

**At måle vaner: et kritisk blik på slips-of-action-testen**

Freja Waade Biermann

Aalborg Universitet

Kandidatspeciale

Vejleder: Thomas Alrik Sørensen

01.02.2022

Abstract.....	2
Indledning - mod en ny form for diagnostik .....	2
Vaner og psykisk lidelse .....	8
Operant betingning.....	11
Vaner defineret.....	12
Vaner er S-R-associationer.....	12
Operationalisering af vaner .....	14
Vaner skabes gennem gentagelse af målrettede handlinger .....	16
Automaticitet .....	17
Delkonklusion.....	18
Neurobiologien bag vaner .....	19
Modelfri og modelbaseret læring.....	21
Modelfri og modelbaseret læring i relation til vaner .....	26
Er vaner og modelfri indlæring det samme? .....	27
Delkonklusion.....	30
Slips of action testen.....	31
Diskussion – problemer med slips-of-action testen .....	32
Model-fri læring som mulig forstyrrende variabel .....	32
Forskellig indkodning af udfald og stimuli .....	34
Forslag til ændringer i Slips-Of-action-testen .....	36
Delkonklusion.....	41
Konklusion.....	42
Referencer.....	43

## **Abstract**

The Slips-Of-Action test is listed as a paradigm under the construct of habits in the RDoC matrix. This paper examines the validity of the Slips-Of-Action test as a measure of habits. It is proposed that the construct of model-free learning, rather than being equal to habits, is actually a possible confounding variable in the Slips-Of-Action-test. It is suggested that stimuli and udfald are likely to be encoded differently, increasing the risk of the Slips-Of-Action test being a measure of model-free learning rather than habits. A new version of the slips-of-action is considered, which provides a better control of the construct of model free learning.

## **Indledning - mod en ny form for diagnostik**

De to mest kendte og brugte diagnostiske systemer inden for psykologien i dag er *Diagnostic and Statistical Manual of Mental Disorders* (DSM) udgivet af American Psychological Association og kapitel 6 om psykiske- og adfærdsmæssige lidelser i World Health Organisation's (WHO) *International Classification of Diseases* (ICD). ICD-manualen er det officielle verdensomspændende diagnosesystem. Det er designet til at kunne bruges af sundhedspersonale med forskellige fagligheder og niveauer af ekspertise, og på tværs af forskellige kulturer, med forskellige opfattelser af psykiske lidelsers natur og ætiologi. DSM er derimod designet til brug af højtuddannet sundhedspersonale og benyttes mest i USA. Disse forskelle afspejler sig i manualernes opbygning (Cuthbert & Insel, 2013). Fælles for de to er, at de har til formål at standardisere klassificeringen af psykiske lidelser, så terminologi og kategorisering bliver ensrettet på tværs af kulturelle praksisser og landegrænser, for på den måde at facilitere klinisk praksis (Clark, Watson & Reynolds, 1995).

I 2009 påbegyndte National Institute of Mental Health (NIMH) et projekt, som adskiller sig markant fra de to andre systemer. Projektet hedder *Research Domain Criteria* (forkortes oftest RDoC) og det adskiller sig fra de andre to klassifikationssystemer fordi det retter sig

mod forskning snarere end klinisk brug (Cuthbert & Insel, 2013). Motivationen for udviklingen af dette nye system er en række kritikpunkter der er blevet fremstillet omkring ICD og DSM-systemerne.

Et af kritikpunkterne er at ICD og DSM-systemerne er baseret udelukkende på symptompræsentation og lidelsernes udvikling (Clark, Watson & Reynolds, 1995; Hyman, 2007). Sat på spidsen kan man sige at når man klassificerer psykiske lidelser udelukkende på baggrund af symptomer, kan det sammenlignes med at man går til lægen med ondt i hovedet og får fortalt at man lider af hovedpine. De fleste ville nok være lidt skuffede i den situation, da de havde håbet på at få undersøgt hvad hovedpinen skyldes, for eksempel migræne, influenza, tømmermænd, dårligt syn eller spændinger i nakken, og med udgangspunkt i det, få en passende behandling. Patienten kender selv symptomerne, men er interesseret i *årsagen*. Derudover er de blevet kritiseret for at overse spørgsmål om ætiologi og patogenese (Clark, Watson & Reynolds, 1995). Som beskrevet, har de primære formål med ICD og DSM heller ikke været at kortlægge ætiologi. Formålet har været at standardisere klassificeringerne og dermed skabe større reliabilitet i klinisk praksis (Cuthbert & Insel, 2013; Clark, Watson & Reynolds, 1995; Miller & Rockstroh, 2013). Dette fokus på diagnostisk reliabilitet i klinisk praksis betyder imidlertid at patogenetisk og ætiologisk vidt forskellige lidelser bliver samlet under samme klassificering (Hyman, 2007; Miller & Rosckstrog, 2013). Der er således en høj grad af heterogenitet i diagnoserne, i form af forskellige patogenetiske og ætiologiske karakteristikker (Clark, Watson & Reynolds, 1995). På baggrund af dette er det ikke overraskende at meget af den medicin, der bliver fremstillet, kun er effektiv for halvdelen af en given patientgruppe; i de fleste tilfælde er den blevet testet på heterogene syndromer, med forskellige ætiologier og patofysiologiske mekanismer (Wong et al., 2010). Dertil kommer at det meste medicin i dag bruges til behandling af flere forskellige diagnoser (Miller, 2010). Til sidst er det blevet fremhævet at de klassiske diagnosesystemer behandler lidelserne

kategorisk. Det vil for eksempel sige at 'syg' og 'rask' anses for at være kvalitativt adskilte (Hyman, 2007; Miller & Rosckstrog, 2013). Du er enten syg eller også er du rask, og der er ikke nogen glidende overgang. Ifølge meget nyere forskning kan mange psykiatriske diagnoser dog formentligt bedre forstås som dimensionelle (Fanous et al., 2001; Hoekstra et al., 2007; Meehl, 1989; Kendler & Gardner, 1998; Tsuang et al., 2003; Widiger & Costa, 1994). Det betyder at de snarere ligger på et kontinuum hvor symptomernes karakter og varighed bevæger sig i en glidende overgang fra rask til syg (Hyman, 2007; Kendler & Gardner, 1998; Yee, Javitt & Miller, 2015).

Det er vigtigt at pointere at de diagnosesystemer vi opererer med idag naturligvis er værdifulde og brugbare. Familie og tvillingestudier har vist stor arvelighed i mange af de diagnoser der indgår i systemerne, så de er langt fra arbitrære (Hyman, 2007). Blot upræcise. Problemet med dem er at det er svært at finde sammenhænge mellem biologiske systemer og adfærd med udgangspunkt i de diagnoser vi har i dag (Cuthbert & Insel, 2013). Og den viden er vigtig hvis vi skal udvikle mere præcise og effektive behandlinger til psykiatriske diagnoser (Cuthbert & Insel, 2013). Når diagnoserne ikke præcist afspejler de underliggende biologiske mekanismer (Hyman, 2007) kan én mekanisme være involveret i flere forskellige lidelser. Samtidig er de fleste lidelser associerede med mere end én underliggende neural mekanisme (Cuthbert & Insel, 2013). Det gør det praktisk talt umuligt at undersøge biologiske mekanismer inden for de traditionelle diagnostiske systemer (Cuthbert & Insel, 2013). Alligevel er ICD og DSM-systemerne stadig de primære systemer, hvis man vil opnå forskningsmidler og godkendelser til forskning i felter som medicinudvikling og ætiologi (Cuthbert & Insel, 2013). Dette er højst sandsynligt en del af årsagen til den manglende fremgang inden for psykiatri og neurovidenskab, sammenlignet med andre grene af medicin (Cuthbert & Insel, 2013; Yee, Javitt & Miller, 2015).

På baggrund af disse problematikker formulerede *National Institute of Mental Health* (NIMH) i 2008 en målsætning om at udvikle nye måder at klassificere psykiske lidelser. De nye klassificeringer skulle være på baggrund af objektive mål af adfærd og neurobiologiske mekanismer og sammenhængen mellem de to (National Institute of Mental Health, 2008). Det mål manifesterede sig i et nyt klassificeringssystem, under navnet *Research Domain Criteria*, som bliver forkortet RDoC. RDoC har to trin; det første går ud på at kortlægge de fundamentale adfærdsmæssige funktioner, som hjernen har udviklet sig til at udføre, og at specificere de neurale systemer, der implementerer dem. Det andet trin går ud på at undersøge hvordan psykopatologi hænger sammen med forskellige grader af- og former for dysfunktion i de specifikke systemer (Cuthbert & Insel, 2013, pp. 4). Inden for RDoC er det altså i første omgang den normale hjerne der undersøges og der antages ikke på forhånd noget kausalt forhold til de kategoriske diagnoser (Yee, Javitt & Miller, 2017). RDoC tager altså udgangspunkt i processer i den normale hjerne, og forstår den patologiske hjerne som dysfunktioner i dimensionelle konstrukter.

Der er formuleret nogle konkrete overordnede mål for RDoC-projektet, som består i 1) at samle relevante eksperter, både klinikere og forskere, for at identificere fundamentale komponenter i adfærd, som er kompatible med neurovidenskabelige tilgange, 2) kortlægge variationen i disse komponenter, fra normal til abnormal, 3) udvikle pålidelige og valide mål for disse både i forskning og klinisk brug, samt 4) at integrere genetik, neurobiologi, adfærd, miljø og oplevelse i forståelsen af disse komponenter og mental sygdom (Cuthbert & Insel, 2013).

En adfærdskomponent eller dimension bliver, i flugt med punkt 1, kun inkluderet i RDoCs matrice, hvis den lever op til to kriterier. Der skal være stærk evidens for det psykologiske konstrukt's validitet i sig selv. Og der skal være stærk evidens for at

konstruktet kan associeres med et specifikt biologisk system, som for eksempel et neuralt kredsløb (Yee, Javitt & Miller, 2015).

I henhold til punkt 2 arbejdes der i RDoC med en dimensionel tilgang til hjernens funktion og dysfunktion (Yee, Javitt & Miller, 2015). De kategorielle diagnoser er altså erstattet med dimensioner som både er relevante for den normale og den syge hjerne og ofte også relevant for mere end én af de traditionelle diagnoser. De valide målemetoder, der skal udvikles ifølge punkt 3 er altså mål for disse dimensioner. Det har en betydning for forskningsdesignet og sampling-strategierne i RDoC-studier. I klassiske kliniske studier er den uafhængige variabel ofte en diagnose fra ICD eller DSM. I et RDoC studie vælger forskeren den uafhængige variabel på et hvilket som helst analyseniveau, så det bedst besvarer det specifikke eksperimentelle spørgsmål (Cuthbert & Insel, 2013).

Produktet er en matrice hvor en række konstrukter eller dimensioner er samlet i nogle overordnede domæner (fx 'positiv valens' eller 'kognition'). Hvert konstrukt er beskrevet på 8 analyseniveauer: gener, molekyler, celler, kredsløb (neurale), fysiologi, adfærd, selv-rapport og paradigmer (National Institute of Mental Health, 2016). Hvert konstrukt er altså defineret på et adfærdsmæssigt og et biologisk niveau. Derudover er de bedst validerede målemetoder i form af selv-rapporteringsundersøgelser og eksperimentelle paradigmer inkluderet (Carcone & Ruocco, 2017; Cuthbert & Insel, 2013)

Projektet blev påbegyndt af en række relevante eksperter som mødtes i fem forskellige workshops og blev enige om, hvilke fundamentale konstrukter, der skulle inkluderes i RDoC (National Institute of Mental Health, 2016). Det er dog vigtigt at pointere at RDoCs dimensioner er konstrukter og ikke fuldendte og fikserede definitioner. Tværtimod lægges der op til at de konstant valideres, revideres eller helt fjernes, i takt med at de udfordres af ny viden. På samme måde hævder RDoC ikke at dække alt; der vil på ethvert

givent tidspunkt kun være inkluderet de konstrukter, der har mødt kriteriet om at være ordentligt valideret og være associeret med et specifikt neuralt kredsløb. Der vil således løbende blive tilføjet eller fjernet konstrukter i takt med at nye konstrukter bliver grundigt valideret og der bliver fundet nye sammenhænge mellem konkret adfærd og neurale kredsløb (Cuthbert & Insel, 2013). Og målet er netop at udvikle et nyt begrebsapparat og det gøres ved hele tiden at arbejde på at opbygge en bedre, mere præcis viden om domænerne og konstrukterne, end den vi har i dag (Carcone & Ruocco, 2017). Det skal også pointeres at målet med RDoC ikke er at erstatte DSM og ICD som diagnosemanual i klinisk praksis. Det er derimod et forskningsinitiativ, og har til hensigt at integrere koncepter og resultater fra psykologien, neurovidenskaben og genetikken (Yee, Javitt & Miller, 2015).

Dette speciale er i høj grad formet af RDoCs principper. Blandt andet fordi det ikke omhandler diagnoser fra DSM og ICD. Ikke dermed sagt at klassificeringer fra de systemer ikke nævnes eller indgår i opgaven. Men omdrejningspunktet er et enkelt konstrukt, der allerede indgår i RDoC, nemlig *vaner* under domænet *positiv valens*. Der findes imidlertid teorier om vaners rolle i konkrete diagnoser. Den teori vil jeg naturligvis belyse. Men hovedfokus i denne opgave er ikke på sammenhængen mellem kategoriske kliniske diagnoser og vane-konstruktet. I RDoCs ånd lægges de klassiske diagnoser her på hylden. I stedet er fokus på vaner i sig selv, som dimensionelt konstrukt. Yderligere er fokus indskrænket til paradigme-niveauet. Jeg forholder mig i opgaven kritisk til validiteten af et eksperimentelt paradigme, som mål for vaner. Paradigmet indgår i vane-konstruktet i RDoCs matrice. Validiteten af en test som mål for vaner afhænger af definitionen af vaner, og jeg vil i den første del af opgaven belyse, hvordan vaner defineres i litteraturen. På nogle områder er der stor konsensus i litteraturen om, hvordan vaner skal defineres og operationaliseres psykologisk. Det har faciliteret store fremskridt i viden om, hvilke neurale mekanismer, der hænger sammen med vaner. Men i nyere tid har machine learning inspireret nye måder at



operationalisere vaner, som bygger på en sidestilling af vanekonstruktet med et andet konstrukt, 'model-fri indlæring'. Denne nye operationalisering har vundet indpas i litteraturen, men sidestillingen af vaner med 'model-fri indlæring' er omdiskuteret. Jeg vil argumentere for at model-fri indlæring ikke kan sidestilles med vaner, men at vaneforskningen bør være opmærksom på konstruktet som forstyrrende variabel. Det lægger op til den sidste del af opgaven. Her diskuterer jeg validiteten af slips-of-action-testen, som for nuværende indgår i RDoCs matrice. Diskussionen er med særligt fokus på model-fri læring som mulig confound i testen. Til sidst kommer jeg med nogle overvejelser om, hvordan og om, en mere valid slips-of-action-test kan bygges op. Specialet skal ses som startskuddet til en raffinering af vanekonstruktet. På trods af at positiv valens domænet er et af de mest undersøgte i RDoCs matrice, havde der ifølge et systematisk review lavet i 2017 ikke været lavet et eneste studie, der undersøgte vanekonstruktet (Carcone & Ruocco, 2017). Der er, så vidt jeg ved heller ikke udført studier om vaner i RDoC-regi siden. Særligt taget i betragtning at vanedomænet er hovedstenen i flere nye teorier om psykisk lidelse (Gillan et al., 2011; Everitt & Robbins, 2006), er det nu på tide at domænet kigges efter i sømmene.

### **Vaner og psykisk lidelse**

Kompulsivitet refereres til som tendens til repetitive, habituelle handlinger, som gentages på trods af ugunstige konsekvenser (Robbins et al., 2012; Grant et al., 2016), eller gentagelse af handlinger, som ikke producerer værdifulde udfald (Gillan, Robbins, Sahakian, van den Heuvel, & van Wingen, 2016). Kompulsivitet har fået meget opmærksomhed både inden for psykologien, psykiatrien og neurovidenskaben (Gillan & Robbins, 2014) og det er blevet foreslået at kompulsivitet danner grundlag for en lang række psykiatriske diagnoser, herunder stof-afhængighed (Robbins & Everitt, 1999), spiseforstyrrelser (Smith & Robbins, 2013) og *obsessiv-kompulsiv lidelse* (OCD, af engelsk 'obsessive compulsive disorder')

(Fineberg et al., 2010). Disse lidelser er forskellige i deres fænotypiske udtryk (Fineberg et al., 2016). De ser imidlertid ud til at ligne hinanden med hensyn til neurokognitive og neurale deficits (Fineberg et al., 2016; Figeo et al., 2016). Kompulsivitet er blevet associeret med afficeringer i striatum og prefrontal cortex, samt en generel frontostriatal dysregulering i både OCD, stof- samt adfærdsafhængighed (Figeo et al., 2016). Neurokognitivt findes forskellige teorier om, hvilke deficitter, der fører til kompulsiv adfærd (Fineberg et al., 2014) da kompulsivitet både er associeret med en svækket processering af belønning og straf, mangel på kognitiv og adfærdsmæssig fleksibilitet og tendens til vanedannelse og vanerespons. Formentligt er der tale om at et komplekst samspil mellem forskellige deficitter ligger til grund for kompulsiv adfærd. Dette er foreneligt med at kompulsivitet ikke er inkluderet som et konstrukt i RDoCs matrix. I stedet passer de ovenstående neurokognitive endofænotyper på forskellige af RDoCs dimensioner, som for eksempel *kognitiv kontrol* under det kognitive domæne og *belønningsprocessering* og *vaner* under positiv valens-domænet (Figeo, 2016). Kompulsivitet kan altså forstås som et produkt af deficits i flere forskellige konstrukter, som alle er associeret med frontostriatal dysfunktion i forskellige afskygninger.

Der er fortsat behov for en bedre forståelse og definition af kompulsivitet og hvordan det bedst måles. Det indbefatter en bedre forståelse af de neurokognitive, neurale og neurokemiske mekanismer bag kompulsivitet, herunder overgangen fra det funktionelle til det dysfunktionelle (Grant et al., 2016). En bedre forståelse for kompulsivitet som multidimensionelt og transdiagnostisk koncept har potentiale til at bane vejen for nye behandlingsmetoder, der er effektive på tværs af psykiatriske diagnoser (Grant et al., 2016).

En måde at opnå bedre forståelse af kompulsivitet som konstrukt er ved at opbygge mere præcis viden om de konstrukter der forbindes med kompulsivitet. Der er fortsat et stort behov for bedre viden omkring relationen mellem vaner og kompulsivitet (Grant et al., 2016). Der er for eksempel flere neurokognitive ruter til udviklingen,

udførelsen, og hæmningen af vaner og en forståelse af disse forskellige processer kan bidrage til forståelsen af vaners involvering i kompulsivitet på tværs af lidelser (Gillan et al., 2016). Vane-hypotesen om kompulsivitet har banet vejen for nye, effektive behandlingsmetoder inden for kompulsive lidelser. Et eksempel er Habit Reversal Therapy hvor fokus rettes mod selve kompulsionerne idet patienten træner at være opmærksom på de stimuli, der udløser den kompulsive adfærd, for derefter at igangsætte en anden, konkurrerende adfærd, som skal hæmme den uønskede adfærd (Grant et al., 2012). Denne terapiform har vist sig effektiv overfor kompulsive lidelser som *trikotillomani* (kompulsiv trang til at plukke sine hår ud) (Rothbart & Stein, 2014; McGuire et al., 2014) og *Tourettes syndrom* (en lidelse præget af tics i form af ufrivillige bevægelser og lyde) (Deckersbach, 2014). Et styrket viden om sammenhængen mellem vaner og kompulsivitet har formentligt potentiale til at bane vejen for flere nye behandlingsformer. Til sidst ændrer vaneteorien opfattelsen af ætiologien bag konkrete lidelser, som OCD (Gillan & Robbins, 2014) og afhængighed (Everitt & Robbins, 2005) markant, hvorfor den synes vigtig at undersøge.

Teorien om at vaner ligger til grund for kompulsiv adfærd, blev første gang foreslået, da Graybiel og Rauch (2000) argumenterede for at OCD kan forstås som dysfunktionel vanedannelse. Der er to overordnede teorier angående de underliggende mekanismer bag OCD (Gillan & Robbins, 2014). Inden for den traditionelle kognitive skole mener man, at den kompulsive adfærd, der ses ved OCD, er målrettet i den forstand, at den har en funktion for individet, fx i form af at berolige angstprovokerende tanker eller minimere risikoen for en opfattet trussel (Gillan & Robbins, 2014). Det primære problem er den kognitive dysfunktion og adfærden er sekundær. Den anden teori, som Graybiel og Rauch skriver sig ind i, foreslår at OCD ikke opstår på grund af kognitiv dysfunktion, men snarere at OCD-patienterne forstår og oplever den negative konsekvens af deres kompulsioner, men ikke har fuld kontrol over deres handlinger (Gillan & Robbins, 2014).

Inden for denne skole, er kompulsionerne således karakteriserede ved ikke at være målrettede eller meningsfulde. En mulig forklaring på, hvad der forårsager denne manglende kontrol over handling, kunne være en ubalance mellem to associative læringssystemer: vaner og målrettet adfærd. Forslaget understøttes af store overlap i de neurale kredsløb, der er i spil i henholdsvis vanedannelse og OCD (Gillan & Robbins, 2014; Graybiel & Rausch, 2000).

En lignende bevægelse har fundet sted med hensyn til afhængighed. Studier har vist at selvadministrering af stoffer går fra at være målrettet til, efter tilstrækkelig gentagelse, at blive vanestyret (Everitt, Dickinson & Robbins, 2001; Robbins & Everitt, 1999).

### **Operant betingning**

For at kunne forstå den psykologiske teori omkring vaner må man kende til nogle af grundbegreberne inden for operant betingning. De begreber er *forstærkning*, *udslukning* og *diskriminering* (i betydningen 'at skelne'). De er gennem årene blevet så grundigt efterprøvet eksperimentelt at de i dag er mere eller mindre anerkendte som love inden for psykologien (O'Donohue & Fergusson, 2001).

*Operant betingning* handler om samspillet mellem adfærd og miljø. En *operant respons* er en adfærd, der ændrer miljøet (O'Donohue & Fergusson, 2001): jeg skriver på min computer, så nu er der bogstaver på min computerskærm. Jeg tager en jakke på, så nu er rummet et sted, hvor jeg har en jakke på, jeg taler, så nu er der lyd i rummet. Disse miljøændringer er *udfald* af adfærd. Disse udfald kan omvendt påvirke adfærd, ved at øge eller sænke sandsynligheden for at den bliver gentaget og det kaldes *forstærkning* (Adams & Dickinson, 1981). Der er ifølge klassisk betingning 3 typer udfald. *Neutrale udfald* hverken hæver eller sænker hyppigheden af adfærd i fremtiden. *Forstærkere* øger hyppigheden af handlingen i fremtiden. *Straf* sænker sandsynligheden. (O'Donohue &

Fergusson, 2001). Hvis jeg for eksempel kommer for sent til en aftale, med det udfald at min veninde bliver meget sur på mig (straf) vil det ifølge operant betingning få mig til at komme mindre for sent for fremtiden. Jeg vil ikke risikere at min veninde bliver sur igen. Hvis jeg arbejder meget hårdt på mit arbejde med det udfald at jeg får en lønforhøjelse for min flid (forstærkning), vil jeg være mere tilbøjelig til at fortsætte med at arbejde meget hårdt. Hvis en forstærkning ophører og den respons der er blevet forstærket efterfølgende også ophører kaldes det *udslukning* (Blackman, 2019; O'Donohue & Fergusson, 2001). Det kunne for eksempel være at jeg i længden holdt op med at få lønforhøjelser og med tiden holdt op med at arbejde så hårdt. En respons forholder sig altid til et miljø. Jeg skriver kun på min computer, hvis den står foran mig og er tændt, jeg tager min jakke på, hvis det er køligt i vejret og jeg taler fordi der er nogen at tale til eller fordi nogen stiller mig et spørgsmål. Disse faktorer i miljøet, som giver anledning til en respons udgør det man inden for betingning kalder *diskriminative stimuli* (O'Donohue & Fergusson, 2001). 'Diskriminativ' står i modsætning til 'genereliserende' og henviser til at to stimuli skelnes fra hinanden. *Diskriminativ indlæring* er for eksempel når et barn lærer at det ikke er alle mænd, men kun én bestemt mand, der skal kaldes 'far' eller hvis jeg lærer at en kontakt, men ikke en anden tænder lyset i køkkenet.

Ved operant betingning er der altså tale om at en respons i sammenhæng med en stimulus udløser et udfald. Det udfald enten sænker eller øger hyppigheden af den respons i sammenhæng med den stimulus. Den association mellem en stimulus, en respons og et udfald stilles op således:  $S \rightarrow R \rightarrow U$  (Eng.  $S \rightarrow R \rightarrow O$  eller  $S \rightarrow A \rightarrow O$ ) (Blackman, 2019).

## **Vaner defineret**

**Vaner er S-R-associationer**

Der er generel konsensus om at instrumentel adfærd kontrolleres af to forskellige systemer; et vane-system og et målrettet system (Yin & Knowlton, 2006). Derfor handler definitionen af vaner i høj grad om at afgrænse dem fra målrettet handling. Målrettede handlinger er karakteriseret ved at være intentionelle. Det betyder at de er baseret på et *ønske* eller *mål* og en *formodning*. Et ønske om at opnå noget og en antagelse om vejen til at opnå det. For at adfærd kan betragtes som målrettet skal det altså leve op til kriterierne om at bygge på et *mål* og en *formodning* (Dickinson & Pérez, 2018). Målrettet læring er dermed styret af forventningen om et udfald. I kontrast til målrettet adfærd udføres vaner uafhængigt af noget mål. I stedet er vaner styret af simple associationer mellem kontekst og adfærd, eller stimulus og respons (Yin & Knowlton, 2006; Gillan et al., 2016; Balleine & O'Doherty, 2009; Dolan & Dayan, 2013; Dickinson & Pérez, 2018). Hvis du fx kører fra arbejde hjem til en ven du skal besøge, er det en målrettet handling fordi det bygger på et mål om at nå frem på din vens adresse, og en tro på at den rute du kører, vil få dig frem på adressen. Det kan også være at du fejlagtigt kommer til at dreje til højre i et kryds, fordi det er vejen hjem til dig selv og du plejer at dreje til højre der. Det er ikke målrettet; du har ikke et ønske om at komme hjem og du tror ikke at dit højresving vil bringe dig frem til din vens adresse. Handlingen er en vane. Den bliver udløst af konteksten, at du lige har fået fri og ankommer til netop det kryds. Dickinson bruger et lignende eksempel: en person, går ind på sit soveværelse for at skifte til middagstøj, men kommer i stedet til at skifte til nattøj. Det at skifte til nattøj om aftenen har mange gange været en handling, der har været passende i forhold til et mål om at komme i seng. Men målet er nu at blive klar til middag og derfor upassende at skifte til nattøj. Så i det tilfælde er det ikke en målrettet handling, men en vanestyret respons på at være på sit soveværelse om aftenen (Dickinson, 1985).

Målrettet handling og læring er altså styret af associationer mellem en handling (action) og et udfald (outcome). Hvis jeg gør A så opnår jeg B. Derfor beskriver man den som

A-O-adfærd (Yin & Knowlton, 2006). Vaner bliver derimod igangsat af en forudgående stimulus: en bestemt kontekst, som at være i et bestemt lyskryds, igangsætter en bestemt adfærd, som at køre til højre. Vaner bygger altså på simple associationer mellem en kontekst (eller stimulus) og en adfærd (eller respons) og bliver betegnet stimulus-respons-adfærd (S-R) (Yin & Knowlton, 2006). Man kan sige det sådan at målrettet adfærd er handlinger, mens vaner er simple responser på stimuli (Dickinson, 1985). En stimulus i denne sammenhæng kan godt være kompleks fx i form af en kontekst (Dickinson & Pérez, 2018).

### **Operationalisering af vaner**

Målrettet handling og vaner er altså forskellige idet vaner er uafhængige af deres udfald, mens målrettet handling netop er styret af forventninger om et bestemt udfald (Dickinson, 1985, Yin & Knowlton, 2006; Gillan et al., 2016; Balleine & O'Doherty, 2009; Dolan & Dayan, 2013). Denne forskel gør at de to også adskiller sig fra hinanden i, hvordan de reagerer på ændringer i mål eller ønske, eller i konsekvenser af handlingen (Dickinson, 1985; Gillan et al., 2016; Balleine & O'Doherty, 2009; Dolan & Dayan, 2013). Altså skift i miljøets eller situationens A-O-forhold. Det kan for eksempel være hvis jeg har en stikkontakt ved indgangen til stuen, som plejer at tænde loftslýset. Normalt vil jeg trykke på stikkontakten når jeg går ind ad døren, for at tænde lyset. Men nu træder jeg ind i rummet og ved at pæren i lampen er sprunget. En målrettet handling vil være at lade være med at trykke på stikkontakten, men i stedet tænde et stearinlys. Det er målrettet fordi jeg har et ønske om lys i stuen, og ved at jeg kan opnå det med et stearinlys, men ikke fra loftslampen. Der er altså sket en ændring i A-O-forholdene. Før ville et tryk på kontakten (handlingen/A) føre til lys i stuen (udfaldet/O). Men fordi pæren er sprunget, giver den samme handling ikke længere samme udfald, og for at opnå samme mål, må jeg tilpasse min adfærd. Det kunne imidlertid også være at jeg trådte ind i lokalet og, af ren vane, trykkede på stikkontakten, selv

om jeg vidste at den ikke virkede. At trykke på kontakten er en simpel S-R-respons, en reaktion på at træde ind i min mørklagte stue og afhænger ikke af noget mål om lys.

Adfærden tilpasses derfor ikke ændringerne i A-O-forholdene. Måltrettet handling er altså fleksibel og sensitiv over for ændringer i A-O-forhold, mens vaner er kendetegnet ved manglende sensitivitet og fleksibilitet overfor ændringer i konsekvenserne af handlinger.

Denne karakteristika ved vaner udnytter man til eksperimentelt at skelne vaner fra måltrettet adfærd. Eksperimentelt er vaner defineret ved at være upåvirkede af ændringer i mål og A-O-associationer (Graybiel, 2008; Dickinson, 1985, Gillan et al., 2016; Dolan and Dayan, 2013; Daw, Niv & Dayan, 2005). Omvendt, for at være måltrettet må en handling være følsom overfor ændringer i værdien af udfaldet (ændringer i mål og ønske) og samtidig sensitiv overfor ændringer i den kausale sammenhæng mellem handling og udfald (ændringer i antagelse) (Yin & Knowlton, 2006; (De fleste tests, der måler vaner, manipulerer altså variablerne *mål* og *udfald* (Gillan et al., 2016). Overordnet findes to typer:

*udfaldsdevaluering-tests* og *contingency degradation-tests* (Gillan et al., 2016).

I udfaldsdevalueringstests undersøger man sensitiviteten i adfærden over for motivationelle ændringer, altså ændringer i målet. I et typisk udfaldsdevaluering forsøg foretages en ændring i værdien af en *forstærker* (oftes sænkes værdien, men i sjældne tilfælde forøges den). Dette kan ske enten gennem en intern ændring af værdien, fx ved at fremprovokere overmætning af en bestemt madvare hos individet. Alternativt kan værdien af forstærkeren justeres gennem ændringer i dens eksterne værdi (Gillan et al., 2016). Hvis adfærden er følsom overfor disse ændringer, vil den anses for at være måltrettet, og hvis den ikke er, ses den som habituel (Gillan et al., 2016; Dolan & Dayan, 2013; Dickinson, 1985). En almindelig type udfaldsdevalueringstest er reversal learning-opgaver, der benytter sig af ekstern udfaldsdevaluering. I disse opgaver indlæres en association mellem to stimuli og deres associerede udfald, gennem at deltageren skal reagere på en af to stimuli for at opnå en



belønning, undgå straf eller begge dele. Dette efterfølges af en reversal-del, hvor disse associationer vendes om. Tendensen til perseveration, altså at adfærden ikke ændres som reaktion på de ændrede forhold, er den afhængige variabel (Fineberg, 2014).

Den anden type test kaldes contingency degradation. I stedet for at ændre i værdien af udfaldet, skrues man i denne type test på sammenhængen mellem handling og udfald, således at en given handling ikke længere er nødvendig for et givent udfald (Gillan et al., 2016), og handlingen altså ikke længere belønnes. Adfærden anses for at være målrettet, hvis deltageren stopper den ikke længere belønnede handling (Gillan et al., 2016).

### **Vaner skabes gennem gentagelse af målrettede handlinger**

Der er generel konsensus om at vaner udvikles gennem målrettet handling (Dickinson, 1985). Det vil sige at en målrettet handling, som at trykke på kontakten, når man vil have lys i stuen, udvikler sig til en vane, når den er blevet gentaget nok gange.

Vaner opstår således af handlinger, drevet af ønsket om at opnå et bestemt mål, som efter en mængde gentagelser bliver responser, der nu er uafhængige af målet, og i stedet udløses af de stimuli, som handlingen er blevet udført i sammenhæng med (Dickinson, 1985). Det tidligere eksempel fra soveværelset illustrerer dette fint. Det at skifte til nattøj som en respons til at være på sit værelse sent om aftenen, er blevet en vane, fordi det har været gjort adskillige gange, som led i en plan om at opnå et bestemt mål (at blive klar til at komme i seng).

Overgangen fra handling til respons, eller målrettet handling til vane, blev undersøgt af Adams (1982) i et eksperiment, hvor to grupper af rotter blev trænet i at trykke på et håndtag/knap for at opnå en belønning i form af sukkertabletter. Grupperne adskilte sig fra hinanden i mængden af træning, således at den ene gruppe fik 100 belønnede tryk på

håndtaget/knappen, mens den anden gruppe fik 500 tryk. Forventningen var, at det at trykke på håndtaget, ville være blevet en vane for den gruppe med øget træning, og dermed insensitiv til udfald-devaluering, mens det for den mindre trænede gruppe stadig ville være en handling, altså fleksibelt i forhold til ændringer i mål og målets værdi. I halvdelen af de to grupper blev sukker-belønningen parret med indsprøjtning af lithiumklorid, som fik rotterne til at få det dårligt. Dermed blev belønningen devalueret. Den anden halvdel fungerede som kontrolgruppe, og de fik lithiumkloridindsprøjtningerne uafhængigt af sukker-belønningen. I gruppen med mindre træning reagerede rotterne på devalueringen, således at de, der havde fået lithium-indsprøjtningen sammen med sukker-pillerne trykkede mindre på håndtaget, end de, der havde fået lithium-indsprøjtningen uafhængigt af sukker-pillerne. Den samme effekt af devaluering sås ikke hos gruppen med en høj mængde træning. Resultaterne indikerer at ekstensiv træning af en handling kan føre til, at den bliver måluafhængig, og dermed går fra at være en handling, til at være en vane eller respons (Adams, 1982).

Det er lidt omdiskuteret, hvilken rolle udfald spiller i dannelsen af vaner. Traditionelt har man henvist til Thorndikes *effektlov*, ifølge hvilken udfaldet af en handling forstærker associationen mellem handlingen og stimuluskonteksten, men uden at udfaldet bliver indkodet (Thorndike, 1911). Andre mener at blot det, at der er en sammenhæng mellem handling og stimuluskontekst er nok til at forstærke associationen (Guthrie, 1959). Altså hvis en bestemt handling gentages mange gange i den samme kontekst. Dette vil oftest være tilfældet hvis handlingen lykkes med at opnå et bestemt mål. Fremfor at forholde sig til et fremtidigt mål relaterer vaner sig altså til et tidligere mål. De udgør med andre ord den adfærd, der tidligere var effektiv til at opnå et tidligere mål (Graybiel, 2008).

### **Automaticitet**

Somme tider bliver det fremhævet at vaner er *automatiske*. I RDoC formuleres det som at når en vane først er sat i gang kan den fuldføres uden ”*constant conscious*

*oversight*” (National Institute of Mental Health, 2016). Klassiske vanestudier har også defineret vaner som produkt af *procedural* indlæring, i kontrast til *deklarativ* indlæring (Graybiel, 2008). Procedural indlæring er netop kendetegnet ved at være ubevidst og implicit i modsætning til deklarativ indlæring, som er eksplicit og bevidst (Gade, 2021). Med ’automatisk’ menes der altså at de er ubevidste, implicitte processer, som kan gennemføres mens opmærksomheden er optaget andetsteds.

S-R-associationer er imidlertid også karakteriseret ved at være non-deklarative, altså implicitte og ubevidste (Gade, 2021). De kan tilmed karakteriseres som automatiske i den forstand at de er automatiske associationer mellem stimulus og respons (Gade, 2021).

Hvis S-R-associationer er implicitte, ubevidste og automatiske er det svært at se hvad begrebet *automatisk* egentlig tilføjer til definitionen af vaner. Snarere er det en givet følge af at vaner er S-R-associationer. En måde at måle om adfærd kræver kognitive ressourcer er ved at tilføje en ekstra opgave, der trækker på de kognitive ressourcer, oven i den adfærd, der skal undersøges. Hvis adfærden eller opgaven kan udføres samtidig med den ekstra kognitive belastning, betragtes adfærden som implicit og ubevidst (Economides et al., 2015). Eksperimentelt er der altså forskel på operationaliseringen af vaner som S-R-associationer og vaner som automatisk adfærd, der ikke kræver kognitive ressourcer. Det kunne tale for at automaticitet alligevel har sin berettigelse i vane-definitionen. Langt de fleste tests der er bygget til at måle vaner, operationaliserer dog vaner som ufølsomhed overfor ændringer i mål og udfald (Gillan et al., 2011; DeWit et al., 2014; D’Cruz et al., 2014; Swainson et al., 2000). Eksperimentelt lægges der altså større vægt på vaner som S-R-associationer, end på at vaner er automatiske og kræver få kognitive ressourcer.

## **Delkonklusion**

Jeg har i ovenstående afsnit forsøgt at give en indføring i, hvordan vaner defineres. Jeg har fremhævet flere forskellige karakteristika ved vaner. Vaner er automatiske og sparer kognitiv energi. Vaner udvikles gennem gentagelse. Vaner er S-R-associationer og i modsætning til målrettet læring er de altså ikke følsomme over for A-O-sammenhænge. Derfor er de heller ikke følsomme overfor ændringer i mål og ønsker eller sammenhænge mellem en handling og dets udfald. Det udnytter man eksperimentelt når man operationaliserer vaner som ufølsomhed overfor ændringer i mål eller i handlingers udfald. Alle disse karakteristika er indkapslet i RDoCs definition af vaner (National Institute of Mental Health, 2016).

### **Neurobiologien bag vaner**

Forskning peger på at overgangen fra målrettet kontrol til vanekontrol medieres af et skift mellem parallelle kortikostriatale kredsløb (Yin & Knowlton, 2006). *Striatum*, som er en del af basalganglierne, modtager input fra stort set alle områder i hjernebarken og fra thalamus. Striatum fungerer som moderator ved enten at hæmme eller forstærke de indkommende projiceringer gennem en excitatorisk og en inhibitorisk strøm (Albin, Young & Penney, 1989; Wickens & Koetter, 1995; Yin & Knowlton, 2006). Basalganglierne og særligt striatum spiller således en vigtig rolle i form af styring af adfærd.

Der er identificeret 5 dissociative 'loops', der forbinder separate områder i pandelappen med separate områder i basalganglierne. Hvert loop løber gennem striatum, globus pallidus, substantia nigra, og hjernebarken, i hver deres ikke-overlappende rute. Hvert loop integrerer input fra flere områder i hjernebarken, som projiceres tilbage til et enkelt område i hjernebarken (Alexander, DeLong & Strick, 1986). De enkelte kredsløb modtager information fra områder i hjernebarken, som er indbyrdes forbundne og funktionelt relaterede (Alexander, DeLong & Strick, 1986; Cheatwood, Reep, & Corwin, 2003). Det oculomotoriske

loop modtager eksempelvis input fra tre forbundne områder, der alle er involveret i kontrol over øjenbevægelser (Alexander, DeLong & Strick, 1986). Særligt det sansemotoriske og det associative loop forbindes med balancen mellem vaner og målrettet læring (Yin & Knowlton, 2006). Det associative loop består blandt andet af dorsolateral prefrontal cortex (dPFC) og *nucleus caudatus* (en del af *striatum*) (Yin & Knowlton, 2006; Joel & Weiner, 2000). Begge strukturer er involveret i forudsigelse af- og forventning om belønning (Leon & Shadlen, 1999; Hassani, Cromwell & Schultz, 2001). Det sansemotoriske loop rummer, ikke overraskende, sansemotoriske områder i hjernebarken, som projicerer til putamen i en somatotopisk organisering, sådan at motoriske og sensoriske områder ansvarlige for samme legemsdel integreres i samme område i putamen (Kunzle 1975; Kunzle 1977; Joel & Weiner, 2000). Kredsløbet ender i motoriske områder i hjernebarken og hjernestammen (Yin & Knowlton, 2006). Det sansemotoriske loop er relateret til bevægelse og processeringen af diskriminative stimuli (Kanazawa, Murata & Kimura, 1993; Kimura, Aosaki, & Ishida, 1993) altså de stimuli, der er involveret i forstærkningen af en adfærd.

Studier har for eksempel vist at rotter med læsioner på *dorsolateral striatum* (DLS), som svarer til putamen hos primater, handler målrettet på opgaver, hvor de ikke-skadede kontrolrotter var styret af vaner (Yin, Knowlton & Ballaine, 2004; Yin, Knowlton & Ballaine, 2006). Det indikerer at putamen er nødvendig for udviklingen af vaner.

Andre rottestudier har omvendt vist at inaktivering af dorsomedial striatum (DMS), som svarer til *caudatus* hos mennesker, resulterer i hæmmet A-O-læring, sammenlignet med kontroller med intakte DLS (Yin & Knowlton, 2004; Yin, Knowlton & Ballaine, 2005; Yin, Ostlund & Balleine, 2005). Derudover har man fundet at aktivitet i *caudatus* i mennesker hænger sammen med opfattelsen af sammenhæng mellem handling og udfald, sådan at *caudatus* kun blev aktiveret når deltagerne troede at udfaldet hang sammen med deres handlinger (Tricomi, Delgado & Fiez, 2004). Disse resultater indikerer at *caudatus* spiller en

afgørende rolle for målrettet læring. Lignende studier peger på at PFC også er afgørende for målrettet læring og kontrol (Balleine & Dickinson, 1998; Leon & Shadlen, 1999; Tsujimoto & Sawaguchi, 2004). Vanedannelse og vanekontrol medieres altså af det sansemotoriske kortikostriatale loop, som udgøres blandt andet af motoriske områder i hjernebarken og putamen, mens målrettet læring og kontrol medieres af det associative loop, der består blandt andet af præfrontal cortex og caudatus (Yin & Knowlton, 2006).

De forskellige loops er parallelle, men anatomisk forbundne blandt andet via substantia nigra (Haber, Fudge & McFarland, 2000) og aktivitet i ét loop kan brede sig til det næste og så videre (Haber, 2003). Man har fundet at indlæringen af nye responser hos mennesker aktiverer områder i det associative loop, mens aktiviteten skiftede til det sansemotoriske loop når sekvenserne var gennemøvede (Jueptner et al., 1997; Jueptner, Frith & Brooks, 1997). Det indikerer at det associative loop i overgangen fra målrettet læring til vaner overdrager kontrollen til det sansemotoriske loop (Yin & Knowlton, 2006).

### **Modelfri og modelbaseret læring**

Maskinlæring (eng. Machine learning) er computermødelles som bruges til forudsigelse og genkendelse ved hjælp af sandsynlighedsteori og statistik. Forstærkningsindlæring (eng. Reinforcement learning) er en form for maskinlæring, der adskiller sig fra andre modeller ved at være målrettet og interagere med miljøet (Bertsekas & Tsitsiklis 1996; Gershman & Daw, 2017; Kaebing, Littman & Moore, 1996; Sutton & Barto, 1998). At den er målrettet vil sige at den forsøger at maksimere belønning. At den interagerer med miljøet betyder at den lærer gennem egen erfaring med miljøet (Gershman & Daw, 2017; Kaebing, Littman & Moore, 1996 Sutton & Barto, 1998). Dermed er det også den form for maskinlæring, der bedst ligner indlæring hos mennesker og dyr og mange af algoritmerne i forstærkningsindlæring er inspireret af biologiske læringssystemer, som klassisk og operant

betingning. Omvendt har forstærkningsindlæring også bidraget til opbygningen af modeller og eksperimenter om mennesker og dyrs læring (Sutton & Barto, 1998). Et eksempel på sådanne algoritmer fra forstærkningsindlæring der har fået betydning for studiet af vaner og målrettet adfærd, er dem, der skelner mellem modelbaseret og modelfri læring (Sutton & Barto, 1998; Dolan and Dayan, 2013; Daw et al., 2005; Otto et al., 2015). Koncepterne modelbaseret og modelfri læring er som sagt oprindeligt opstået ud af maskinlæring, men har senere bevæget sig ind på psykologiens domæne (Miller et al., 2018), og det er det jeg fokuserer på her.

Distinktionen mellem modelbaseret og modelfri indlæringsstrategier beror på, hvorvidt agenten gør brug af en indre model af opgavens opbygning (Sutton & Barto, 1998). Et hverdageksempel kunne være at jeg kender vejen fra der hvor jeg bor, hjem til min veninde. Den vej kan jeg have lært enten modelbaseret eller modelfrit. Jeg kan fx have lært den modelbaseret ved at jeg har boet i byen længe og nemt kan finde rundt. Jeg har altså en indre model over vejnettet og vejen hjem til min veninde har jeg ”lært” ved at få hendes adresse at vide og selv tænke mig frem til den hurtigste vej derhen. Hvis jeg får at vide at der er kø eller afspærret på en af de veje jeg ville bruge kan jeg også tænke mig frem til en alternativ rute. Og hvis jeg ændrer mening og beslutter mig for at hente vores tredje veninde først kan jeg også hurtigt ændre ruten. Jeg kunne også have lært ruten modelfrit. Lad os sige at jeg lige er flyttet til byen, og generelt ikke kan finde rundt. Jeg kender vejen til min veninde fordi jeg har fulgtes med andre derhen en del gange. Så jeg ved af erfaring at den rute fører hjem til min veninde. Jeg har indlært en sekvens af handlinger (ud ad hoveddøren og til venstre. Til højre ved tredje sidegade. Til venstre ved supermarkedet, osv.). Men min viden bygger ikke på nogen overordnet viden om vejnettet og hvordan de forskellige gader hænger sammen så hvis jeg får at vide at et sted på min rute er spærret, kan jeg ikke bruge informationen til at lave en ny rute. Og hvis jeg får at vide at arrangementet er blevet rykket hjem til min anden veninde, kan jeg heller ikke bare ændre ruten.

'Model' henviser her til sammenhængen mellem opgavens eller miljøets forskellige *stadier* (punkter på ruten), agentens *handlinger* (drej til højre, venstre eller gå ligeud) og *udfaldene* af disse (det næste punkt på ruten/det næste sted jeg når frem til) (Kaelbling, Littman & Moore, 1996)

Modelbaserede strategier er prospektive processer, og benytter indre repræsentationer af disse sammenhænge, til at forudsige, hvilken handling, eller transition, der fører til det mest værdifulde udfald i et givent stadie. De har med andre ord en indre model over miljøet (Sutton & Barto, 1998; Dolan & Dayan, 2013; Miller et al., 2018). Modelfri strategier er kendetegnet ved ikke at gøre brug af modeller for opgavens forskellige transitioner og sammenhængen mellem de forskellige stadier (Sutton & Barto, 1998). Modelfri læring er i stedet en retrospektiv proces (Dolan & Dayan, 2013) hvor adfærden baseres på de forstærkninger, der tidligere er forbundet med en given adfærd, altså på direkte erfaring med miljøet (Miller et al., 2018).

De to koncepter bliver typisk målt i opgaver, hvor en agent skal vælge mellem handlinger i flere (oftest to) konsekutive stadier. Det kunne for eksempel være en labyrint, hvor en rotte i første stadie skal vælge mellem to gange. Hver gang fører til et nyt rum, andet stadie, med to nye valgmuligheder, hver med en belønning af forskellig værdi, for eksempel værdierne 0, 2, 3 og 4. Hvis rotten vælger at gå til højre på første stadie, fører det til et andet stadie, hvor rotten skal vælge mellem to gange, den ene med værdien 2 og den anden med værdien 3. Hvis rotten vælger at gå til venstre i første stadie, vil mulighederne i andet stadie have værdierne 0 og 4. For at opnå den maksimale værdi skal rotten træffe de beslutninger på stadie 1 og 2, som i sidste ende fører til belønningen med værdien 4 (Sutton & Barto, 1998). Det valg, der træffes i det første stadie, bestemmer transitionen til det næste stadie, hvor der på ny skal træffes en beslutning. Valget på det sidste stadie bestemmer det endelige udfald. En modelfri strategi kan være indlæringen af *handlingsværdier*, som henviser til den højest mulige værdi en handling i



et bestemt stadie kan føre til, i enden af opgaven. I ovenstående eksempel vil handlingsværdien af at gå til venstre i stadie 1 være 4 mens handlingsværdien af at gå til højre vil være 3. Der er altså tale om associationer mellem enkelte stadier og handlinger, eller stimuli og responser og disse associationer indlæres gennem gentagne forsøg med den samme opgave, indtil de til sidst er blevet så pålidelige, at de kan bruges til at træffe beslutninger med maksimal værdi. Modelbaseret viden beror, fremfor simple stimulus-respons-associationer på viden om sammenhænge mellem et stadie, en handling, det deraf følgende stadie osv., altså på stimulus-respons-udfald-associationer (Sutton & Barto, 1998).

Hvis labyrinten ændres sådan at rummet med en værdi af 4 nu har en værdi af 0, vil rotten reagere forskelligt på denne ændring afhængigt af, om den benytter en modelfri eller en modelbaseret indlæringsstrategi. Hvis rotten placeres i det rum, der tidligere havde en belønning med værdien fire, men nu oplever at belønningen er blevet ændret til 0, vil den, hvis den benytter en modelbaseret strategi opdatere hele modellen. Det betyder at den, uden at have direkte erfaring fra resten af labyrinten ved, at der ikke er nogen belønning at hente ved at gå til venstre i stadie 1. I en modelfri strategi kræver en sådan opdatering af adfærden, at de nye forhold indlæres gennem gentagne og direkte erfaringer med miljøet, hvor ændringerne opleves af aktøren (Sutton & Barto, 1998). De to strategiers forskellige forhold til miljøet medfører således at de reagerer forskelligt på ændringer i miljøet (Sutton & Barto, 1998). Ved ændringer i miljøets værdiforhold vil en model blive opdateret og have direkte og øjeblikkelig indvirkning på adfærden, mens en modelfri strategi først vil føre til en adfærdsændring efter at aktøren har erfaret de nye forhold direkte gennem interaktion med miljøet (Sutton & Barto, 1998; Doal & Dayan, 2013). Lad os tage endnu et hverdageksempel. Jeg kender en opskrift på nogle rigtig gode boller som jeg har lært modelfrit. Jeg har prøvet mig lidt tilfældigt frem gennem noget tid. Tilføjet lidt mere hvedemel, lidt mindre rugmel, mere gær og skruet op og ned for varmen i ovnen, indtil de en dag var perfekte. Jeg ved, fordi jeg har erfaret det, at præcis denne

fremgangsmåde resulterer i gode morgenboller, men jeg ved ikke hvorfor. Jeg ved ikke om valget af mel har en betydning for hvor varm ovnen skal være. En dag sker der en ændring i miljøet; en af de meltyper jeg bruger kan ikke længere købes i supermarkederne. For at finde en ny god opskrift må jeg igennem hele processen igen med at prøve mig lidt tilfældigt frem, tilføje lidt af det ene og lidt af det andet og skrue op og ned for varmen i ovnen. Jeg skal med andre ord gøre min nye erfaringer med miljøet og den proces er tidskrævende. Dette er karakteriserende for modelfri læring: efter ændringer i miljøet vil den modelfri strategi først blive opdateret efter at jeg har opbygget nye erfaringer gennem interaktion med miljøet (prøvet mig fremmed ingredienser) (Sutton & Barton, 1998). Jeg kunne også have lært opskriften modelbaseret ved at jeg havde en masse viden om bagning og samspillet mellem forskellige ingredienser og fremgangsmåden. Hvis miljøet eller målet nu ændrer sig, ved at jeg for eksempel finder ud af at jeg er allergisk overfor hvedemel kan jeg hurtigt bruge min viden om ingrediensernes samspil (min indre model) til at finde på en ny opskrift, som formentligt vil være god. Ved ændringer i miljøet kan jeg opdatere min model og ændre adfærd uden at have interageret med miljøet og gjort mig nye erfaringer først.

Den mest brugte test til at måle balancen mellem modelbaseret og modelfri læring hos mennesker er Daw's *twostep sequential decision making task* (Daws two-step) (Daw et al., 2011). Testen er bygget op af to stadier. I det første stadium skal deltageren vælge mellem to stimuli. Afhængig af valget, sker en transition videre til et af to andet-stadier. Transitionen er probabilistisk, det vil sige, at hvert førstestadie-valg er forbundet med høj sandsynlighed for ét andet-stadie og lav sandsynlighed for et andet andet-stadie. Valgene på andet stadium er forbundet med hver en sandsynlighed for belønning. Disse sandsynligheder ændrer sig igennem eksperimentet. Man er interesseret i, hvordan deltageren reagerer på belønning eller fravær af belønning efter højsandsynligheds-transitioner og lavsandsynligheds-transitioner. Ved modelfri indlæring forventes det, at en belønning efter en lavsandsynlighedstransition

fører til, at der tages det samme førstestadie-valg i den følgende runde. Hvis der ikke opnås belønning, forventes deltageren at skifte første-stadie. Deltageren gentager altså den handling, der har ført til belønning, ligegyldigt om der er høj eller lav sandsynlighed for, at det sker igen. Ved en modelbaseret strategi forventes det, at deltageren tager højde for disse sandsynligheder. Ved belønning efter en lavsandsynlighedstransition vil deltageren skifte valg i første stadie i næste runde (Daw et al., 2011).

### **Modelfri og modelbaseret læring i relation til vaner**

Man har i nyere tid hævdet at model-fri strategier implementeres af vane-systemet, mens model-baserede strategier implementeres af det målrettede system (Daw, Niv & Dayan, 2005; Dolan & Dayan, 2013). Faktisk er begreberne i dag så tæt sammenvævede, at modelbaseret læring bruges mere eller mindre synonymt med målrettet læring, og det samme gælder for model-fri læring og habituel adfærd (Miller et al., 2018). For eksempel har et studie undersøgt tendensen til model-fri og modelbaseret læring hos forskellige kompulsive patientgrupper. Folk med OCD, nervøs tvangsspisning og amfetaminmisbrug gennemførte en Daw two-step opgave og der sås et bias hos disse patientgrupper mod model-fri læring sammenlignet med raske kontrolpersoner. Forskerne konkluderer at have påvist en bias mod vanedannelse hos disse patientgrupper, sammenlignet med kontrolgruppen (Voon et al., 2015). Mange andre artikler konkluderer på samme måde at have målt vaner gennem modelfri indlæring, ved brug af tests som Daw Two-step (Daw, Niv & Dayan., 2005; Daw et al., 2011; Otto et al., 2015). Det er ofte gjort eksplicit at konklusionen bygger på antagelsen om en sidestilling af modelfri indlæring og vaner. Men dette er ikke nødvendigvis tilfældet når andre artikler henviser til dem. Gillan og kolleger (2016) skriver eksempelvis med reference til Voon og kollegers fund, at deficits i målrettet kontrol ses hos psykiatriske diagnoser præget af kompulsivitet som OCD, nervøs tvangsspisning og amfetaminmisbrug (Gillan et al., 2016).

Antagelsen om sidestilling af modelfri og modelbaseret indlæring med henholdsvis vaner og målrettet adfærd, har altså fået stor indflydelse på vaneforskningen. Antagelsen har imidlertid mødt nogen kritik.

### **Er vaner og modelfri indlæring det samme?**

Sidestillingen af modelfri-modelbaseret-dimensionen med vane-målrettet adfærd dimensionen er dog blevet kritiseret med argumenter for at modelfri læring deler flere karakteristika med målrettet adfærd, end det gør med vaner (Miller et al., 2018).

Der bør ifølge kritikere skelnes mellem vaner og målrettet adfærd med henvisning til den rolle værdier spiller i styringen af adfærden. Målrettet adfærd er styret af repræsentationer af en forventet værdi. Vaner fungerer uafhængigt af sit udfald og dermed værdifrit. Modelfri adfærd er styret af forstærkning og er dermed værdibaseret. Ideen om et forstærknings- og dermed værdibaseret modelfrit system er således ikke foreneligt med konceptualiseringen af vaner som er uafhængige af adfærdens udfald og dermed værdifri. Idet både modelbaseret og modelfri adfærd er styret af forventningen om en belønning, er de værdibaserede og dermed målrettede (Miller et al., 2018).

At skelne mellem vaner og målrettet adfærd ved at henvise til en værdifri/værdibaseret-taksonomi, er identisk med den klassiske forståelse af vaner og målrettet adfærd, som styret af henholdsvis stimuli og mål. Den bidrager altså i sig selv ikke med noget nyt. Det, der er vigtigt i denne sammenhæng, er imidlertid forslaget om to separate dimensioner; en værdifri/værdibaseret dimension og en modelbaseret-modelfri-dimension. De to dimensioner har før været regnet for ens (Voon et al; Daw, Niv & Dayan., 2005; Daw et al., 2011; Otto et al., 2015). Men i Millers forståelse beskriver modelfri-model-baseret-dimensionen i stedet variationer i værdibaseret, målrettet handling, således at modelfri og modelbaserede strategier skal forstås som subkategorier af målrettet handling (Miller, 2018).

Derudover bør man overveje om modelfri læring kan være en forurenende faktor i de tests, der er tænkt til selektivt at måle vaner. Ligesom vaner og målrettet adfærd, adskiller modelfri og modelbaserede strategier sig fra hinanden ved deres forhold til ændringer i udfaldet af handlinger. Modelfri strategier er mindre fleksible overfor ændringer i udfald relativt til modelbaserede strategier. Vaner derimod er direkte uafhængige og ufleksible. Ved brug af en rent modelbaseret strategi i et klassisk udfalds-devalueringstudie vil man forvente at se en nær øjeblikkelig ændring i adfærden, som reaktion på ændringer i miljøet. En model-fri strategi vil ikke opdateres i samme tempo. Dette ligner mønsteret man ser ved vaner og målrettet adfærd i disse tests. Men den manglende effekt vil hos den modelfri læring forventes at være forsinket, hvor den hos vanen forventes at være fraværende. I forhold til fleksibilitet overfor ændringer i miljøet er der altså tale om et gradsspørgsmål mellem de forskellige konstrukter, hvor modelbaseret læring, modelfri læring og vaner bevæger sig fra at være fleksible til mindre fleksible til autonome fra deres udfald.

I et af de første udfalds-devalueringstudier blev rotter trænet til at trykke på en knap, for at få sukrose-tabletter. Efterfølgende blev tabletterne devalueret for halvdelen af rotterne, ved indsprøjtning af et giftstof, der inducerede kvalme. Rotterne havde samtidig fri adgang til sukrose-tabletterne via tryk på knappen, men de injicerede rotter holdt til sidst op med at spise dem, hvilket indikerede at devalueringen havde fundet sted. Til sidst, dagen efter devalueringen blev rotterne igen placeret i træningskammeret og udsat for udslukningstræning, det vil sige at de skulle trykke på knappen, men uden at modtage belønning. Resultatet var, at kontrolrotterne blev ved med at trykke på knappen, mens forsøgsrotterne allerede fra første træning havde signifikant lavere responsrater end kontrolrotterne. Dette blev af forfatterne tolket som, at rotterne havde indlært opgaven gennem en indre model over opgaven, hvor pres på knappen blev associeret med tabletter og tabletter med kvalme. Så selvom rotterne ikke havde nogen direkte erfaring med at pres på knappen fører til tabletter, der giver kvalme, kunne

rotterne lave denne kobling, ved hjælp af den simulerede model over opgaven. De havde således anlagt en modelbaseret strategi (Adams & Dickinson, 1981. I tilfældet af en modelfri strategi ville man først forvente, at adfærden ændrede sig, når rotten havde opnået direkte erfaring med, at knappen nu førte til sukkerpiller, der gav kvalme. I tilfælde af, at trykket på knappen var blevet vanebetinget ville man også forvente, at rotterne fortsat trykkede på knappen, på trods af devaluering, idet handlingen skulle være autonom af en sådan devaluering. Adfærden må således være målrettet og modelbaseret.

Havde resultatet derimod været sådan at der ingen forskel havde været på de to gruppers respons efter devalueringsprocessen, ville man umiddelbart ikke ville kunne sige noget om, hvorvidt der er tale om vaner eller modelfri strategi. I begge tilfælde ville man forvente fravær af en øjeblikkelig ændring i adfærd. At adfærden ikke umiddelbart tilpasses ændringer i opgavens værdiforhold, er altså ikke i sig selv nok til at skelne mellem adfærd styret af modelfri strategier og vaner.

Dette er ikke specifikt for opgaver der, som denne, lægger op til en modelbaseret strategi. Til eksempel kan bruges probabilistic reversal learning-opgaven (PRLT)(Swainson et al., 2000), som benytter sig af reversal indlæring som devaluering. I denne opgave lærer deltageren en S-A-O-sammenhæng idet en bestemt respons på en bestemt stimulus resulterer i enten belønning eller fravær af belønning. Efter indlæring af denne sammenhæng ændres forholdene sådan at de responser, der før gav belønning ved en bestemt stimulus, nu giver belønning ved en anden stimulus og omvendt. Man er interesseret i, hvor hurtigt deltageren tilpasser sin adfærd de nye forhold. En manglende fleksibilitet overfor de ændrede forhold antages at indikere vanestyret adfærd fordi den er autonom fra ændringer i udfaldts værdi (Swainson et al., 2000; D’Cruz et al., 2013). Deltageren får altså feedback gennem interaktion med opgaven og får således besked om at forholdene er ændret, ved at opleve, at de indlærte responser nu giver et andet resultat. En modelfri strategi er tilstrækkelig til at løse denne

opgave. Igen opstår det samme problem. Ved en modelfri indlæringsstrategi vil man forvente en forsinkelse af tilpasningen til de ændrede forhold og ved en vanestyret strategi vil man forvente ingen tilpasning. Men der er ingen regler for, hvor længe der skal gå, før man er sikker nok på, at adfærden ikke ændrer sig, til at kunne udelukke modelfri læring.

Jeg anfægter altså validiteten af udfalds-devalueringsopgaver, som de bruges og ser ud i dag, som mål for vaner overfor målrettet adfærd. Jeg mener at de i bedste tilfælde kan pege på brugen af modelbaserede strategier, men ikke udelukke brugen af andre målrettede strategier, eksempelvis modelfri læring fordi vaner og modelfri læring potentielt kan give ens resultater.

### **Delkonklusion**

Jeg har i det ovenstående argumenteret imod en sidestilling af vaner og modelfri indlæring. Jeg har argumenteret for at vaner adskiller sig fra den modelfri indlæringsstrategi ved at være værdifri. Modelfri indlæring adskiller sig fra vanen ved at være målrettet, og dermed ikke værdifri. De to konstrukter ligner hinanden ved begge at være mindre fleksible over for ændringer i værdiforhold i miljøet, relativt til modelbaseret indlæring. De adskiller sig imidlertid fra hinanden i dette aspekt, ved at vaner i deres definition er autonome overfor ændringer i værdiforhold mens modelfri indlæringsstrategier, selvom der er en vis forsinkelse sammenlignet med modelfri strategier, har en fleksibilitet overfor dem. Til sidst har jeg argumenteret for at modelfri læring er en mulig og sandsynlig confound i udfaldsdevalueringsstudier.

Jeg vil nu forholde mig til slips-of-action-testen. Med udgangspunkt i det ovenstående, vil jeg argumentere for, at modelfri læring er en mulig og sandsynlig confound i slips-of-action-testen.

## Slips of action testen

Slips-Of-Action-testen, henvises i RDoC til som *devalueringssopgave*. Det er en computerbaseret test, som består af flere faser, hvor den sidste er udviklet med henblik på at undersøge deficits i målrettet læring og vanedomineret kontrol over adfærden hos OCD-patienter (Gillan et al., 2011).

I den første fase, *diskriminationstrænings-fasen* bliver deltageren præsenteret for nogle forskellige kasser én ad gangen. Kasserne kan skelnes fra hinanden, ved at de har forskellige frugter tegnet udenpå. Deltageren skal nu vælge mellem to forskellige responser (to forskellige taster på tastaturet). Begge responser åbner kassen, men den ene respons viser en tom boks, mens den anden respons viser en frugt og udløser point til deltageren (DeWit et al., 2007). Deltageren skal altså i denne fase lære en  $S \rightarrow R \rightarrow U$  association. Deltageren gennemfører et fast antal runder (eng. Trials) og hver  $S \rightarrow R \rightarrow U$ -association bliver trænet 8 gange (DeWit et al., 2007).

Derefter kommer en *udfaldsdevalueringss-fase*. Her bliver deltageren præsenteret for 2 kasser ad gangen. Kasserne er åbne og har begge en frugt inden i. De to frugter var associeret med hver deres respons i diskriminationstræningsfasen. Et af frugterne har et stort kryds over sig der indikerer at det er devalueret (det giver ikke længere point). Deltagerne bliver instrueret i at give den respons, der tidligere var associeret med den ikke-devaluerede frugt. Fasen består af 14 runder, og hvert udfald bliver devalueret 2 gange (Gillan et al., 2011).

Deltagerne får også et *spørgeskema om deres eksplicitte viden om responserne og udfaldene*. Deltageren bliver vist alle stimulus-frugterne en ad gangen og skal svare på hvilken respons, der er korrekt og hvilket udfald en korrekt respons giver (Gillan et al., 2011).



Til sidst er der *slips-of-action fasen*. I denne fase bliver deltagerne først præsenteret for en skærm med alle de forskellige åbne kasser med udfaldsfrugter. To af dem har et rødt kryds over som indikerer at de nu er devaluerede. Billedet bliver vist i 10 sekunder. Efterfølgende bliver kasserne med stimuli-frugter vist en ad gangen hurtigt efter hinanden. Deltageren tjener points ved at give den respons, de indlærte i træningsfasen, til de stimuli, der stadig har værdifulde udfald. Hvis de giver en respons på de frugter, der nu har devaluerede udfald mister de point (Gillan et al., 2011). En sådan respons er en slips-of-action og indikerer vanekontrol. Hvis adfærden derimod er målrettet vil deltageren respondere selektivt på de frugter, der stadig har værdifulde udfald. Det er i denne fase at Gillan og kolleger mener at give et mere sensitivt mål for vaner, end i mere simple devalueringsopgaver, som udfaldsdevalueringfasen, fordi at vanesystemet her direkte konkurrerer med det målrettede system (Gillan et al, 2011).

## **Diskussion – problemer med slips-of-action testen**

### **Model-fri læring som mulig forstyrrende variabel**

Slips-of-action er en devalueringstest hvor en indlært instrumentel handling devalueres, så den ikke længere fører til et værdifuldt udfald. Slips-of-action adskiller sig fra mange andre devalueringsopgaver såsom probabilistic reversal learning (PRLT) (Swainson et al., 2000), ved sin relativt komplekse opbygning, og ved at korrekt præstation kræver en modelbaseret strategi. Det gør den, fordi test-delen foregår ved *udslukning*, hvilket resulterer i at deltageren i denne del af opgaven ikke har mulighed for den interaktion med miljøet, som er nødvendig for at kunne tilpasse adfærden gennem en modelfri indlæringsstrategi. Deltageren oplever ikke at handling 'X' i situation 'Y' nu giver et andet udfald end i træningsfasen. Deltageren får derimod noget isoleret information om ændringer i bestemte udfald, som ikke længere vil være værdifulde. For at kunne omsætte denne information til ændret adfærd må

deltageren benytte en indre model af opgaven, der kobler de specifikke udfald sammen med stimulus og respons. En succesfuld gennemførelse af træningsfasen kræver imidlertid ikke en modelbaseret tilgang. Det kan endda se ud til, at den lægger særligt op til en modelfri strategi. En modelbaseret strategi i træningsfasen ville se sådan ud, at deltageren havde et kognitivt kort, der associerer en bestemt respons med et bestemt frugt-udfald og det frugt-udfald med point. Fx et kort, der hedder ananas → 'v'-tast → kirsebær → point. Hvis deltageren havde en sådan mental model over opgaven og fik at vide at kirsebærret nu blev devalueret, ville deltageren vide, at en respons på ananassen ikke længere førte til point. Hvis deltageren ikke har lavet denne kobling, vil det at kirsebærret er blevet devalueret ikke have nogen betydning for deltageren, for personen vil ikke vide, at det betyder fravær af point. Noget lignende vil gælde hvis deltageren ikke har lavet en kobling mellem ananas, v-tast og kirsebær. Da vil deltageren ikke vide, at det er netop ananassen, der nu ikke giver point. Deltageren behøver netop ikke at have et sådant kort og lave disse koblinger for at gennemføre træningsfasen. Fordi deltagerens mål i opgaven er at opnå point, og fordi deltageren får sin pointscore at vide sammen med frugterne, er frugterne overflødige for at gennemføre opgaven. Deltageren behøver blot at vide at en 'v'-respons ved ananas-stimulussen giver point. Faktisk vil indlæringen af udfald-frugterne blot tilføje et ekstra led, som gør opgaveløsningen unødigt kompleks.

Meget peger på at træningsfasen i slips-of-action-testen ofte bliver løst modelfrit. I et forsøg hvor man har sammenlignet OCD-patienter og raske kontrolpersoners præstation på slips-of-action testen (Gillan et al., 2011), gav man deltagerne et spørgeskema, som blandt andet undersøgte deltagernes viden om sammenhængen mellem respons og udfald. Forsøget viste at deltagernes performance på slips-of-action testen flugtede med deres viden om disse sammenhænge. Alle deltagerne havde mere viden om sammenhænge i den kongruente diskrimination end i de to andre. Denne diskrimination var også den, hvor deltagerne klarede sig bedst i slips-of-action-fasen. Der var en mindre viden hos OCD-patienter sammenlignet

med raske kontrolpersoner (Gillan et al., 2011), hvilket flugter med effekten af testgruppe på mængden af slips-of-action i testen. Der var tilmed en korrelation mellem viden om udfaldene og mængden af slips-of-action (Gillan et al., 2011).

Nogle deltagere og nogle konditioner er altså forbundet med mindre viden om koblingen mellem bestemte S-R-par og deres udfald. Det kan tolkes som udtryk for, at træningsfasen er gennemført via modelfri indlæring. Hvis træningsfasen indlæres model-frit vil deltageren ikke kunne gennemføre slips-of-action-fasen, som kræver en model over opgaven. Dette er foreneligt med at viden om udfaldene i spørgeskemaet korrelerer med performance på slips-of-action-fasen og kunne indikere, at den effekt der måles i test-fasen, egentlig er deltagernes tendens til modelfri strategier i træningsfasen.

Udover den generelle problematik omkring adskillelsen af modelfri læring og vaner i udfald-devalueringstudier, er der altså yderligere en problematik knyttet til slips-of-action-testen, nemlig at testfasen kræver en modelbaseret strategi, mens træningsfasen kan løses modelfrit. Gillan et al's studie (2011) viser tilmed en sammenhæng mellem modelfri strategi i træningsfasen og performance i testfasen, hvilket understøtter hypotesen om modelfri strategi som en alvorlig confound i testen. Hypotesen bliver yderligere støttet af muligheden for at stimuli og udfald indkodes forskelligt i slips-of-action testen. Det vil jeg uddybe nedenfor.

### **Forskellig indkodning af udfald og stimuli**

Man har fundet ud af at det er sværere for både raske deltagere og deltagere med OCD at indlære den inkongruente diskrimination, sammenlignet med både den kongruente og bikonditionelle. Dette er blevet forklaret ved, at der i diskriminationerne dannes R-O-associationer. -Det vil sige at deltageren indlærer en sammenhæng mellem en respons og et udfald. Når dette udfald i en anden diskrimination fungerer som stimulus, men nu fordrer en

anden respons, opstår der responskonflikt, og derfor er denne type diskriminationer sværere at indlære end de andre typer, hvor der ikke opstår denne konflikt (de Wit et al., 2007; Gillan et al., 2011). Resultaterne er forenelige med respons-udfald-teori.

I et andet lignende forsøg med rotter har man fx brugt sukrose-opløsninger og madpiller som både stimulus og udfald i en inkongruent diskriminations-opgave og i dette studie fandt man imidlertid ikke den forventede effekt af inkongruens, altså ingen responskonflikt (Dickinson & de Wit, 2003). Det kunne pege på, at der ikke dannes de R-O-associationer, som skal resultere i responskonflikt i de inkongruente diskriminationer. De Wit et al. foreslår, at den manglende responskonflikt skyldes at stimuli og udfald indkodes forskelligt, således at stimuli indkodes ved deres sensoriske attributter, mens udfald indkodes ved deres motiverende karakteristika, altså deres evne til henholdsvis at slukke tørst og mætte (de Wit et al., 2007). Denne mulighed bakkes op af et studie, der peger på at repræsentationer af netop sukrose-opløsninger og madpiller aktiveres i form af deres motiverende karakteristika, når de aktiveres af en stimulus (Dickinson & Dawson, 1987). Hvis det er tilfældet, at stimuli og udfald indkodes forskelligt, vil der kunne være en R-O-association, der ikke skaber responskonflikt, fordi den samme genstand indkodes forskelligt afhængigt af, om den udgør en stimulus eller et udfald (de Wit et al., 2007). Den manglende viden om udfaldene i spørgeskemaet i Gillans slips-of-action-studie (Gillan et al., 2011) kunne altså forklares ved, at deltagerne har tilegnet sig viden om udfaldet i form af motiverende karakteristika, men uden at have indkodet en sensorisk karakteristik, som er det, spørgeskemaet beror på. Associationen kunne være i form af 'kirsebær → z-tast → point'. Dette ville tilmed være foreneligt med en manglende sensitivitet til devaluering af udfaldet, når denne devaluering præsenteres udelukkende gennem de sensoriske karakteristika. Hvis dette er tilfældet, betyder det, at man ikke kan sige, at man har målt vaner i testen, da man ikke har testet deltagerens sensitivitet til

udfald-devaluering. I stedet har man fået et mål for deltagernes indkodning af motiverende versus sensoriske karakteristika.

I slips-of-action-testen har man, for at undgå forskellig indkodning af stimuli og udfald, forsøgt at lave en opstilling, hvor udfald og stimuli kun adskiller sig fra hinanden i deres sensoriske karakteristika, altså deres farve, form og lignende. Mere specifikt har man villet undgå at udfald og stimuli kunne differentieres på baggrund af motiverende karakteristika (de Wit et al., 2007). Spørgsmålet er så, om dette er lykkedes. Eksempelvis står det ikke klart, hvorvidt de point deltageren får sammen med udfaldet tilføjer en motiverende dimension. Dermed kan det heller ikke udelukkes at udfaldet indkodes via disse karakteristika i stedet for de sensoriske. Altså at deltageren registrerer sammenhængen mellem en respons og points, snarere end sammenhængen mellem en respons og en given frugt.

Jeg har i det ovenstående foreslået to confounds i Slips-of-action-testen; at diskriminationen i testen løses gennem S-R-associationer, altså modelfrit, eller at den løses gennem S-R-O-associationer, men at de indkodes på baggrund af en anden dimension end den, devalueringen indikeres igennem. Begge forhold er forenelige med resultater fra studier, der har brugt slips-of-action-testen og begge forhold ville kunne forklare insensitivitet til devaluering af udfald i slips-of-action-testen.

### **Forslag til ændringer i Slips-Of-action-testen**

Jeg har peget på et hovedproblem i SOA-testen, som er at brugen af en modelfri strategi i træningsfasen, er en confound i testfasen, hvor der kræves modelbaseret læring. Hvis deltageren indlærer opgaven modelfrit opnår han måske ikke den nødvendige viden om udfaldene, til at vi med god sikkerhed kan sige at det er vaner der måles.

Jeg har peget på to mulige problemer/confounds i slips-of-action testen: 1) brugen af en modelfri strategi i træningsfasen kan udgøre en confound, fordi testfasen kræver en modelbaseret strategi og 2) indkodning baseret på såkaldte motiverende aspekter af udfaldene i testfasen kan udgøre en confound, fordi testfasen kræver viden om udfaldenes sensoriske karakteristika. Hvorvidt dette reelt er tilfældet, er fortsat et åbent spørgsmål og bør undersøges empirisk. Men hvis SOA-testen skal være et validt mål for vaner, bør disse mulige confounds fjernes eller kontrolleres for. En måde at minimere disse confounds er ved at justere testen således at disse 'uoverensstemmelser' mellem trænings- og test-fasen så vidt muligt minimeres eller helt fjernes, ved enten at sikre en modelbaseret indlæring i træningsfasen eller fjerne kravet om en indre model i testfasen. Testfasen kan udelukkende løses modelbaseret fordi den foregår in extinction. Hvis testfasen ændres til ikke længere at være in extinction vil den kunne løses modelfrit fordi deltageren modtager feedback på sine handlinger. Denne ændring ville imidlertid gøre slips-of-action-fasen tæt ved identisk med devalueringsfasen. Gillan og kolleger (2011) har netop tilføjet slips-of-action-fasen for at skabe konkurrence mellem vanedomineret og målrettet kontrol og få et mere sensitivt mål for deres relative kontrol. Der er altså ikke meget idé i at mindske confounds ved at ændre i slips-of-action-fasen. Derfor synes den bedste mulighed at være at forsøge at sikre brugen af en modelfri læringsstrategi i træningsfasen. Den mulighed vil jeg undersøge i det følgende afsnit.

Meget RL-forskning peger på, at det modelfri og det modelbaserede system konkurrerer om kontrollen over læring (Kool, Cushman & Gershman, 2018). I Daw two step-opgaven har det vist sig, at deltagerne typisk benytter en blanding af de to strategier, og forskelle i balancen mellem dem er associeret med en række faktorer som alder (Decker et al., 2016), psykiatriske symptomer (Gillan, Kosinski et al., 2014) og bestemte personlighedstræk (Skavota, Chan & Daw, 2015). En bestemt test eller situation lægger altså ikke entydigt op til den ene eller anden strategi eller en bestemt form for samspil mellem dem.

Der findes dog forskellige teorier om, hvordan der vælges mellem de to strategier i en given situation (Kool, Cushman & Gershman, 2018). Teorierne har det tilfælles, at de påstår at arbitreringen mellem systemerne foregår på baggrund af den ene eller anden form for trade-off, hvor fordele og ulemper ved hver strategi afvejes, og den mest effektive eller bedst egnede vælges. Fælles for de forskellige modeller er altså, at det system, der har de største fordele i den givne situation, får kontrollen over adfærden. Teorierne adskiller sig fra hinanden i detaljerne om, hvilke dimensioner disse fordele vurderes ud fra (Kool, Cushman & Gershman, 2018).

Fordele ved modelbaseret læring er dens fleksibilitet og en statistisk større sikkerhed end modelfri læring (Kool, Cushman & Gershman, 2018; Daw, Niv & Dayan, 2005). Dette kommer på bekostning af høje processeringsomkostninger (computational cost). Meget empiri peger på, at modelbaseret læring er eksekutivt krævende sammenlignet med modelfri læring. I et forsøg har man brugt Transmagnetisk Stimulation (TMS) til at forstyrre aktiviteten i dorsolateral prefrontal cortex (dlPFC) på deltagere, der gennemførte Daw two-step-opgaven og fundet at balancen mellem model-baseret og model-fri læring skiftede i retning af større modelfri dominans (Smittenaar et al., 2013). I et andet forsøg med Daw two-step har man på nogle af trialsene givet deltagerne en numerisk stroop task, som de skulle klare samtidigt med two-step-opgaven og fundet at deltagerne på de trials var mere tilbøjelige til at benytte en modelfri strategi. Forfatterne foreslår, at dette skyldes rekrutteringen af kognitive ressourcer til den eksekutivt krævende stroop-opgave (Otto, Gershman et al., 2013). Andre studier har vist sammenhænge mellem ringe performance på eksekutive prøver og tendens til model-fri strategier på Daw Two-step (Otto et al., 2015; Otto, Raio et al., 2013). Disse og andre studier (se fx Daw et al., 2011; Otto et al., 2015) peger på at modelbaserede strategier er kognitivt kostelige sammenlignet med modelfri strategier og særligt trækker på eksekutive funktioner. Modelfri læring er således mere effektiv idet det kommer med en lavere

processeringsomkostninger end modelbaseret læring (Kaelbling, Littman, & Moore, 1996). Til gengæld betyder den langsomme tilpasning til ændringer i miljøet, som er karakteristisk for modelfri læring, at det er mindre fleksibelt end modelbaseret læring (Kool, Cushman & Gershman, 2018). (Kool, Cushman & Gershman, 2018).

Nogle modeller foreslår, at arbitreringen mellem de to systemer er styret af, hvilken strategi der er forbundet med lavest usikkerhed i en given situation (Daw et al., 2005; Lee et al., 2014). Ifølge Daw og kolleger er sikkerheden i modelbaseret læring begrænset af kognitive ressourcer, mens den i modelfri læring er begrænset af mængden og tilgængeligheden af erfaring med miljøet (Daw et al., 2005).

Andre teorier påstår, at der vælges mellem de to systemer på baggrund af en cost-benefit afvejning; at modelbaseret læring vælges, hvis den værdi strategien kan afføde sammenlignet med en modelfri strategi, opvejer den kognitive indsats (Kool, Cushman & Gershman, 2018).

Denne diskussion er et speciale i sig selv værdigt, og jeg vil ikke uddybe den yderligere her. Ovenstående antyder dog, at der sker en konkurrence mellem de to systemer på baggrund af nogle parametre, der udspringer af de to systemers indbyrdes forskelligheder. En mere sikker viden om disse sammenhænge ville potentielt kunne øge mulighederne for at opbygge tests på måder, der øger sandsynligheden for brugen af én strategi fremfor en anden.

I opbygningen af slips-of-action-testens træningsfase, ser det ud til, at en modelbaseret strategi er forbundet med en ekstra indsats, som ikke tilføjer nogen ekstra værdi eller større sikkerhed. En indre model over sammenhænge mellem stimuli- og udfaldsfrugterne giver altså ikke større chance for point. Deltageren kan løse opgaven blot ved at danne stimulus-respons-associationer, det vil sige ved at lære hvilke responser, der belønnes ved hvilke stimuli. Information om hvilken frugt, der modtages sammen med det givne point, er



blot ekstra information, der kræver ekstra kognitive ressourcer, sammenlignet med blot at holde styr på om en handling giver point eller ikke.

Brugen af en modelbaseret strategi i denne fase er således kognitivt dyrere end en modelbaseret, samtidigt med, at den ikke bidrager men nogen fordel i form af større sikkerhed. Således vil man forvente en høj sandsynlighed for brugen af en modelfri strategi i træningsfasen hos både de modeller, der mener at arbitreringen mellem de to systemer er på baggrund af en cost-benefit-afvejning (Kool, Cushman & Gershman, 2018), såvel som dem, der mener, at det er på baggrund af en sammenligning af de to systemers sikkerhed (Daw et al., 2005; Lee et al., 2014). Denne sandsynlighed kan antageligt mindskes, ved at sænke den kognitive pris for en modelbaseret strategi, og/eller ved at gøre en modelbaseret strategi mere sikker end en modelfri strategi.

Et forslag til, hvordan dette kan gøres, er at "sammensmelte" de to informationer, altså udfald-frugten og pointet. Man kunne for eksempel lade bestemte frugter repræsentere point. Helt konkret kunne man tilføje en træningsfase i begyndelsen af testen, hvor deltageren indlærer associationer mellem forskellige frugter og en værdi. Frugterne præsenteres på skærmen én ad gangen og deltageren kan reagere ved at trykke på mellemrumstasten. Ved halvdelen af frugterne fører respons til et point, mens det ved de andre frugter fører til et minuspoint. Træningen slutter når deltageren har indlært frugternes værdi og kun trykker på mellemrumstasten ved de frugter, der giver point. I den efterfølgende træningsfase skal deltageren lære S-O-A-associationerne mellem stimulusfrugt, respons og udfald. Ved korrekt respons, ser deltagerne ikke længere deres points skrevet på skærmen, men blot den associerede udfald-frugt; en af de frugter, der i første træningsfase førte til point. Ved forkert respons ser deltageren en af de frugter som i forrige fase var forbundet med et minuspoint. Idet frugt-udfaldene repræsenterer point vil en stimulus-respons-udfald-model være nødvendig for at løse træningsfasen. En modelbaseret strategi vil således være forbundet med betydeligt højere

sikkerhed end en modelfri. Og den kognitive pris for en modelbaseret strategi vil i højere grad opvejes af større sandsynlighed for belønning, end i den oprindelige opbygning af testen.

Denne opbygning vil imidlertid ikke udelukke brugen af modelfri læring. Deltageren kan stadig indlære S-R-sammenhænge, uden at have koblet de konkrete udfald på S-R-parrene. Selvom pointene ikke præsenteres selvstændigt, men via frugtbillederne, kan stimulussen måske fortsat indkodes via en motivationel og en sensorisk dimension. Hvis deltageren indkoder den motiverende og ikke den sensoriske, kan resultatet være at deltageren fortsat ikke har kendskab til den specifikke frugt i stimulus-respons-udfald-parrene. Opbygningen ville dog højne sandsynligheden for indkodning af sensoriske karakteristika idet disse nødvendigvis må registreres, i modsætning til i den oprindelige opbygning, hvor det eneste, der nødvendigvis må registreres, er tallene på skærmen, der angiver pointene.

Ændringen kommer med den ulempe at pointene er nødt til at have samme værdi hver gang. Når pointene ikke skrives på skærmen, men repræsenteres via frugtbillederne, er det ikke muligt at skalere pointgivningen efter, hvor hurtigt responsen falder, ligesom Gillan og kolleger gør det (Gillan et al., 2011). De skriver ikke eksplicit hvilken funktion det tænkes at have, men man må formode, at det er med det formål at få deltagerne til at svare hurtigt. Det er ikke umiddelbart tydeligt, hvorfor dette er vigtigt i træningsfasen. Måske kan en hurtig responstid i træningsfasen endda mindske sandsynligheden for, at deltageren bruger en mere krævende modelbaseret strategi. Ligeledes øger det formentligt heller ikke sandsynligheden for en grundig indkodning af udfaldets sensoriske karakteristika. I andre devalueringstests, som *probabilistic reversal learning* har deltageren enten betydeligt længere tid til at svare, eller ingen tidsbegrænsning overhovedet (D'Cruz et al., 2013; Swainson et al., 2000). Det virker derfor ikke umiddelbart til, at denne ændring er af større betydning.

## **Delkonklusion**

Jeg har foreslået at validiteten af slips-of-action-testen som mål for sensitivitet i adfærd til ændringer i udfald, kan øges ved at lave en justering i testen, som nødvendiggør viden om de specifikke frugter ved hvert udfald i testfasen. Min hypotese er, at en sådan ændring vil øge sandsynligheden for en modelbaseret indlæringsstrategi i træningsfasen, samt indkodning af udfaldenes sensoriske karakteristika. Begge dele er afgørende for at kunne måle vaner i slips-of-action-testen, idet de begge er nødvendige for at løse testfasen.

### **Konklusion**

Dette speciale har forholdt sig kritisk til slips-of-action testen som validt mål for vaner. Det er blevet foreslået at modelfri læring er en forstyrrende variabel i slips-of-action testen. Jeg viser at problemet ligger i at træningsfasen lægger op til en modelfri strategi, mens testfasen lægger op til en modelbaseret strategi. Derudover er det blevet pointeret at muligheden for forskellig indkodning af stimuli og udfald i testen øger mistanken om at den måler modelfri læring snarere end vaner. Til sidst er det blevet foreslået at en ny udgave af testen, med bedre kontrol af indkodningen af stimuli og bedre kontrol over brugen af indlæringsstrategi (modelfri og modelbaseret), kan give et mere validt mål for vaner.

Hensigten med specialet er ikke at give en endegyldig dom over slips-of-action testen. Dels skal påstandene der er blevet præsenteret her efterprøves empirisk. Det bør eksempelvis bekræftes eller afkræftes empirisk om træningsfasen i slips-of-action-testen bliver indlært modelfrit. Ligeledes bør det belyses hvorvidt stimuli og udfald indkodes forskelligt. Derudover er en krydsvalidering af den klassiske slips-of-action, og den reviderede udgave som er foreslået her et oplagt næste skridt herfra.

Påstandene i dette speciale bygger på antagelsen om at modelfri læring og vaner ikke er sammenlignelige. Dette er der noget uenighed om i litteraturen og bør få mere

opmærksomhed, for at vi kan blive klogere på om modelfri læring er en hjælp eller en udfordring i undersøgelsen af vaner og dermed klogere på, hvordan vi skal bygge tests op for at imødekomme det.

Dette speciale skal således ses som en åbning til at begynde at forfine vane-konstruktet. Det er et gammelt konstrukt, som traditionelt er grundigt efterprøvet empirisk og hvor der har været stor konsensus om dens definition og operationalisering. I nyere tid er der imidlertid kommet tilføjelser til definitionen, som at vaner og modelfri indlæring er det samme (Daw et al., 2005) og nye paradigmer til at måle dem, som Slips-Of-Action (Gillan et al., 2011). Disse nye forslag bør modtage den samme grundige empiriske og teoretiske undersøgelse.

## Referencer

- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience*, 9(1), 357–381.  
<https://doi.org/10.1146/annurev.ne.09.030186.002041>
- Adams, C. D., & Dickinson, A. (1981). Instrumental Responding following Reinforcer Devaluation. *The Quarterly Journal of Experimental Psychology Section B*, 33(2b), 109–121. <https://doi.org/10.1080/14640748108400816>
- Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, 12(10), 366–375. [https://doi.org/10.1016/0166-2236\(89\)90074-x](https://doi.org/10.1016/0166-2236(89)90074-x)

- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience*, 9(1), 357–381.  
<https://doi.org/10.1146/annurev.ne.09.030186.002041>
- Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4–5), 407–419. [https://doi.org/10.1016/s0028-3908\(98\)00033-1](https://doi.org/10.1016/s0028-3908(98)00033-1)
- Balleine, B. W., & O’Doherty, J. P. (2009). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*, 35(1), 48–69. <https://doi.org/10.1038/npp.2009.131>
- Blackman, D. E. (2019). *Operant Conditioning: An Experimental Analysis of Behaviour (Psychology Revivals)* (1st ed.). Routledge.
- Blazquez, P. M., Fujii, N., Kojima, J., & Graybiel, A. M. (2002). A Network Representation of Response Probability in the Striatum. *Neuron*, 33(6), 973–982.  
[https://doi.org/10.1016/s0896-6273\(02\)00627-x](https://doi.org/10.1016/s0896-6273(02)00627-x)
- Bertsekas, D. P., Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Sci.
- Carcone, D., & Ruocco, A. C. (2017). Six Years of Research on the National Institute of Mental Health’s Research Domain Criteria (RDoC) Initiative: A Systematic Review. *Frontiers in Cellular Neuroscience*, 11. <https://doi.org/10.3389/fncel.2017.00046>
- Cheatwood, J., Reep, R., & Corwin, J. (2003). The associative striatum: cortical and thalamic projections to the dorsocentral striatum in rats. *Brain Research*, 968(1), 1–14.  
[https://doi.org/10.1016/s0006-8993\(02\)04212-9](https://doi.org/10.1016/s0006-8993(02)04212-9)
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711. <https://doi.org/10.1038/nn1560>

- Daw, N., Gershman, S., Seymour, B., Dayan, P., & Dolan, R. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- D'Cruz, A. M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, *27*(2), 152–160. <https://doi.org/10.1037/a0031721>
- Deckersbach, T., Chou, T., Britton, J. C., Carlson, L. E., Reese, H. E., Siev, J., Scahill, L., Piacentini, J. C., Woods, D. W., Walkup, J. T., Peterson, A. L., Dougherty, D. D., & Wilhelm, S. (2014). Neural correlates of behavior therapy for Tourette's disorder. *Psychiatry Research: Neuroimaging*, *224*(3), 269–274. <https://doi.org/10.1016/j.pscychresns.2014.09.003>
- Dickinson, A. (1985). Actions and habits: the development of behavioral autonomy. *Royal Society*, *308*(1135), 67-78. <https://doi.org/10.1098/rstb.1985.0010>
- Dickinson, A., & de Wit, S. (2003). The Interaction between Discriminative Stimuli and Udfalds during Instrumental Learning. *The Quarterly Journal of Experimental Psychology Section B*, *56*(1b), 127–139. <https://doi.org/10.1080/02724990244000223>
- Dickinson, A., & Dawson, G. R. (1987). Pavlovian processes in the motivational control of instrumental performance. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *39*(B), 201–213.
- Dickinson, A & Pérez, O. D. (2018). Actions and Habits: Psychological Issues in Dual-System Theory. In Morris, R. W., & Bornstein, A (Ed.), *Goal-Directed Decision Making: Computations and Neural Circuits* (pp. 407-428). Academic Press.
- Dolan, R., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. <https://doi.org/10.1016/j.neuron.2013.09.007>

- Economides, M., Kurth-Nelson, Z., Lübbert, A., Guitart-Masip, M., & Dolan, R. J. (2015). Model-Based Reasoning in Humans Becomes Automatic with Training. *PLOS Computational Biology*, 11(9), e1004463.  
<https://doi.org/10.1371/journal.pcbi.1004463>
- Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, 8(11), 1481–1489.  
<https://doi.org/10.1038/nn1579>
- Fanous, A., Gardner, C., Walsh, D., & Kendler, K. S. (2001). Relationship Between Positive and Negative Symptoms of Schizophrenia and Schizotypal Symptoms in Nonpsychotic Relatives. *Archives of General Psychiatry*, 58(7), 669.  
<https://doi.org/10.1001/archpsyc.58.7.669>
- Figeo, M., Pattij, T., Willuhn, I., Luigjes, J., van den Brink, W., Goudriaan, A., Potenza, M. N., Robbins, T. W., & Denys, D. (2016). Compulsivity in obsessive–compulsive disorder and addictions. *European Neuropsychopharmacology*, 26(5), 856–868.  
<https://doi.org/10.1016/j.euroneuro.2015.12.003>
- Fineberg, N. A., Chamberlain, S. R., Goudriaan, A. E., Stein, D. J., Vanderschuren, L. J. M. J., Gillan, C. M., Shekar, S., Gorwood, P. A. P. M., Voon, V., Morein-Zamir, S., Denys, D., Sahakian, B. J., Moeller, F. G., Robbins, T. W., & Potenza, M. N. (2014). New developments in human neurocognition: clinical, genetic, and brain imaging correlates of impulsivity and compulsivity. *CNS Spectrums*, 19(1), 69–89.  
<https://doi.org/10.1017/s1092852913000801>
- Fineberg, N., Menchon, J., Zohar, J., & Veltman, D. (2016). Compulsivity—A new trans-diagnostic research domain for the Roadmap for Mental Health Research in Europe (ROAMER) and Research Domain Criteria (RDoC) initiatives. *European*

*Neuropsychopharmacology*, 26(5), 797–799.

<https://doi.org/10.1016/j.euroneuro.2016.04.001>

- Gade, A. (2021). De kliniske amnesisyndromer. In Randi, S., Christian, G. & Anders, G. (Eds), *Klinisk psykologi*, (2 ed., pp. 146-177). Frydenlund.
- Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, 68(1), 101–128. <https://doi.org/10.1146/annurev-psych-122414-033625>
- Graybiel, A. M., & Rauch, S. L. (2000). Toward a Neurobiology of Obsessive-Compulsive Disorder. *Neuron*, 28(2), 343–347. [https://doi.org/10.1016/s0896-6273\(00\)00113-6](https://doi.org/10.1016/s0896-6273(00)00113-6)
- Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal Pathways in Primates Form an Ascending Spiral from the Shell to the Dorsolateral Striatum. *The Journal of Neuroscience*, 20(6), 2369–2382. <https://doi.org/10.1523/jneurosci.20-06-02369.2000>
- Hassani, O. K., Cromwell, H. C., & Schultz, W. (2001). Influence of Expectation of Different Rewards on Behavior-Related Neuronal Activity in the Striatum. *Journal of Neurophysiology*, 85(6), 2477–2489. <https://doi.org/10.1152/jn.2001.85.6.2477>
- Hoekstra, R. A., Bartels, M., Verweij, C. J. H., & Boomsma, D. I. (2007). Heritability of Autistic Traits in the General Population. *Archives of Pediatrics & Adolescent Medicine*, 161(4), 372. <https://doi.org/10.1001/archpedi.161.4.372>
- Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, 8(9), 725–732. <https://doi.org/10.1038/nrn2218>
- Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, 96(3), 451–474. [https://doi.org/10.1016/s0306-4522\(99\)00575-8](https://doi.org/10.1016/s0306-4522(99)00575-8)



- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4, 237–285.  
<https://doi.org/10.1613/jair.301>
- Kanazawa, I., Murata, M., & Kimura, M. (1993). Roles of dopamine and its receptors in generation of choreic movements. *Advances in Neurology*, 60, 107–112.
- Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, 139(4), 665–682. <https://doi.org/10.1037/a0020198>
- Kendler, K. S. & Gardner, C. O. Jr. (1998). Boundaries of major depression: an evaluation of DSM-IV criteria. *American Journal of Psychiatry*, 155, 172–177.
- Kimura, M., Aosaki, T. & Ishida, A. (1993). Neurophysiological aspects of the differential roles of the putamen and caudate nucleus in voluntary movement. *Advanced Neurology*, 60, 62–70
- Künzle, H. (1975). Bilateral projections from precentral motor cortex to the putamen and other parts of the basal ganglia. An autoradiographic study in *Macaca fascicularis*. *Brain Research*, 88(2), 195–209. [https://doi.org/10.1016/0006-8993\(75\)90384-4](https://doi.org/10.1016/0006-8993(75)90384-4)
- Künzle, H. (1977). Projections from the primary somatosensory cortex to basal ganglia and thalamus in the monkey. *Experimental Brain Research*, 30(4).  
<https://doi.org/10.1007/bf00237639>
- Leon, M. I., & Shadlen, M. N. (1999). Effect of Expected Reward Magnitude on the Response of Neurons in the Dorsolateral Prefrontal Cortex of the Macaque. *Neuron*, 24(2), 415–425. [https://doi.org/10.1016/s0896-6273\(00\)80854-5](https://doi.org/10.1016/s0896-6273(00)80854-5)
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). Oxford University Press.

- McGuire, J. F., Ung, D., Selles, R. R., Rahman, O., Lewin, A. B., Murphy, T. K., & Storch, E. A. (2014). Treating trichotillomania: A meta-analysis of treatment effects and moderators for behavior therapy and serotonin reuptake inhibitors. *Journal of Psychiatric Research*, 58, 76–83. <https://doi.org/10.1016/j.jpsychires.2014.07.015>
- Meehl, P. E. (1989). Schizotaxia Revisited. *Archives of General Psychiatry*, 46(10), 935. <https://doi.org/10.1001/archpsyc.1989.01810100077015>
- Miller, G. A. (2010). Mistreating Psychology in the Decades of the Brain. *Perspectives on Psychological Science*, 5(6), 716–743. <https://doi.org/10.1177/1745691610388774>
- Miller, J. K., Ludvig, E. A., Pezzulo, G., & Shenhav, A., (2018): Realigning Models of Habitual and Goal-Directed Decision-Making. In Morris, R. W., & Bornstein, A (Ed.), *Goal-Directed Decision Making: Computations and Neural Circuits* (pp. 407-428). Academic Press.
- Miller, G. A., & Rockstroh, B. (2013). Endophenotypes in Psychopathology Research: Where Do We Stand? *Annual Review of Clinical Psychology*, 9(1), 177–213. <https://doi.org/10.1146/annurev-clinpsy-050212-185540>
- National Institute of Mental Health. (2016). Behavioral Assessment Methods for RDoC Constructs. <https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/behavioral-assessment-methods-for-rdoc-constructs>
- Orbell, S., & Verplanken, B. (2010). The automatic component of habit in health behavior: Habit as cue-contingent automaticity. *Health Psychology*, 29(4), 374-383. doi:10.1037/a0019596
- Otto, A. R., Skatova, A., Madlon-Kay, S. & Daw, N. D. (2015): Cognitive control predicts the use of model-based reinforcement learning. *Journal of Cognitive Neuroscience*, 27(2), 319-333. Doi: 10.1162/jocn\_a\_00709

- Otto, A. R., Gersman, S. J., Markman, B. A., & Daw, N. D. (2013). The curse of planning: Dissecting Multiple Reinforcement-Learning Systems by Taxing the Central Executive. *Psychological Science*, 24(5) 751–761. doi: 10.1177/0956797612463080
- O'Donohue, W. T., & Ferguson, K. E. (2001). *The Psychology of B F Skinner* (1st ed.). SAGE Publications, Inc.
- Peterson, G. B., Linwick, D., & Overmier, J. (1987). On the comparative efficacy of memories and expectancies as cues for choice behavior in pigeons. *Learning and Motivation*, 18(1), 1–20. [https://doi.org/10.1016/0023-9690\(87\)90020-8](https://doi.org/10.1016/0023-9690(87)90020-8)
- Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. *Trends in Cognitive Sciences*, 16(1), 81-91. <https://doi.org/10.1016/j.tics.2011.11.009>
- Rothbart, R., & Stein, D. J. (2014). Pharmacotherapy of trichotillomania (hair pulling disorder): an updated systematic review. *Expert Opinion on Pharmacotherapy*, 1–11. <https://doi.org/10.1517/14656566.2015.972936>
- Swainson, R., Rogers, R. D., Sahakian, B. J., Summers, B. A., Polkey, C. E., & Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia*, 38(5), 596-612. [https://doi.org/10.1016/S0028-3932\(99\)00103-7](https://doi.org/10.1016/S0028-3932(99)00103-7).
- Larsen, I. U., Gausbøl, U. K. (2011). Gentagelser, vaner og tvang – om OCD og OC-spektrumforstyrrelser. Frydenlund.
- Thorndike, E. L., & Jelliffe. (1912). Animal Intelligence. Experimental Studies. *The Journal of Nervous and Mental Disease*, 39(5), 357. <https://doi.org/10.1097/00005053-191205000-00016>

- Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of Caudate Activity by Action Contingency. *Neuron*, *41*(2), 281–292. [https://doi.org/10.1016/s0896-6273\(03\)00848-1](https://doi.org/10.1016/s0896-6273(03)00848-1)
- Tsujimoto, S. (2004). Neuronal Representation of Response-Udfald in the Primate Prefrontal Cortex. *Cerebral Cortex*, *14*(1), 47–55. <https://doi.org/10.1093/cercor/bhg090>
- Tsuang, M. T., Stone, W. S., Tarbox, S. I. & Faraone, S. V. (2003). Insights From Neuroscience for the Concept of Schizotaxia and the Diagnosis of Schizophrenia. In K. A., Phillips, M. B. First, & H. A. Pincus (Eds.), *Advancing DSM: Dilemmas in Psychiatric Diagnosis* (pp. 105-128). American Psychiatric Association.
- Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R. N., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., & Bullmore, E. T. (2014). Disorders of compulsivity: a common bias towards learning habits. *Molecular Psychiatry*, *20*(3), 345–352. <https://doi.org/10.1038/mp.2014.44>
- Wickens, J. R. & Koetter, R. (1995). Cellular Models of Reinforcement. In J. L. Davis, D. G. Beiser (Eds), *Models of Information Processing in the Basal Ganglia* (pp. 187–214). MIT Press.Y
- Yee, C. M., Javitt, D. C., & Miller, G. A. (2015). ReplacingDSM Categorical Analyses With Dimensional Analyses in Psychiatry Research. *JAMA Psychiatry*, *72*(12), 1159. <https://doi.org/10.1001/jamapsychiatry.2015.1900>
- Yin, H. H., & Knowlton, B. J. (2004). Contributions of Striatal Subregions to Place and Response Learning. *Learning & Memory*, *11*(4), 459–463. <https://doi.org/10.1101/lm.81004>
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. <https://doi.org/10.1038/nrn1919>

- Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve udfald expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*(1), 181–189. <https://doi.org/10.1111/j.1460-9568.2004.03095.x>
- Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action-udfald learning in instrumental conditioning. *European Journal of Neuroscience*, *22*(2), 505–512. <https://doi.org/10.1111/j.1460-9568.2005.04219.x>
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, *22*(2), 513–523. <https://doi.org/10.1111/j.1460-9568.2005.04218.x>
- Yin, H., Knowlton, B., & Balleine, B. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–udfald contingency in instrumental conditioning. *Behavioural Brain Research*, *166*(2), 189–196. <https://doi.org/10.1016/j.bbr.2005.07.012>