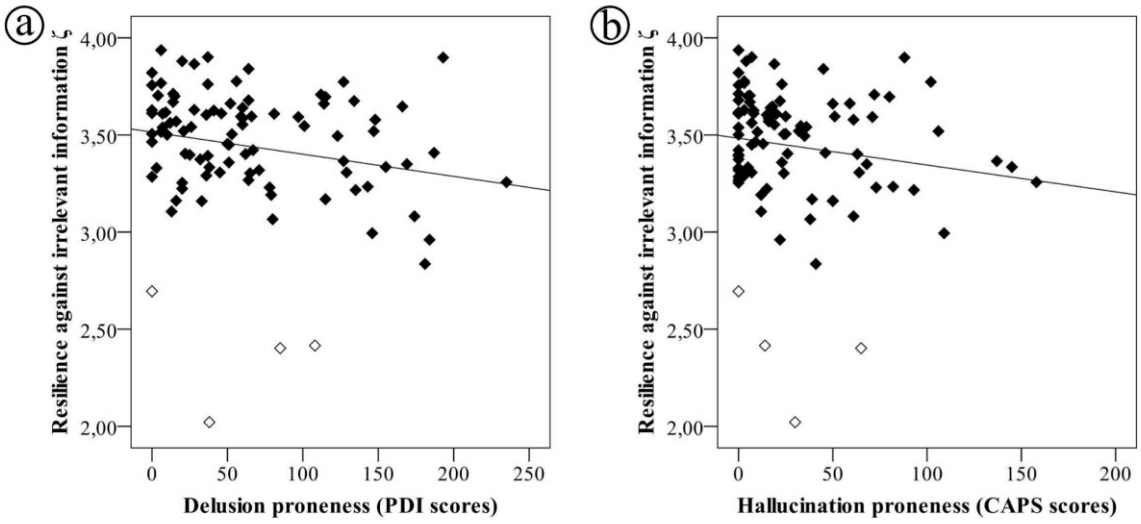


**Fig 5. Association between resilience against irrelevant information in the probabilistic reasoning task and psychotic experiences.** (a) association with delusion proneness (PDI scores, rho = -0.235, p = 0.022, Spearman's correlation), (b) association with hallucination proneness (CAPS scores, rho = -0.198, p = 0.056). Empty squares mark outliers with $\zeta$ values two standard deviations below mean. Exclusion of these outliers yielded magnified effect sizes.

doi:10.1371/journal.pcbi.1005328.g005

stronger effects (rho = -0.261, p = 0.014 for the correlation between PDI scores and $\zeta$, rho = -0.212, p = 0.044 for the correlation between CAPS scores and $\zeta$).

These correlations remained unchanged when controlling for the unspecific learning rate obtained from the linear model: A Spearman partial correlation including the linear learning rate as a covariate revealed similar results to those reported above (r = -0.234, p = 0.024 for correlation between PDI scores and $\zeta$, r = -0.196, p = 0.062 for correlation between CAPS scores and $\zeta$). This shows that effects between psychosis proneness and belief updating specifically affect the treatment of irrelevant information and cannot be explained by the nonspecific linear learning rate.

When p-values were adjusted for multiple comparisons (two tests with correlated outcomes for CAPS and PDI scores), the relationship between PDI scores and $\zeta$ values remained significant (adjusted p of two-sided Spearman rank correlation = 0.027) and the relationship between CAPS scores and $\zeta$ values a statistical trend (adjusted p of two-sided Spearman rank correlation = 0.068).

## Discussion

In the present study, we tested the claim put forward by predictive coding models [1–3] that psychotic experiences may be linked to maladaptive learning, i.e., an aberrant encoding of precision, that results from a reduced resilience against irrelevant information and leads to overhasty and erroneous inferences. In line with our hypothesis, we found that delusional ideation and hallucinatory experiences of healthy individuals were predicted by a low resilience against irrelevant information in a probabilistic reasoning task.

In order to quantify the resilience against irrelevant information, we applied a novel computational learning model that allowed for a non-linear relationship between prediction error and learning. Compared to a linear relationship, individuals with a non-linear processing of the prediction error are relatively resilient against an overestimation of excessively surprising information, since the relationship between prediction error and learning is flattened for high prediction errors. On the other hand, they are still capable of rapidly building predictive beliefs about the world, since learning of moderately surprising information (low prediction errors) is not relevantly impaired. By this approach, we were thus able to disentangle inter-individual differences in the *general* speed of learning (that is captured in the learning rate) from specific differences in the impact of former beliefs on learning (that is captured in the resilience against irrelevant information). Our results suggest that specifically the latter is related to an increased proneness for delusional and hallucinatory experiences. Considering that every incoming signal is noisy and naturally contains both relevant and irrelevant information, it seems plausible that an attenuation of specifically the excessively surprising and hence irrelevant information constitutes an effective protection against overhasty and erroneous inferences, while a weakening of this attenuation in turn predisposes for delusional beliefs and hallucinatory percepts. This understanding is additionally supported by our finding that the resilience against irrelevant information indeed was the parameter with the strongest association with hasty inferences (jumping-to-conclusions) and that jumping-to-conclusions was, consistent with prior studies [11,31], in turn related to the proneness for delusional convictions. Obviously, the adaptiveness of a non-linear prediction error processing depends on the particular task: In volatile tasks with frequent changes of the underlying probabilities, large prediction errors might in fact provide vital information, namely that the context of learning has changed. Our task however included no volatility in that sense, because in one round, there were no changes in the task probabilities (the lake, from which fishes were being angled remained the same). Thus, the degree of non-linearity indeed provides a measure for adaptive learning in the adapted beads task.

The idea that a core alteration in psychosis lies in the weighting of new information with regard to prior beliefs has a longstanding history in cognitive schizophrenia research evolving from the suggestion that "the basic disturbance in schizophrenia is 'a weakening of the influences of stored memories of regularities of previous input on current perception'" [32] via hypotheses of an aberrant attribution of salience to stimuli [8,33] to predictive coding frameworks that embed these hypotheses into a broader framework of Bayesian information processing in the brain [2,3]. In line with this, schizophrenia and/or psychotic symptoms have been consistently linked to the aberrant attribution of salience to stimuli ([34,35] for reviews). Similarly, schizophrenia and psychotic symptoms have been associated with a decreased influence of prior beliefs in perceptual inference ([36], see [37] for a review on visual illusions), although recent work suggests a complex interplay between prior beliefs and perceptual inference in psychosis-related conditions [38,39]. By the use of a feasible and interpretable model, our current study provides a formal description for the interaction between prior beliefs and new information, thereby elucidating the computational mechanisms underlying maladaptive learning and inference in psychosis.

Intriguingly, we found that our participants showed without exception a resilience against irrelevant information that cannot be captured in models that assume a linearly processed prediction error. Especially considering the significant clustering of parameter values in a region with a marked non-linearity in our experiment, this finding suggests that learning in some tasks might not be driven linearly by prediction errors. From a more general perspective, a growing non-linearity between prediction error and learning implies that the impact that a certain new information has on the belief (i.e., the learning) becomes increasingly *independent* from the current belief itself: Whilst under the assumption of a linearly processed prediction error, one and the same information (e.g., a certain fish in our task) has a massively different impact on the learning depending on whether it is surprising or not, a strong non-linearity implies that every fish is treated more or less equally, regardless of the current belief. Similar concepts have previously been proposed in terms of precision-weighted prediction errors, where the learning of strongly surprising information is attenuated if a marked and precise opposing belief has already been built, e.g., if the precision of the belief is high and the precision of the new information low [19]. Compared to these frameworks, our approach has the advantage of simplicity and that the degree of resilience against irrelevant information is captured in one single and easily interpretable parameter: It is noteworthy, that a reduced non-linearity of the relationship between prediction error and learning that could be proven to be associated with psychosis-proneness in this study effectively and straightforwardly models what has been theoretically proposed as a core alteration behind psychotic experiences, namely "a reduction in the precision of prior beliefs, relative to sensory evidence" [1]. Nevertheless, whilst providing a substantial model fit in rather simple tasks like the one carried out in this study, it is questionable if our non-hierarchical model can sufficiently account for more complex environments (e.g., environments with changing volatility). Based on the continuity view of psychosis, we studied psychotic experiences in a sample of non-clinical participants. Mounting evidence suggests that clinical and non-clinical psychotic experiences reflect different expressions of a continuously distributed trait, as they share a common factor structure [40], similar risk factors and demographics [21] as well as a co-clustering in relatives [41,42]. It could moreover be prospectively demonstrated that an increased, but non-clinical proneness for psychotic experiences massively increases the risk of developing a "full" clinical psychosis in the future [43–45], further indicating that non-clinical and clinical psychotic experiences can be explained in terms of similar underlying mechanisms. Importantly, studying psychotic experiences in non-clinical participants does preclude potential confounds associated with clinical diseases and their pharmacological treatment. Hence, although future work is needed

to confirm whether our current findings generalize to patients suffering from psychotic disease, the link between maladaptive learning and psychotic experiences established here might generally shed light on the computational mechanisms underlying both non-clinical psychotic experiences and psychosis.

One of the studies limitations is that we only yielded a modest and non-significant association between conventional JTC measures (draws to decision) and psychosis proneness. This is however consistent with previous reports on the relationship between JTC and psychosis proneness in healthy individuals that yielded small effects and mixed results [46–49] and suggest that conventional JTC measures such as draws to decision might not provide a sufficiently fine-grained measure for individual psychosis-related differences in learning and reasoning in healthy individuals.

Taken together, our current findings suggest that a less non-linear processing of prediction error gives rise to overhasty and erroneous inferences, thereby leading to delusional ideas and hallucinatory experiences. Our current work thus empirically substantiates theories that link maladaptive learning to psychotic experiences both in health and disease.

## Supporting Information

**S1 Text. Mathematical proof that the belief in the non-linear model is bounded between zero and one.**
(PDF)

## Author Contributions

**Conceptualization:** HeS KS VAW.

**Formal analysis:** HeS HaS KS.

**Funding acquisition:** KS.

**Investigation:** HeS.

**Methodology:** HeS HaS KS.

**Project administration:** HeS VAW KS.

**Resources:** KS.

**Software:** HeS HaS KS.

**Supervision:** KS.

**Validation:** HeS KS.

**Visualization:** HeS KS.

**Writing – original draft:** HeS.

**Writing – review & editing:** HeS HaS VAW KS.

## References

1. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013) The computational anatomy of psychosis. Front Psychiatry 4: 47. doi: 10.3389/fpsyt.2013.00047 PMID: 23750138
2. Corlett PR, Frith CD, Fletcher PC (2009) From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology (Berl) 206: 515–530.

3. Fletcher PC, Frith CD (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 10: 48–58. doi: 10.1038/nrn2536 PMID: 19050712
4. Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360: 815–836. doi: 10.1098/rstb.2005.1622 PMID: 15937014
5. Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci 27: 712–719. doi: 10.1016/j.tins.2004.10.007 PMID: 15541511
6. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66: 241–251. PMID: 1540675
7. Heinz A (2002) Dopaminergic dysfunction in alcoholism and schizophrenia—psychopathological and behavioral correlates. Eur Psychiatry 17: 9–16. PMID: 11918987
8. Kapur S (2003) Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry 160: 13–23. doi: 10.1176/appi.ajp.160.1.13 PMID: 12505794
9. Galdos M, Simons C, Fernandez-Rivas A, Wichers M, Peralta C, et al. (2011) Affectively salient meaning in random noise: a task sensitive to psychosis liability. Schizophr Bull 37: 1179–1186. doi: 10.1093/schbul/sbq029 PMID: 20360211
10. Vercammen A, de Haan EH, Aleman A (2008) Hearing a voice in the noise: auditory hallucinations and speech perception. Psychol Med 38: 1177–1184. doi: 10.1017/S0033291707002437 PMID: 18076771
11. McLean BF, Mattiske JK, Balzan RP (2016) Association of the Jumping to Conclusions and Evidence Integration Biases With Delusions in Psychosis: A Detailed Meta-analysis. Schizophr Bull.
12. Ross K, Freeman D, Dunn G, Garety P (2011) A randomized experimental investigation of reasoning training for people with delusions. Schizophr Bull 37: 324–333. doi: 10.1093/schbul/sbn165 PMID: 19520745
13. So SH-w, Siu NY-f, Wong H-l, Chan W, Garety PA (2016) 'Jumping to conclusions' data-gathering bias in psychosis and other psychiatric disorders—Two meta-analyses of comparisons between patients and healthy individuals. Clinical Psychology Review 46: 151–167. doi: 10.1016/j.cpr.2016.05.001 PMID: 27216559
14. Fine C, Gardner M, Craigie J, Gold I (2007) Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. Cogn Neuropsychiatry 12: 46–77. doi: 10.1080/13546800600750597 PMID: 17162446
15. Moore SC, Sellen JL (2006) Jumping to conclusions: a network model predicts schizophrenic patients' performance on a probabilistic reasoning task. Cogn Affect Behav Neurosci 6: 261–269. PMID: 17458441
16. Moritz S, Scheu F, Andreou C, Pfueller U, Weisbrod M, et al. (2016) Reasoning in psychosis: risky but not necessarily hasty. Cogn Neuropsychiatry 21: 91–106. doi: 10.1080/13546805.2015.1136611 PMID: 26884221
17. Moritz S, Woodward TS, Lambert M (2007) Under what circumstances do patients with schizophrenia jump to conclusions? A liberal acceptance account. Br J Clin Psychol 46: 127–137. doi: 10.1348/014466506X129862 PMID: 17524208
18. Rescorla RA, Wagner AW (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors. Classical Conditioning II: Current Research and Theory: Appleton-Century-Crofts. pp. 64–99.
19. Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011) A bayesian foundation for individual learning under uncertainty. Front Hum Neurosci 5: 39. doi: 10.3389/fnhum.2011.00039 PMID: 21629826
20. Nelson MT, Seal ML, Pantelis C, Phillips LJ (2013) Evidence of a dimensional relationship between schizotypy and schizophrenia: a systematic review. Neurosci Biobehav Rev 37: 317–327. doi: 10.1016/j.neubiorev.2013.01.004 PMID: 23313650
21. van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L (2009) A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. Psychol Med 39: 179–195. doi: 10.1017/S0033291708003814 PMID: 18606047
22. Peters ER, Joseph SA, Garety PA (1999) Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). Schizophr Bull 25: 553–576. PMID: 10478789
23. Bell V, Halligan PW, Ellis HD (2006) The Cardiff Anomalous Perceptions Scale (CAPS): a new validated measure of anomalous perceptual experience. Schizophr Bull 32: 366–377. doi: 10.1093/schbul/sbj014 PMID: 16237200
24. Phillips LD, Edwards W (1966) Conservatism in a simple probability inference task. J Exp Psychol 72: 346–354. PMID: 5968681

25. Garety PA, Freeman D, Jolley S, Dunn G, Bebbington PE, et al. (2005) Reasoning, emotions, and delusional conviction in psychosis. J Abnorm Psychol 114: 373–384. doi: 10.1037/0021-843X.114.3.373 PMID: 16117574

26. Dubey SD. Adjustment of p-values for multiplicities of intercorrelating symptoms; 1985; Germany.

27. Sankoh AJ, Huque MF, Dubey SD (1997) Some comments on frequently used multiple endpoint adjustment methods in clinical trials. Stat Med 16: 2529–2542. PMID: 9403954

28. Sutton RS, Barto AG (1998) Reinforcement learning: An introduction: MIT press.

29. Daunizeau J, Adam V, Rigoux L (2014) VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comput Biol 10: e1003441. doi: 10.1371/journal.pcbi.1003441 PMID: 24465198

30. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46: 1004–1017. doi: 10.1016/j.neuroimage.2009.03.025 PMID: 19306932

31. Ross RM, McKay R, Coltheart M, Langdon R (2015) Jumping to Conclusions About the Beads Task? A Meta-analysis of Delusional Ideation and Data-Gathering. Schizophr Bull 41: 1183–1191. doi: 10.1093/schbul/sbu187 PMID: 25616503

32. Hemsley DR (1993) A simple (or simplistic?) cognitive model for schizophrenia. Behav Res Ther 31: 633–645. PMID: 8216165

33. Heinz A (1999) [Anhedonia—a general nosology surmounting correlate of a dysfunctional dopaminergic reward system?]. Nervenarzt 70: 391–398. PMID: 10407834

34. Heinz A, Schlagenhauf F (2010) Dopaminergic dysfunction in schizophrenia: salience attribution revisited. Schizophr Bull 36: 472–485. doi: 10.1093/schbul/sbq031 PMID: 20453041

35. Nelson B, Whitford TJ, Lavoie S, Sass LA (2014) What are the neurocognitive correlates of basic self-disturbance in schizophrenia? Integrating phenomenology and neurocognition: Part 2 (aberrant salience). Schizophr Res 152: 20–27. doi: 10.1016/j.schres.2013.06.033 PMID: 23863772

36. Schmack K, Schnack A, Priller J, Sterzer P (2015) Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. Schizophrenia Research: Cognition 2: 72–77.

37. Notredame CE, Pins D, Deneve S, Jardri R (2014) What visual illusions teach us about schizophrenia. Front Integr Neurosci 8: 63. doi: 10.3389/fnint.2014.00063 PMID: 25161614

38. Schmack K, Gomez-Carrillo de Castro A, Rothkirch M, Sekutowicz M, Rossler H, et al. (2013) Delusions and the role of beliefs in perceptual inference. J Neurosci 33: 13701–13712. doi: 10.1523/JNEUROSCI.1778-13.2013 PMID: 23966692

39. Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, et al. (2015) Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci U S A 112: 13401–13406. doi: 10.1073/pnas.1503916112 PMID: 26460044

40. Shevlin M, McElroy E, Bentall RP, Reininghaus U, Murphy J (2016) The Psychosis Continuum: Testing a Bifactor Model of Psychosis in a General Population Sample. Schizophrenia Bulletin.

41. Fanous A, Gardner C, Walsh D, Kendler KS (2001) Relationship between positive and negative symptoms of schizophrenia and schizotypal symptoms in nonpsychotic relatives. Arch Gen Psychiatry 58: 669–673. PMID: 11448374

42. Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, et al. (1993) The Roscommon Family Study. III. Schizophrenia-related personality disorders in relatives. Arch Gen Psychiatry 50: 781–788. PMID: 8215802

43. Chapman LJ, Chapman JP, Kwapil TR, Eckblad M, Zinser MC (1994) Putatively psychosis-prone subjects 10 years later. J Abnorm Psychol 103: 171–183. PMID: 8040487

44. Hanssen M, Bak M, Bijl R, Vollebergh W, van Os J (2005) The incidence and outcome of subclinical psychotic experiences in the general population. Br J Clin Psychol 44: 181–191. doi: 10.1348/014466505X29611 PMID: 16004653

45. Poulton R, Caspi A, Moffitt TE, Cannon M, Murray R, et al. (2000) Children's self-reported psychotic symptoms and adult schizophreniform disorder: a 15-year longitudinal study. Arch Gen Psychiatry 57: 1053–1058. PMID: 11074871

46. Freeman D, Pugh K, Garety P (2008) Jumping to conclusions and paranoid ideation in the general population. Schizophr Res 102: 254–260. doi: 10.1016/j.schres.2008.03.020 PMID: 18442898

47. So SH, Kwok NT (2015) Jumping to conclusions style along the continuum of delusions: delusion-prone individuals are not hastier in decision making than healthy individuals. PLoS One 10: e0121347. doi: 10.1371/journal.pone.0121347 PMID: 25793772

48. Van Dael F, Versmissen D, Janssen I, Myin-Germeys I, van Os J, et al. (2006) Data gathering: biased in psychosis? Schizophr Bull 32: 341–351. doi: 10.1093/schbul/sbj021 PMID: 16254066

49. Warman DM, Lysaker PH, Martin JM, Davis L, Haudenschield SL (2007) Jumping to conclusions and the continuum of delusional beliefs. Behav Res Ther 45: 1255–1269. doi: 10.1016/j.brat.2006.09.002 PMID: 17052687

# 3. Diskussion

## *3.1 Die präsentierten Ergebnisse im Kontext der dual system Theorie der Abhängigkeit*

Die unter 2.1 und 2.2 präsentierten Ergebnisse (Kunas, Stuke, Heinz, Strohle, & Bermpohl, 2022; Stuke et al., 2016) stützen die incentive salience Theorie der Abhängigkeit (s. Abschnitt 1.3) vollumfänglich, indem sowohl eine verstärkte Reaktivität auf substanzbezogene Belohnungen, als auch eine reduzierte Reaktivität auf nicht-substanzbezogene alternative Belohnungen nachgewiesen wurde. Während eine verstärkte Reaktivität auf substanzbezogene Belohnungen bei Abhängigkeit ein erwartbarer und vielfach replizierter Befund ist (Lin et al., 2020), rückt die Bedeutung, die eine reduzierte Reaktivität auf nicht-substanzbezogene Belohnungen in der Entstehung der Abhängigkeit spielt, erst langsam in den Fokus des Forschungsinteresses (Versace et al., 2017). Ein reduziertes Ansprechen belohnungsassoziierter Areale auf nicht-substanzbezogene Reize passt klinisch zu der bei Abhängigkeitserkrankungen häufig vorliegenden Anhedonie (Garfield, Lubman, & Yucel, 2014; Hatzigiakoumis, Martinotti, Giannantonio, & Janiri, 2011) sowie zu den Ergebnissen verschiedener psychologisch-behavioraler Tests, die eine geringere Erregbarkeit durch nicht-substanzbezogene Verstärker bei Patient*innen mit Abhängigkeit zeigen (Acuff, Dennhardt, Correia, & Murphy, 2019). Anhedonie psychotherapeutisch oder pharmakologisch zu adressieren, kann daher ein relevanter Bestandteil einer multimodalen Therapie bei Abhängigkeit sein (Hatzigiakoumis et al., 2011) und ihre individuelle Erfassung könnte zur Personalisierung von Therapieplänen beitragen (s. Abschnitt 3.3). Eine weitere vielversprechende Translation der incentive salience Theorie in die klinische Behandlung von Abhängigkeit stellt das sogenannte bias modification training dar (Schoenmakers et al., 2010). Dieses geht aus von der Beobachtung, dass bei Patient*innen mit Alkoholabhängigkeit ein „bias“ besteht, der eine automatische präferentielle Prozessierung, Allokation von Aufmerksamkeit und eine Annäherungstendenz an alkoholbezogene Stimuli umfasst. Quantifiziert werden kann diese automatische Annäherungstendenz beispielsweise mit einer sogenannten impliziten Annäherungs-Vermeidungs-Aufgabe. Bei dieser werden den Patient*innen Bilder von alkoholischen und nichtalkoholischen Getränken präsentiert, die sie, abhängig von einem von der Art des Getränks unabhängigen Kriterium (z.B. Quer- oder Hochformat des Bildes) mit einem Joystick zu sich „hinziehen“ oder von sich „wegschieben“

sollen (z.B. müssen Bilder im Querformat stets „weggeschoben“ werden). Mit diesen Aufgaben konnte gezeigt werden, dass Patient*innen mit Alkoholabhängigkeit schnellere Reaktionszeiten beim „Hinziehen“ von alkoholischen Getränken zeigen (verglichen mit nichtalkoholischen Getränken sowie gesunden Kontrollproband*innen) und dass mit diesem approach bias eine gesteigerte Aktivität in Arealen des mesolimbischen Dopaminsystems einhergeht (Wiers et al., 2014). Als approach bias modification Training wird eine Variante dieser Aufgabe bezeichnet, bei der überproportional häufig (z.B., zu 80%) Bilder von alkoholischen Getränken „weggeschoben“ werden müssen (d.h., im obigen Beispiel im Querformat präsentiert werden). Durch wiederholtes Training auf diese Art wird sich eine Verbesserung der automatischen Annäherungstedenz verprochen. Trotz des zunächst spekulativen Wirkprinzips ist die klinische Effektivität dieses Ansatzes, vor allem in der Rückfallprävention bei Alkoholabhängigkeit, inzwischen in mehreren größeren Studien belegt worden (Kakoschke, Kemps, & Tiggemann, 2017; Lammel et al., 2014) und eine fMRT-Studie legt Reduktionen in der Aktivierung der Amygdala durch alkoholbezogene Stimuli als möglichen Wirkmechanismus nahe (Wiers et al., 2015).

Die Befunde unserer Arbeiten für das in dual system Theorien postulierte Defizit in Kontrollprozessen und den mit ihnen verbundenen Hirnarealen sind weniger eindeutig. In der unter 2.1 präsentierten Studie zeigte sich paradoxerweise eine verstärkte Aktivierung von mit kognitiver Kontrolle in Verbindung gebrachten Arealen (DLPFC) mit zunehmender Trinkschwere während Entscheidungen für ein alkoholisches Getränk (Stuke et al., 2016). Eine weitere Studie unserer Arbeitsgruppe fand später ebenfalls keinen Unterschied zwischen Raucher*innen und Nichtraucher*innen in der Aktivierung kontrollassoziierter Areale während kognitiver Kontrollprozesse (Herunterregulierung von Konsumverlangen für Nikotin, (Kunas, Stuke, Plank, et al., 2022)). Demgegenüber stehen zahlreiche Studienergebnisse, die Defizite von kognitiver Kontrolle bei Patient*innen mit Abhängigkeit dokumentieren sowohl in verschiedenen Verhaltensmaßen (Selbstauskunftsskalen zur Impulsivität, Reaktionszeiten in stop-signal Aufgaben, Entscheidungen in reversal learning Aufgaben, (Jentsch & Pennington, 2014)), als auch in fMRT- und EEG-Studien (Luijten et al., 2014). Auch der Befund der unter 2.3 dargestellten Studie, dass eine Präsentation von Bildern negativer Konsequenzen des Rauchens nachfolgend die Aktivität kontrollassoziierter Areale erhöht (und mit einem subjektiv reduzierten Konsumverlangen einhergeht) spricht für eine Bedeutung von kognitiver Kontrolle

im Sinne der dual system Theorie (d.h., unter anderem bei der Suppression von Konsumverlangen basierend auf langfristigen Zielstellungen).

Derartig inkonsistente Befunde können auftreten, wenn funktionell hoch komplexe Areale wie der DLPFC untersucht werden, deren Aktivität weit darüber hinausgeht, ein Korrelat von Selbstkontrollprozessen zu sein und deren Veränderung bei Abhängigkeitserkrankungen zahlreichen Faktoren unterliegt wie der Schwere der Abhängigkeit, dem aktuellen Konsumstatus (intoxikiert versus entzügig) und der Art der konsumierten Substanz (Goldstein & Volkow, 2011). Zu beachten ist außerdem, dass in unseren Studien neu entwickelte tasks eingesetzt wurden (eine alkoholspezifische Entscheidungsaufgabe in (Stuke et al., 2016) und eine Aufgabe zur bewussten Suppression von Konsumverlangen in (Kunas, Stuke, Plank, et al., 2022)), die sich von den in (Luijten et al., 2014) zusammengefassten, unspezifischeren Kontroll- und Inhibitionsaufgaben (Stopp-Signal-Aufgabe, Go-NoGo-Aufgabe) unterscheiden. Diese Aufgaben bilden also unterschiedliche Aspekte von kognitiver Kontrolle ab: Stopp-Signal-Aufgaben und Go-NoGo-Aufgaben werden hierbei als „cold" cognitive control tasks bezeichnet, bei denen es um die Realisierung von Intentionen und Instruktionen geht, wenn diese im Konflikt mit automatisierten oder habituellen Reaktionen stehen. Demgegenüber wird von „hot" control tasks gesprochen, wenn es bei der Verfolgung langfristiger Ziele notwendig ist, Emotionen, kurzfristige Anreize, konkurrierende Motivationstendenzen oder Craving zu regulieren (Chan, Shum, Toulopoulou, & Chen, 2008). Die von uns eingesetzten tasks fallen dementsprechend in die Kategorie der "hot" control tasks und es ist denkbar (wenn auch beim aktuellen Forschungsstand spekulativ), dass diese auf verglichen mit den bislang hauptsächlich untersuchten „cold" tasks auf unterschiedliche Art und Weise mit Aktivierungen im DLPFC einhergehen, was zur Diskrepanz zwischen unseren Ergebnissen und den Vorbefunden beitragen könnte. Die Ergebnisse der unter 2.1 präsentierten Studie basieren zudem auf korrelativen Analysen, d.h., sie setzen Aktivierungen bei der Entscheidung für alkoholische Getränke ins Verhältnis zur Trinkschwere bei Proband*innen mit unterschiedlich ausgeprägtem Alkoholkonsum. Eine denkbare Erklärung für die (entgegen der Hypothese) mit zunehmender Trinkschwere erhöhte Aktivität des DLPFC während Entscheidungen für alkoholische Getränke ist daher, dass die durch starke DLPFC-Aktivität angezeigten Bemühungen um Selbstkontrolle erst bei einer höheren Trinkschwere notwendig werden (im Sinne eines zunehmenden Konflikts zwischen Trinkverlangen und Problembewusstsein bei missbräuchlichem, aber noch nicht

abhängigen Alkoholkonsum). Es bleibt eine Aufgabe weiterer Forschung, Veränderungen in Kontrollprozessen und assoziierten Aktivierungsmustern im fMRT bei Abhängigkeiten detaillierter und theoriegestützter zu differenzieren („cold“ versus „hot“ cognitive control, generelles versus substanzspezifisches Kontrolldefizit).

Ein entscheidender Aspekt bei der Bewertung von aus Bildgebungsstudien gewonnenen Markern ist ihr Potenzial, prospektiv den Erkrankungsverlauf (Rückfallrisiko, Erfolgschancen bestimmter Therapien) vorherzusagen. Einzelne diesbezügliche Untersuchungen weisen bislang darauf hin, dass Bildgebungskorrelate einer verstärkten incentive salience und einer reduzierten Inhibitionsfähigkeit klinische Vorhersagen ermöglichen können, zum Beispiel zum individuellen Rückfallrisiko bei Alkoholabhängigkeit (Moeller & Paulus, 2018). Einer klinischen Nutzung dieser Ergebnisse stehen aber noch mehrere Hürden entgegen. Zum einen mangelt es an großen, präregistrierten prospektiven Studien, die den tatsächlichen Nutzen einer basierend auf neurophysiologischen Markern individualisierten Therapie zeigen. Zum anderen sind die technischen Voraussetzungen für die Akquise der beschriebenen Marker oft zu hoch, um diese in routinemäßige klinische Praxis zu überführen (z.B. Verfügbarkeit von MRT-Messzeiten, komplexe Auswerteprozeduren). In der psychiatrischen Forschung und den klinischen Neurowissenschaften gibt es zunehmende Diskussionen über die Nutzbarmachung von grundlagenwissenschaftlichen Erkenntnissen für die klinische Psychiatrie. Einige diesbezügliche Ansätze werden in Abschnitt 3.3 vorgestellt.

## *3.2 Die präsentierten Ergebnisse im Kontext der Bayesianischen Modelle der Psychose*

Die „aberrant salience“ Theorie der Psychose postuliert eine durch eine erhöhte Dopaminausschüttung bedingte Hypersalienz eigentlich nicht relevanter Reize als Kernpathologie bei der Entstehung psychotischer Symptome. Die der statistischen Informatik entlehnten Bayesianischen Modelle der Psychose gehen darüber hinaus von einer damit zusammenhängenden pathologischen Veränderung der Integration von Vorannahmen und aktuellen sensorischen Informationen aus (Abschnitt 1.4).

Während die unter 2.5 präsentierte Studie in diesem Zusammenhang einen reduzierten Einfluss kurzfristig erlernter probabilistischer Vorannahmen bei Proband*innen mit höherer

Psychoseneigung zumindest in der Domäne der perzeptuellen Entscheidungsfindung nahelegt, zeigt die unter 2.6 präsentierte Studie eine *verstärkte* Vorannahme der Präsenz von Gesichtern mit höherer Psychoseneigung. Diese Inkonsistenz passt zu der generellen Uneindeutigkeit der Bayesianischen Psychosemodelle, die man schematisch als weak prior und strong prior Hypothesen fassen kann. Wie unter 1.4 dargestellt, lassen sich beide Hypothesen grundsätzlich basierend auf Bayesianischen Modellen theoretisch entwickeln. Empirische Bestätigung für die weak prior Hypothese fand sich sowohl in weiteren Arbeiten unserer Arbeitsgruppe (Schmack, Schnack, Priller, & Sterzer, 2015; Weilnhammer et al., 2020) als auch in denjenigen anderer Arbeitsgruppen (Jardri, Duverne, Litvinova, & Deneve, 2017; Valton et al., 2019). Umgekehrt wurden auch mehrfach zur strong prior Hypothese passende Befunde berichtet (Alderson-Day et al., 2017; Powers et al., 2017; Schmack, Bosc, Ott, Sturgill, & Kepecs, 2021; Teufel et al., 2015). Manche Studien berichteten sogar entgegengesetzte Effekte je nach Art der Vorannahmen und / oder Krankheitsstadium (Haarsma, Knolle, et al., 2020; Schmack et al., 2013).

Es gab verschiedene Versuche, diese gegensätzlichen Befunde in Einklang zu bringen, indem verschiedene Arten von Vorannahmen unterschieden wurden und je nach Art der Vorannahme eine unterschiedliche Veränderung bei Psychosen postuliert wurde. Unterschieden wurden hierbei beispielsweise Vorannahmen bei perzeptuellen versus kognitiv-probabilistischen Entscheidungen (so in der unter 2.4 dargestellten Studie), kurzfristige und transiente versus langfristig stabile Vorannahmen (Zeki & Chen, 2020), sowie Vorannahmen auf unterschiedlichen Stufen der Informationsverarbeitung (Schmack et al., 2013; Sterzer et al., 2018). Diesen Kategorien folgend wurde vorgeschlagen, dass auf unteren Ebenen der Informationsverarbeitung Vorannahmen wenig oder dysfunktional genutzt werden, was die Unsicherheit der auf höheren Verarbeitungsebenen prozessierten Informationen erhöht. Um diese erhöhte Unsicherheit zu kompensieren, ist der Einfluss von Vorannahmen auf diesen höheren Ebenen der Informationsverarbeitung erhöht. Da diese „higher level priors“ häufig allgemeine Annahmen kodieren (beispielsweise die Präsenz einer Stimme in akustischen sensorischen Informationen), kann dieser kompensatorische Rückfall auf Vorannahmen der höheren Verarbeitungsebenen zur Entstehung von Halluzinationen beitragen (Schmack et al., 2013; Sterzer et al., 2018). Diese Theorie, die eine Kombination aus weak prior und strong prior Hypothese darstellt, empirisch zu überprüfen, ist gegenwärtig jedoch kaum möglich, da

messbare neurophysiologische Korrelate der jeweiligen Vorannahmen noch nicht ausreichend bestimmt sind.

Während die Ergebnisse der unter 2.4 präsentierten Studie zunächst mit einem ähnlichen Modell in einer klinischen Stichprobe repliziert wurden (Adams, Napier, Roiser, Mathys, & Gilleen, 2018), fand eine Untersuchung einer sehr großen Stichprobe aus der Allgemeinbevölkerung keinen Anhalt für einen Zusammenhang zwischen Psychoseneigung und veränderten Lernmechanismen (Croft et al., 2021). Eine weitere mögliche Erklärung für die Heterogenität der empirischen Ergebnisse zu Bayesianischen Psychosemodellen betrifft dementsprechend die untersuchten Stichproben, die Patient*innen mit Schizophrenie, teils mit und ohne Einnahme antipsychotischer Medikation, sowie (wie die unter 2.4 – 2.6 dargestellten Studien) gesunde Proband*innen mit unterschiedlicher subklinischer Psychoseneigung umfassten. Hier könnten Grenzen der sogenannten Psychosen-Kontinuum-Hypothese deutlich werden. Diese postuliert, dass die klinischen Manifestationen der Psychose die extremste Form einer Psychoseneigung darstellen, die kontinuierlich in der Allgemeinbevölkerung verteilt ist (van Os, Linscott, Myin-Germeys, Delespaul, & Krabbendam, 2009). Die unter 2.4 – 2.6 präsentierten Studien bauen auf dieser Hypothese auf, indem sie mögliche Mechanismen psychotischer Symptome bei Gesunden mit unterschiedlicher subklinischer Psychoseneigung untersuchen. Es gibt konvergierende Evidenz aus verschiedenen Studien, die für die Psychosen-Kontinuum-Hypothese sprechen: So sind psychotische Erfahrungen nicht auf Patient*innen mit Schizophrenie beschränkt, sondern können in unterschiedlichem Maße in der Allgemeinbevölkerung gefunden werden (Bell, Halligan, & Ellis, 2006; Peters, Joseph, Day, & Garety, 2004). Hierbei zeigen Angehörige von Patient*innen mit psychotischen Störungen eine erhöhte subklinische Psychoseneigung, was auf gemeinsame genetische Grundlagen hindeutet (Fanous, Gardner, Walsh, & Kendler, 2001; Kendler et al., 1993; Tienari et al., 2003). Die Risikofaktoren für subklinische Psychoseneigung und klinische Psychosen ähneln sich (Linscott & van Os, 2013) und die Symptome weisen eine ähnliche Faktorenstruktur auf (Shevlin, McElroy, Bentall, Reininghaus, & Murphy, 2017). Schließlich konnte gezeigt werden, dass ein hohes Maß an subklinischer Psychoseneigung das Risiko für spätere klinische Psychosen erhöht (Chapman, Chapman, Kwapil, Eckblad, & Zinser, 1994; Hanssen, Bak, Bijl, Vollebergh, & van Os, 2005; Welham et al., 2009). Zusammengenommen legen diese Ergebnisse nahe, dass subklinische und klinische psychotische Erfahrungen auf ähnlichen Mechanismen basieren. Es

ist demzufolge ein vielversprechender Ansatz, die Mechanismen der Psychose in subklinischen Populationen zu untersuchen, die häufig eine bessere compliance bezüglich der Studienanforderungen zeigen und weniger potenziell konfundierende Faktoren wie Antipsychotikaeinnahme aufweisen. Die der Psychosen-Kontinuum-Hypothese zugrundeliegende Annahme eines fehlenden *qualitativen* Unterschieds zwischen Patient*innen und Gesunden ist allerdings nicht unumstritten, wobei ihre Kritiker*innen vor allem argumentieren, dass Patient*innen mit Schizophrenie qualitative Einzigartigkeiten aufweisen in den Inhalten ihrer psychotischen Symptome, in der Überzeugung von ihrer Richtigkeit und dem daraus entstehenden Leidensdruck (David, 2010; Lawrie, Hall, McIntosh, Owens, & Johnstone, 2010). Wenn psychosespezifische Veränderungen in Menschen ohne klinischer Schizophrenie untersucht werden, könnten daher bestimmte Veränderungen von Lernen und Entscheidungsfindung fehlen, die für den Übergang zu einer klinischen Erkrankung entscheidend sind. Dies könnte sich schließlich in inkonsistenten Ergebnissen je nach untersuchter Stichprobe widerspiegeln.

Ein daran anschließender Punkt betrifft die Präsenz und Kontrolle von potenziell konfundierenden unspezifischen Faktoren. Patient*innen mit schwersten psychischen Erkrankungen wie der Schizophrenie zeigen typischerweise Veränderungen in zahlreichen neuropsychologischen und sonstigen testpsychologischen Domänen (Palmer, Dawes, & Heaton, 2009), was die Spezifität einzelner Befunde oft schwer interpretierbar macht. Die Akquisition von zeitlich veränderbaren Vorannahmen in experimentellen Kontexten beispielsweise ist auch eine Leistung des Arbeitsgedächtnis, das bei Patient*innen mit Schizophrenie ausgeprägte Defizite aufweist (Forbes, Carrick, McIntosh, & Lawrie, 2009). Befunde, die eine reduzierte Nutzung von zeitlich transienten Vorannahmen zeigen bei Schizophrenie zeigen, könnten dementsprechend durch ein allgemeines Defizit des Arbeitsgedächtnis konfundiert sein.

Insgesamt wird es daher für den Fortschritt der Bayesianischen Psychosemodelle entscheidend sein, rigoros zu überprüfen, welche Veränderungen bei der Nutzung von welchen spezifischen Vorannahmen in welchen klinischen Stadien (wie subklinische Psychoseneigung, akute Erkrankung, chronische Erkrankung) vorliegen und welche neurophysiologischen Veränderungen ihnen korrespondieren (zum Beispiel unter Nutzung von hochauflösendem und modellbasierten fMRT, (Haarsma, Kok, & Browning, 2020)).

Auch unter Bezugnahme auf Bayesianische Psychosemodelle und der verwandten aberrant salience Theorie (Abschnitt 1.4), die den vorgestellten eigenen Arbeiten zugrunde liegen, gibt es Versuche, theoriebasiert Therapiekonzepte für psychotische Erkrankungen zu entwickeln. Der in diesem Zusammenhang wohl am besten untersuchte Ansatz ist das metakognitive Training. Dieses versucht, kognitive Verzerrungen, die auf eine gesteigerte Zuschreibung von Bedeutung zurückgeführt werden (wie der in Abschnitt 1.4 dargestellte jumping to conclusions bias), in einem gruppen- oder einzeltherapeutischen setting bewusst zu machen und damit möglichst zu korrigieren (Moritz et al., 2014). Die klinische Wirksamkeit dieser Therapieform gegen verschiedene psychotische Symptome ist inzwischen durch mehrere Studien belegt und eine Integration des metakognitiven Trainings in klinische Behandlungsleitlinien wird diskutiert (Penney et al., 2022).

Das Beispiel des metakognitiven Trainings zeigt, wie grundlagenwissenschaftliche Forschung zur Entwicklung innovativer Therapiekonzepte beitragen kann. Wie im folgenden Abschnitt ausgeführt wird, ist zudem ein entscheidendes Kriterium für die externe Validität der aus den theoretischen Modellen abgeleiteten Maße ihre Fähigkeit, den klinischen Verlauf vorherzusagen und damit potenziell die Therapie individuell zu steuern.

## *3.3 Nächste Schritte: vom theoretischen Modell zur individualisierten Behandlung in der Psychiatrie*

Die Fortschritte der letzten Jahrzehnte im Verständnis der Entstehung psychischer Erkrankungen wie Abhängigkeit oder Psychose konnten bislang nicht substanziell zu einer Verbesserung ihrer klinischen Behandlungsmöglichkeiten beitragen (Cuthbert & Insel, 2013; Kapur, Phillips, & Insel, 2012; Stephan et al., 2016). Aus klinisch-psychiatrischer Perspektive erhofft man sich von der Grundlagenforschung die Entwicklung von Markern, die das Vorliegen einer Erkrankung, ihre individuelle Prognose und idealerweise die Beeinflussbarkeit dieser Prognose durch verschiedene mögliche Therapieformen anzeigen: Um eine Translation grundlagenwissenschaftlicher Ergebnisse in die klinische Praxis zu ermöglichen, ist es also entscheidend, ihr allgemeines und therapiespezifisches prädiktives Potenzial zu erforschen. Dies könnte im Sinne einer personalisierten Medizin dabei helfen, Therapiekonzepte auf die individuellen biologischen, klinischen und psychosozialen Marker der Patient*innen abzustimmen (in Abgrenzung zur aktuellen Behandlung, die zumeist eine einheitliche Strategie

für Patient*innen mit identischer Diagnose umfasst).

Einige der in den hier vorgestellten Arbeiten berichteten krankheitsspezifische Veränderungen, die das Potenzial für derartige individuelle Marker haben könnten. In der Rauchentwöhnung beispielsweise könnte die psychotherapeutische Begleitung entsprechend der individuellen Veränderungen in der Verarbeitung von substanzbezogenen Belohnungsreizen, alternativen Belohnungsreizen und substanzbezogenen aversiven Reizen angepasst werden. Es wäre vorstellbar, dass Patient*innen mit einer stark reduzierten Reaktivität gegenüber alternativen (nicht-substanzbezogenen) Belohnungen eher von Interventionen wie zum Beispiel Genusstraining profitieren, während für Patient*innen, deren Reaktion auf bedrohliche Darstellungen von Rauchfolgen noch intakt ist, der Einsatz dieser Stimuli in der Therapie nützlich sein kann. Die Nutzung von Markern, die, ähnlich wie in der unter 2.6 vorgestellten Studie, aus der Anpassung von statistischen Modellen an individuelle Verhaltensdaten gewonnenen werden, könnte die Vorhersagekraft für Diagnose und Prognose zusätzlich erhöhen („computational assays“, (Huys, Maia, & Frank, 2016)). Schließlich könnte die Kombination mehrerer klinischer, biologischer und behavioraler Parameter mittels Methoden maschinellen Lernens die Genauigkeit von Vorhersagen weiter steigern ((Brandt et al., 2023; Gurevich, Stuke, Kastrup, Stuke, & Hildebrandt, 2017; Stuke, Priebe, Weilnhammer, Stuke, & Schoofs, 2023; Stuke et al., 2021) für eigene Arbeiten in dieser Richtung). Meiner Kenntnis nach gibt es zur Zeit kaum empirische Untersuchungen im Bereich von Abhängigkeitserkrankungen und Psychose, die den Vorteil einer solchen individualisierten Behandlung (verglichen mit einer einheitlichen Standardbehandlung) verglichen haben. Erste randomisiert-kontrollierte Studien im Bereich der Psychotherapie für depressive Erkrankungen weisen darauf hin, dass eine an Hand von Patientencharakteristika individualisierte Therapie einen Vorteil gegenüber evidenzbasierten, aber nicht individualisierten Interventionen haben könnte (Lutz et al., 2022; van Bronswijk et al., 2021). Entsprechende Studien im Bereich von Abhängigkeit und Psychose sowie unter Berücksichtigung eines breiteren Spektrums möglicher prädiktiver Marker erscheinen vielversprechend, um betroffenen Patient*innen ihre individuell optimale Therapie zukommen zu lassen.

## 4. Zusammenfassung

In der Entstehung von Abhängigkeit und Psychosen spielen Veränderungen in der Informationsverarbeitung (Lernen und Entscheidungsfindung) eine Rolle. Bei der Abhängigkeit werden basierend auf Vorarbeiten vor allem postuliert (1) eine verstärkte Belohnungsreaktion auf substanzbezogene Reize, (2) eine reduzierte Belohnungsreaktion auf nicht-substanzbezogene Belohnungen (wie zum Beispiel Speisen), (3) eine reduzierte kognitive Kontrolle gegenüber Belohnungsprozessen und (4) eine reduzierte Reaktion auf Konfrontation mit den negativen Folgen von Substanzkonsum. Bei Psychosen wird im Rückgriff auf Konzepte der Bayes Statistik eine veränderte Gewichtung von Vorannahmen und neuen Informationen postuliert, wobei der exakte Charakter dieser Veränderungen je nach klinischem Stadium und Art der Vorannahme zu variieren scheint und die bislang vorliegende Evidenz uneindeutig ist. Bei beiden Erkrankungen werden die beschriebenen Veränderungen in der Informationsverarbeitung mit spezifischen Pathologien des Dopaminsystems in Verbindung gebracht. Hier steht bei Abhängigkeit vor allem die Beteiligung des Dopamins an Belohnungsprozessen (die zugunsten der Substanz verändert sind) im Vordergrund, während bei Psychosen die Funktion des Dopamins als Lernsignal gestört zu sein scheint.

In eigenen Arbeiten zur Abhängigkeit konnten wir in diesem Kontext zeigen, dass das dopaminerge Belohnungssystem von Raucher*innen stärker auf substanzbezogene Reize und schwächer auf alternative Belohnungsreize reagiert und bei Menschen mit stärkerem Alkoholkonsum stärker aktiviert ist, wenn sie Entscheidungen für den Konsum alkoholischer Getränke treffen. Ferner konnten wir zeigen, dass sich die Reaktivität auf bedrohliche Darstellungen von körperlichen Folgen des Rauchens zwar nicht signifikant zwischen Raucher*innen und Nichtraucher*innen unterscheidet, dass aber bei Raucher*innen durch die vorhergehende Präsentation dieser Darstellungen Kontrollprozesse während der anschließenden Präsentation substanzbezogener Reize evoziert werden. Schließlich fanden wir, konträr zur ursprünglichen Hypothese und einigen Vorbefunden, keine reduzierte Aktivität von kontrollassoziierten Arealen während der Regulierung von Verlangen bei Raucher*innen und während Entscheidungen für Alkoholkonsum bei zunehmender Trinkschwere von Proband*innen mit missbräuchlichem Alkoholkonsum. Im Zusammenhang mit zunehmender (subklinischer) Psychoseneigung fanden wir ein relatives Übergewicht von neuen Informationen

(einen reduzierten Einfluss von Vorannahmen) bei perzeptueller, aber nicht bei probabilistischer Entscheidungsfindung. Desweiteren konnten wir einen verstärkten Einfluss von Vorannahmen bezüglich sozial bedeutsamer Reize (Gesichter und direkter Blick) nachweisen. Schließlich fanden wir im Zusammenhang mit einer stärkeren Psychoseneigung ein verändertes Lernen aus neuen Informationen im Sinne einer reduzierten Suppression unwahrscheinlicher / überraschender Informationen.

Als Limitation der präsentierten Arbeiten muss zunächst die teilweise geringe Fallzahl angeführt werden. Desweiteren waren die Ergebnisse, vor allem der Studien zur Psychose, auch unter Einbeziehung von Vorarbeiten teils widersprüchlich was den Einfluss von Vorannahmen in verschiedenen Aufgaben betrifft, sodass sich das Versprechen der Theorie, die Vielgestaltigkeit psychotischer Symptome auf wenige Kernpathologien zurückzuführen, nicht einzulösen scheint und wiederum komplexere Modelle aufgestellt werden müssen.

Trotz der genannten Limitationen stellen die Arbeiten wichtige Bausteine für die notwendige Unternehmung dar, ein gemeinsames kognitiv-neurobiologisches Verständnis von psychiatrischen Erkrankungen zu entwickeln. Der entscheidende Validierungsschritt für die aus den Theorien abgeleiteten Tests und Markern (sowohl neurophysiologische fMRT-Marker als auch computionale Marker der Verhaltensmodellierung) ist die Translation in die Praxis. Zukünftige Studien werden zeigen müssen, ob diese Marker dazu geeignet sind, Diagnose, Prognose und Ansprechen auf spezifische Therapien vorherzusagen und damit das Leben der betroffenen Patient*innen real zu verbessern.

## 5. Literatur


Acuff, S. F., Dennhardt, A. A., Correia, C. J., & Murphy, J. G. (2019). Measurement of substance-free reinforcement in addiction: A systematic review. *Clin Psychol Rev, 70*, 79-90. doi:10.1016/j.cpr.2019.04.003

Adams, R. A., Napier, G., Roiser, J. P., Mathys, C., & Gilleen, J. (2018). Attractor-like Dynamics in Belief Updating in Schizophrenia. *J Neurosci, 38*(44), 9471-9485. doi:10.1523/JNEUROSCI.3163-17.2018

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Front Psychiatry, 4*, 47. doi:10.3389/fpsyt.2013.00047

Aggelopoulos, N. C. (2015). Perceptual inference. *Neurosci Biobehav Rev, 55*, 375-392. doi:10.1016/j.neubiorev.2015.05.001

Alderson-Day, B., Lima, C. F., Evans, S., Krishnan, S., Shanmugalingam, P., Fernyhough, C., & Scott, S. K. (2017). Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. *Brain, 140*(9), 2475-2489. doi:10.1093/brain/awx206

Baik, J. H. (2013). Dopamine signaling in reward-related behaviors. *Front Neural Circuits, 7*, 152. doi:10.3389/fncir.2013.00152

Baler, R. D., & Volkow, N. D. (2006). Drug addiction: the neurobiology of disrupted self-control. *Trends Mol Med, 12*(12), 559-566. doi:10.1016/j.molmed.2006.10.005

Barlow, H. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Res, 30*(11), 1561-1571. doi:10.1016/0042-6989(90)90144-a

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). The Cardiff Anomalous Perceptions Scale (CAPS): a new validated measure of anomalous perceptual experience. *Schizophr Bull, 32*(2), 366-377. doi:10.1093/schbul/sbj014

Berke, J. D. (2018). What does dopamine mean? *Nat Neurosci, 21*(6), 787-793. doi:10.1038/s41593-018-0152-y

Brandt, L., Ritter, K., Schneider-Thoma, J., Siafis, S., Montag, C., Ayrilmaz, H., . . . Stuke, H. (2023). Predicting psychotic relapse following randomised discontinuation of paliperidone in individuals with schizophrenia or schizoaffective disorder: an individual participant data analysis. *Lancet Psychiatry, 10*(3), 184-196. doi:10.1016/S2215-0366(23)00008-1

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron, 68*(5), 815-834. doi:10.1016/j.neuron.2010.11.022

Campbell, W. G. (2003). Addiction: a disease of volition caused by a cognitive impairment. *Can J Psychiatry, 48*(10), 669-674. doi:10.1177/070674370304801005

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Arch Clin Neuropsychol, 23*(2), 201-216. doi:10.1016/j.acn.2007.08.010

Chapman, L. J., Chapman, J. P., Kwapil, T. R., Eckblad, M., & Zinser, M. C. (1994). Putatively psychosis-

prone subjects 10 years later. *J Abnorm Psychol, 103*(2), 171-183. doi:10.1037//0021-843x.103.2.171

Collins, A. L., & Saunders, B. T. (2020). Heterogeneity in striatal dopamine circuits: Form and function in dynamic reward seeking. *J Neurosci Res, 98*(6), 1046-1069. doi:10.1002/jnr.24587

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R., 3rd. (2019). Hallucinations and Strong Priors. *Trends Cogn Sci, 23*(2), 114-127. doi:10.1016/j.tics.2018.12.001

Croft, J., Teufel, C., Heron, J., Fletcher, P. C., David, A. S., Lewis, G., . . . Zammit, S. (2021). A Computational Analysis of Abnormal Belief Updating Processes and Their Association With Psychotic Experiences and Childhood Trauma in a UK Birth Cohort. *Biol Psychiatry Cogn Neurosci Neuroimaging*. doi:10.1016/j.bpsc.2021.12.007

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med, 11*, 126. doi:10.1186/1741-7015-11-126

David, A. S. (2010). Why we need more debate on whether psychotic symptoms lie on a continuum with normality. *Psychol Med, 40*(12), 1935-1942. doi:10.1017/S0033291710000188

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends Cogn Sci, 22*(9), 764-779. doi:10.1016/j.tics.2018.06.002

Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2016). Psychosis, Delusions and the "Jumping to Conclusions" Reasoning Bias: A Systematic Review and Meta-analysis. *Schizophr Bull, 42*(3), 652-665. doi:10.1093/schbul/sbv150

Fanous, A., Gardner, C., Walsh, D., & Kendler, K. S. (2001). Relationship between positive and negative symptoms of schizophrenia and schizotypal symptoms in nonpsychotic relatives. *Arch Gen Psychiatry, 58*(7), 669-673. doi:10.1001/archpsyc.58.7.669

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci, 10*(1), 48-58. doi:10.1038/nrn2536

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. *Psychol Med, 39*(6), 889-905. doi:10.1017/S0033291708004558

Garfield, J. B., Lubman, D. I., & Yucel, M. (2014). Anhedonia in substance use disorders: a systematic review of its nature, course and clinical correlates. *Aust N Z J Psychiatry, 48*(1), 36-51. doi:10.1177/0004867413508455

Gayet, S., Van der Stigchel, S., & Paffen, C. L. (2014). Breaking continuous flash suppression: competing for consciousness on the pre-semantic battlefield. *Front Psychol, 5*, 460. doi:10.3389/fpsyg.2014.00460

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci, 12*(11), 652-669. doi:10.1038/nrn3119

Gurevich, P., Stuke, H., Kastrup, A., Stuke, H., & Hildebrandt, H. (2017). Neuropsychological Testing and Machine Learning Distinguish Alzheimer's Disease from Other Causes for Cognitive Impairment. *Front Aging Neurosci, 9*, 114. doi:10.3389/fnagi.2017.00114

Haarsma, J., Knolle, F., Griffin, J. D., Taverne, H., Mada, M., Goodyer, I. M., . . . Murray, G. K. (2020). Influence of prior beliefs on perception in early psychosis: Effects of illness stage and hierarchical level of belief. *J Abnorm Psychol, 129*(6), 581-598. doi:10.1037/abn0000494

Haarsma, J., Kok, P., & Browning, M. (2020). The promise of layer-specific neuroimaging for testing predictive coding theories of psychosis. *Schizophr Res*. doi:10.1016/j.schres.2020.10.009

Hammond, D. (2011). Health warning messages on tobacco products: a review. *Tob Control, 20*(5), 327-337. doi:10.1136/tc.2010.037630

Hanssen, M., Bak, M., Bijl, R., Vollebergh, W., & van Os, J. (2005). The incidence and outcome of subclinical psychotic experiences in the general population. *Br J Clin Psychol, 44*(Pt 2), 181-191. doi:10.1348/014466505X29611

Hatzigiakoumis, D. S., Martinotti, G., Giannantonio, M. D., & Janiri, L. (2011). Anhedonia and substance dependence: clinical correlates and treatment options. *Front Psychiatry, 2*, 10. doi:10.3389/fpsyt.2011.00010

Heinz, A. (2002). Dopaminergic dysfunction in alcoholism and schizophrenia--psychopathological and behavioral correlates. *Eur Psychiatry, 17*(1), 9-16. doi:10.1016/s0924-9338(02)00628-4

Heinz, A., Murray, G. K., Schlagenhauf, F., Sterzer, P., Grace, A. A., & Waltz, J. A. (2019). Towards a Unifying Cognitive, Neurophysiological, and Computational Neuroscience Account of Schizophrenia. *Schizophr Bull, 45*(5), 1092-1100. doi:10.1093/schbul/sby154

Hommer, D. W., Bjork, J. M., & Gilman, J. M. (2011). Imaging brain response to reward in addictive disorders. *Ann N Y Acad Sci, 1216*, 50-61. doi:10.1111/j.1749-6632.2010.05898.x

Howes, O. D., & Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr Bull, 35*(3), 549-562. doi:10.1093/schbul/sbp006

Howes, O. D., & Nour, M. M. (2016). Dopamine and the aberrant salience hypothesis of schizophrenia. *World Psychiatry, 15*(1), 3-4. doi:10.1002/wps.20276

Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci, 19*(3), 404-413. doi:10.1038/nn.4238

Jardri, R., Duverne, S., Litvinova, A. S., & Deneve, S. (2017). Experimental evidence for circular inference in schizophrenia. *Nat Commun, 8*, 14218. doi:10.1038/ncomms14218

Jentsch, J. D., & Pennington, Z. T. (2014). Reward, interrupted: Inhibitory control and its relevance to addictions. *Neuropharmacology, 76 Pt B*, 479-486. doi:10.1016/j.neuropharm.2013.05.022

Kakoschke, N., Kemps, E., & Tiggemann, M. (2017). Approach bias modification training and consumption: A review of the literature. *Addict Behav, 64*, 21-28. doi:10.1016/j.addbeh.2016.08.007

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry, 160*(1), 13-23. doi:10.1176/appi.ajp.160.1.13

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry, 17*(12), 1174-1179.

doi:10.1038/mp.2012.105

Keiflin, R., & Janak, P. H. (2015). Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. *Neuron, 88*(2), 247-263. doi:10.1016/j.neuron.2015.08.037

Kendler, K. S., McGuire, M., Gruenberg, A. M., O'Hare, A., Spellman, M., & Walsh, D. (1993). The Roscommon Family Study. III. Schizophrenia-related personality disorders in relatives. *Arch Gen Psychiatry, 50*(10), 781-788. doi:10.1001/archpsyc.1993.01820220033004

Klein, M. O., Battagello, D. S., Cardoso, A. R., Hauser, D. N., Bittencourt, J. C., & Correa, R. G. (2019). Dopamine: Functions, Signaling, and Association with Neurological Diseases. *Cell Mol Neurobiol, 39*(1), 31-59. doi:10.1007/s10571-018-0632-3

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology, 24*(2), 97-129. doi:10.1016/S0893-133X(00)00195-0

Kunas, S. L., Stuke, H., Heinz, A., Strohle, A., & Bermpohl, F. (2022). Evidence for a hijacked brain reward system but no desensitized threat system in quitting-motivated smokers: An fMRI study. *Addiction, 117*(3), 701-712. doi:10.1111/add.15651

Kunas, S. L., Stuke, H., Plank, I. S., Laing, E. M., Bermpohl, F., & Strohle, A. (2022). Neurofunctional alterations of cognitive down-regulation of craving in quitting motivated smokers. *Psychol Addict Behav*. doi:10.1037/adb0000820

Lammel, S., Lim, B. K., & Malenka, R. C. (2014). Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology, 76 Pt B*(0 0), 351-359. doi:10.1016/j.neuropharm.2013.03.019

Lawrie, S. M., Hall, J., McIntosh, A. M., Owens, D. G., & Johnstone, E. C. (2010). The 'continuum of psychosis': scientifically unproven and clinically impractical. *Br J Psychiatry, 197*(6), 423-425. doi:10.1192/bjp.bp.109.072827

Lin, X., Deng, J., Shi, L., Wang, Q., Li, P., Li, H., . . . Lu, L. (2020). Neural substrates of smoking and reward cue reactivity in smokers: a meta-analysis of fMRI studies. *Transl Psychiatry, 10*(1), 97. doi:10.1038/s41398-020-0775-0

Linscott, R. J., & van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychol Med, 43*(6), 1133-1149. doi:10.1017/S0033291712001626

Luijten, M., Machielsen, M. W., Veltman, D. J., Hester, R., de Haan, L., & Franken, I. H. (2014). Systematic review of ERP and fMRI studies investigating inhibitory control and error processing in people with substance dependence and behavioural addictions. *J Psychiatry Neurosci, 39*(3), 149-169. doi:10.1503/jpn.130052

Lutz, W., Deisenhofer, A. K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. *J Consult Clin Psychol, 90*(1), 90-106. doi:10.1037/ccp0000642

Maia, T. V., & Frank, M. J. (2017). An Integrative Perspective on the Role of Dopamine in Schizophrenia. *Biol Psychiatry, 81*(1), 52-66. doi:10.1016/j.biopsych.2016.05.021

Mather, M., & Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspect Psychol Sci, 6*(2), 114-133. doi:10.1177/1745691611400234

McClure, S. M., & Bickel, W. K. (2014). A dual-systems perspective on addiction: contributions from neuroimaging and cognitive training. *Ann N Y Acad Sci, 1327*, 62-78. doi:10.1111/nyas.12561

Moeller, S. J., & Paulus, M. P. (2018). Toward biomarkers of the addicted human brain: Using neuroimaging to predict relapse and sustained abstinence in substance use disorder. *Prog Neuropsychopharmacol Biol Psychiatry, 80*(Pt B), 143-154. doi:10.1016/j.pnpbp.2017.03.003

Moritz, S., Andreou, C., Schneider, B. C., Wittekind, C. E., Menon, M., Balzan, R. P., & Woodward, T. S. (2014). Sowing the seeds of doubt: a narrative review on metacognitive training in schizophrenia. *Clin Psychol Rev, 34*(4), 358-366. doi:10.1016/j.cpr.2014.04.004

Noori, H. R., Cosa Linan, A., & Spanagel, R. (2016). Largely overlapping neuronal substrates of reactivity to drug, gambling, food and sexual cues: A comprehensive meta-analysis. *Eur Neuropsychopharmacol, 26*(9), 1419-1430. doi:10.1016/j.euroneuro.2016.06.013

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45*(1), 75-92. doi:10.1016/j.neuropsychologia.2006.04.025

Palmer, B. W., Dawes, S. E., & Heaton, R. K. (2009). What do we know about neuropsychological aspects of schizophrenia? *Neuropsychol Rev, 19*(3), 365-384. doi:10.1007/s11065-009-9109-y

Pang, B., Saleme, P., Seydel, T., Kim, J., Knox, K., & Rundle-Thiele, S. (2021). The effectiveness of graphic health warnings on tobacco products: a systematic review on perceived harm and quit intentions. *BMC Public Health, 21*(1), 884. doi:10.1186/s12889-021-10810-z

Penney, D., Sauve, G., Mendelson, D., Thibaudeau, E., Moritz, S., & Lepage, M. (2022). Immediate and Sustained Outcomes and Moderators Associated With Metacognitive Training for Psychosis: A Systematic Review and Meta-analysis. *JAMA Psychiatry, 79*(5), 417-429. doi:10.1001/jamapsychiatry.2022.0277

Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). *Schizophr Bull, 30*(4), 1005-1022. doi:10.1093/oxfordjournals.schbul.a007116

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *J Exp Psychol, 72*(3), 346-354. doi:10.1037/h0023653

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science, 357*(6351), 596-600. doi:10.1126/science.aan3458

Rehm, J., Taylor, B., & Room, R. (2006). Global burden of disease from alcohol, illicit drugs and tobacco. *Drug Alcohol Rev, 25*(6), 503-513. doi:10.1080/09595230600944453

Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw, 15*(4-6), 507-521. doi:10.1016/s0893-6080(02)00045-x

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res Brain Res Rev, 18*(3), 247-291. doi:10.1016/0165-0173(93)90013-p

Robinson, T. E., & Berridge, K. C. (2001). Incentive-sensitization and addiction. *Addiction, 96*(1), 103-114. doi:10.1046/j.1360-0443.2001.9611038.x

Schacht, J. P., Anton, R. F., & Myrick, H. (2013). Functional neuroimaging studies of alcohol cue reactivity: a quantitative meta-analysis and systematic review. *Addict Biol, 18*(1), 121-133. doi:10.1111/j.1369-1600.2012.00464.x

Schmack, K., Bosc, M., Ott, T., Sturgill, J. F., & Kepecs, A. (2021). Striatal dopamine mediates hallucination-like perception in mice. *Science, 372*(6537). doi:10.1126/science.abf4740

Schmack, K., Gomez-Carrillo de Castro, A., Rothkirch, M., Sekutowicz, M., Rossler, H., Haynes, J. D., . . . Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. *J Neurosci, 33*(34), 13701-13712. doi:10.1523/JNEUROSCI.1778-13.2013

Schmack, K., Schnack, A., Priller, J., & Sterzer, P. (2015). Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophr Res Cogn, 2*(2), 72-77. doi:10.1016/j.scog.2015.03.005

Schoenmakers, T. M., de Bruin, M., Lux, I. F., Goertz, A. G., Van Kerkhof, D. H., & Wiers, R. W. (2010). Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. *Drug Alcohol Depend, 109*(1-3), 30-36. doi:10.1016/j.drugalcdep.2009.11.022

Schultz, W. (2007). Behavioral dopamine signals. *Trends Neurosci, 30*(5), 203-210. doi:10.1016/j.tins.2007.03.007

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599. doi:10.1126/science.275.5306.1593

Seeman, P., Lee, T., Chau-Wong, M., & Wong, K. (1976). Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature, 261*(5562), 717-719. doi:10.1038/261717a0

Series, P., & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Front Hum Neurosci, 7*, 668. doi:10.3389/fnhum.2013.00668

Shevlin, M., McElroy, E., Bentall, R. P., Reininghaus, U., & Murphy, J. (2017). The Psychosis Continuum: Testing a Bifactor Model of Psychosis in a General Population Sample. *Schizophr Bull, 43*(1), 133-141. doi:10.1093/schbul/sbw067

Speranza, L., di Porzio, U., Viggiano, D., de Donato, A., & Volpicelli, F. (2021). Dopamine: The Neuromodulator of Long-Term Synaptic Plasticity, Reward and Movement Control. *Cells, 10*(4). doi:10.3390/cells10040735

Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., . . . Breakspear, M. (2016). Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *Lancet Psychiatry, 3*(1), 77-83. doi:10.1016/S2215-0366(15)00361-2

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., . . . Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. *Biol Psychiatry, 84*(9), 634-643. doi:10.1016/j.biopsych.2018.05.015

Stuke, H., Gutwinski, S., Wiers, C. E., Schmidt, T. T., Gropper, S., Parnack, J., . . . Bermpohl, F. (2016). To drink or not to drink: Harmful drinking is associated with hyperactivation of reward areas rather than hypoactivation of control areas in men. *J Psychiatry Neurosci, 41*(3), E24-36.

Stuke, H., Priebe, K., Weilnhammer, V. A., Stuke, H., & Schoofs, N. (2023). Sparse models for predicting psychosocial impairments in patients with PTSD: An empirical Bayes approach. *Psychol Trauma, 15*(1), 80-87. doi:10.1037/tra0001279

Stuke, H., Schoofs, N., Johanssen, H., Bermpohl, F., Ulsmann, D., Schulte-Herbruggen, O., & Priebe, K. (2021). Predicting outcome of daycare cognitive behavioural therapy in a naturalistic sample of patients with PTSD: a machine learning approach. *Eur J Psychotraumatol, 12*(1), 1958471. doi:10.1080/20008198.2021.1958471

Tang, D. W., Fellows, L. K., Small, D. M., & Dagher, A. (2012). Food and drug cues activate similar brain regions: a meta-analysis of functional MRI studies. *Physiol Behav, 106*(3), 317-324. doi:10.1016/j.physbeh.2012.03.009

Tanji, J., & Hoshi, E. (2008). Role of the lateral prefrontal cortex in executive behavioral control. *Physiol Rev, 88*(1), 37-57. doi:10.1152/physrev.00014.2007

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., . . . Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc Natl Acad Sci U S A, 112*(43), 13401-13406. doi:10.1073/pnas.1503916112

Theodoridou, A., & Rössler, W. (2010). Disease Burden and Disability-Adjusted Life Years Due to Schizophrenia and Psychotic Disorders. In V. R. Preedy & R. R. Watson (Eds.), *Handbook of Disease Burdens and Quality of Life Measures* (pp. 1493-1507). New York, NY: Springer New York.

Tienari, P., Wynne, L. C., Laksy, K., Moring, J., Nieminen, P., Sorri, A., . . . Wahlberg, K. E. (2003). Genetic boundaries of the schizophrenia spectrum: evidence from the Finnish Adoptive Family Study of Schizophrenia. *Am J Psychiatry, 160*(9), 1587-1594. doi:10.1176/appi.ajp.160.9.1587

Valton, V., Karvelis, P., Richards, K. L., Seitz, A. R., Lawrie, S. M., & Series, P. (2019). Acquisition of visual priors and induced hallucinations in chronic schizophrenia. *Brain, 142*(8), 2523-2537. doi:10.1093/brain/awz171

van Bronswijk, S. C., DeRubeis, R. J., Lemmens, L., Peeters, F., Keefe, J. R., Cohen, Z. D., & Huibers, M. J. H. (2021). Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: cognitive therapy or interpersonal psychotherapy? *Psychol Med, 51*(2), 279-289. doi:10.1017/S0033291719003192

van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol Med, 39*(2), 179-195. doi:10.1017/S0033291708003814

Versace, F., Engelmann, J. M., Deweese, M. M., Robinson, J. D., Green, C. E., Lam, C. Y., . . . Cinciripini, P. M. (2017). Beyond Cue Reactivity: Non-Drug-Related Motivationally Relevant Stimuli Are Necessary to Understand Reactivity to Drug-Related Cues. *Nicotine Tob Res, 19*(6), 663-669. doi:10.1093/ntr/ntx002

Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *Lancet Psychiatry, 3*(2), 171-178. doi:10.1016/S2215-0366(15)00505-2

Weilnhammer, V., Rod, L., Eckert, A. L., Stuke, H., Heinz, A., & Sterzer, P. (2020). Psychotic Experiences in Schizophrenia and Sensitivity to Sensory Evidence. *Schizophr Bull, 46*(4), 927-936. doi:10.1093/schbul/sbaa003

Welham, J., Scott, J., Williams, G., Najman, J., Bor, W., O'Callaghan, M., & McGrath, J. (2009). Emotional and behavioural antecedents of young adults who screen positive for non-affective psychosis: a 21-year birth cohort study. *Psychol Med, 39*(4), 625-634. doi:10.1017/S0033291708003760

Wiers, C. E., Stelzel, C., Gladwin, T. E., Park, S. Q., Pawelczack, S., Gawron, C. K., . . . Bermpohl, F. (2015). Effects of cognitive bias modification training on neural alcohol cue reactivity in alcohol dependence. *Am J Psychiatry, 172*(4), 335-343. doi:10.1176/appi.ajp.2014.13111495

Wiers, C. E., Stelzel, C., Park, S. Q., Gawron, C. K., Ludwig, V. U., Gutwinski, S., . . . Bermpohl, F. (2014). Neural correlates of alcohol-approach bias in alcohol addiction: the spirit is willing but the flesh is weak for spirits. *Neuropsychopharmacology, 39*(3), 688-697. doi:10.1038/npp.2013.252

Wise, R. A. (2004). Dopamine, learning and motivation. *Nat Rev Neurosci, 5*(6), 483-494. doi:10.1038/nrn1406

World Health Organization. (2016). International statistical classification of diseases and related health problems (10th ed.). https://icd.who.int/browse10/2016/en

Zeki, S., & Chen, O. Y. (2020). The Bayesian-Laplacian brain. *Eur J Neurosci, 51*(6), 1441-1462. doi:10.1111/ejn.14540

## Danksagung

Aus Datenschutzgründen in der Internetfassung entfernt

# Erklärung

§ 4 Abs. 3 (k) der HabOMed der Charité

Hiermit erkläre ich, dass

- weder früher noch gleichzeitig ein Habilitationsverfahren durchgeführt oder angemeldet wurde,
- die vorgelegte Habilitationsschrift ohne fremde Hilfe verfasst, die beschriebenen Ergebnisse selbst gewonnen sowie die verwendeten Hilfsmittel, die Zusammenarbeit mit anderen Wissenschaftlern/Wissenschaftlerinnen und mit technischen Hilfskräften sowie die verwendete Literatur vollständig in der Habilitationsschrift angegeben wurden,
- mir die geltende Habilitationsordnung bekannt ist.

Ich erkläre ferner, dass mir die Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis bekannt ist und ich mich zur Einhaltung dieser Satzung verpflichte.

24.10.2023

…… ……..

Datum Unterschrift