



Instytut Farmakologii
im. Jerzego Maja
Polskiej Akademii Nauk

Praca doktorska

**Analiza modeli uczenia się ze wzmocnieniem związanym z
działaniem nagród naturalnych i substancji uzależniających w
zadaniu z probabilistycznym odwróceniem**

mgr Judyta Jabłońska

Promotorzy:

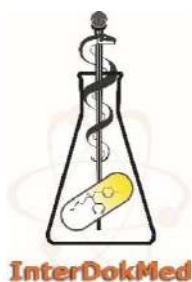
dr hab. Jan Rodriguez Parkitna

Instytut Farmakologii im. Jerzego Maja Polskiej Akademii Nauk

prof. dr hab. Piotr Zieliński

Instytut Fizyki Jądrowej im. Henryka Niewodniczańskiego Polskiej Akademii Nauk

Kraków, 2022



Rozprawa doktorska powstała w trakcie realizacji Środowiskowych Studiów Doktoranckich w ramach Projektu nr POWR.03.02.00-00-Io13/16, "Interdyscyplinarność dla medycyny innowacyjnej" InterDok-Med, realizowanych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój 2014-2020, współfinansowanego ze środków Europejskiego Funduszu Społecznego.



Unia Europejska
Europejski Fundusz Społeczny



Spis treści

Wykaz najczęściej stosowanych skrótów	4
1. Streszczenie	5
Abstract	7
2. Wprowadzenie.....	9
2.1 Zjawisko uczenia ze wzmocnieniem.....	9
2.2 Nagradzające działanie etanolu i sacharyny	10
2.3 Badania funkcji wykonawczych i elastyczności poznawczej z wykorzystaniem testu uczenia z odwróceniem	15
2.4 Neuronalne podłoże uczenia ze wzmocnieniem.....	17
2.5 Modelowanie obliczeniowe uczenia ze wzmocnieniem	24
3. Cele badawcze.....	34
4. Metody.....	35
4.1 Zwierzęta.....	35
4.2 Doświadczenia.....	36
4.3 Podsumowanie doświadczeń	43
4.4 Analiza statystyczna	43
4.5 Modelowanie procesu uczenia ze wzmocnieniem	44
5. Wyniki	47
5.1 Analiza uczenia ze wzmocnieniem zależnie od rodzaju nagrody w teście probabilistycznego odwrócenia w klatkach IntelliCage	47
5.2 Analiza uczenia ze wzmocnieniem zależna od wybranych receptorów w teście probabilistycznego odwrócenia w klatkach IntelliCage	65
5.3 Analiza uczenia ze wzmocnieniem u zwierząt szczepu NR1 ^{Tph2CreERT2} w teście probabilistycznego odwrócenia w klatkach do warunkowania instrumentalnego	81
6. Dyskusja	89
6.1 Uczenie z odwróceniem testowane behawioralnie	90
6.2 Modelowanie uczenia ze wzmocnieniem.....	98
6.3 Zależność między czasem i strategią podejmowania decyzji	101
7. Wnioski	104
8. Piśmiennictwo	106
9. Spis rycin.....	114
10. Spis tabel	115

Wykaz najczęściej stosowanych skrótów

5-HT	ang. 5-hydroxytryptamine, serotonin	5-hydrokсыtryptofan, serotoninina
AgRP	ang. Agouti-related protein	białko powiązane z Agouti
AIC	ang. Akaike Information Criterion	kryterium informacyjne Akaike
AMPA	ang. α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid	kwas α -amino-3-hydrokсы-5-metylo-4-izoksazolopropionowy
GABA	ang. γ -aminobutyric acid	kwas γ -aminomasłowy
GABA _A i GABA _B	ang. GABA receptors	receptory dla kwasu γ -aminomasłowego
NAc	ang. nucleus accumbens	jądro półleżące przegrody
NMDA	ang. N-methyl-D-aspartate	kwas N-metylo-D-asparaginowy
NMDAR	ang. N-methyl-D-aspartate receptor	receptor dla kwasu N-metylo-D-asparaginowego
POMC	ang. proopiomelanocortin	proopiomelanokortyna
REM	ang. rapid eye movement	faza snu, w której występują szybkie ruchy gałek ocznych
SERT	ang. serotonin transporter	transporter serotoniny
Tph2	ang. tryptophan hydroxylase 2	hydroksylaza tryptofanu 2
VTA	ang. ventral tegmental area	brzuszne pole nakrywki

1. Streszczenie

Zachowanie, które w danej sytuacji przynosi subiektywną korzyść, zostaje skojarzone z tą sytuacją i jeśli okoliczności powtórzą się, to wzrasta prawdopodobieństwo, że powtórzone zostanie także to zachowanie. Za wzrost prawdopodobieństwa powtórzenia działania odpowiada uczenie ze wzmocnieniem, które wskazuje się jako podstawowy mechanizm umożliwiający dostosowanie zachowania do zmian zachodzących w otoczeniu. Uczenie ze wzmocnieniem odgrywa kluczową rolę dla funkcji wykonawczych, dlatego też zaburzenia uczenia ze wzmocnieniem obserwowane są w szeregu zaburzeń neurologicznych i psychiatrycznych, np. w schizofrenii, uzależnieniu czy chorobie Parkinsona.

Celem moich badań było opracowanie testu behawioralnego pozwalającego na obserwację uczenia ze wzmocnieniem w warunkach możliwie zbliżonych do naturalnych oraz sprawdzenie czy obserwowane zachowanie w warunkach testu może zostać odwzorowane za pomocą modeli obliczeniowych. Wykorzystałam opracowaną metodykę do porównania wzmacniającego działania nagrody naturalnej (sacharyna) oraz substancji uzależniającej (etanol) oraz analizy zmian w uczeniu ze wzmocnieniem u genetycznie zmodyfikowanych myszy z selektywnymi mutacjami upośledzającymi działanie układu serotoninowego.

W badaniach wykorzystałam test probabilistycznego uczenia ze odwróceniem. Ten test został zaadoptowany do klatki IntelliCage, dzięki czemu przebiegał w naturalnym cyklu okołodobowym i bez deprywacji pokarmu. Co istotne, test przebiegał z minimalnym bezpośrednim udziałem eksperymentatora. Podczas testu przebywające w grupie kilkunastu osobników samice myszy miały dostęp roztworu sacharyny lub etanolu, zależnie od doświadczenia. Jednak dostęp do substancji był probabilistyczny, ponieważ drzwi prowadzące do butelek z nagrodą otwierały się z prawdopodobieństwem 30% lub 90%. W założeniu, uczenie ze wzmocnieniem powinno pociągać za sobą to że zwierzęta będą częściej wybierać opcję związaną z większym prawdopodobieństwem uzyskania nagrody. Istotnie, w przypadku grup testowanych z wykorzystaniem sacharyny, zaobserwowałam znamienne preferencję opcji związanej z pewniejszym uzyskaniem nagrody.

W celu odwzorowania zachowania myszy zaaplikowałam szereg modeli obliczeniowych z równaniami opartymi o założenia algorytmu *Q-learning*. Ze względu na efekty wpływu długości interwału na decyzję oraz potencjalne ograniczenia związane z pamięcią, włączyłam do badań rozszerzenia tego algorytmu. Spośród przeanalizowanych rozwiązań lepszym dopasowaniem do wyborów zwierząt charakteryzowały się modele zakładające aktualizację oczekiwań na podstawie zarówno podjętego i niepodjętego działania oraz różnicujące wrażliwość na pozytywną i negatywną informację zwrotną.

Pomimo różnic w wzmacniających właściwościach alkoholu i sacharyny nie zaobserwowałam wyraźnych różnic w mechanizmie uczenia ze wzmocnieniem u testowanych zwierząt.

Natomiast niezależnie od typu nagrody na prawdopodobieństwo powtórzenia wyboru istotny wpływ miał czas, jaki upłynął od poprzedniej próby uzyskania nagrody. Mutacje w postaci inaktywacji receptorów 5-HT₇ oraz pozbawienia receptorów NMDA na komórkach serotoninowych wykazały pomijalny efekt na uczenie ze wzmocnieniem, chociaż druga modyfikacja pociągnęła za sobą osłabienie tendencji do powtórzenia poprzedniego wyboru notowanej przy dłuższych odstępach czasowych.

Podsumowując, przedstawiłam nowy test uczenia ze wzmocnieniem oraz rozszerzyłam modele obliczeniowe przewidujące wybory dokonywane przez zwierzęta. Opracowana metodologia może posłużyć do badań w dziedzinie neuropsychofarmakologii i wnioskowania o zmianach powodowanych zaburzeniami neuropsychiatrycznymi na funkcje wykonawcze oraz skuteczności leków w przywracaniu normalnego funkcjonowania.

Abstract

If under given circumstances a specific behaviour is associated with subjectively positive consequences, the odds of repeating the same behaviour under similar circumstances will increase. The change in behaviour is driven by reinforcement learning and is essential for the ability to adapt to changes occurring in the environment. Reinforcement learning plays a key role in executive functions, and accordingly, impairments in reinforcement learning emerge in multiple neurological and psychiatric disorders, including schizophrenia, addiction and Parkinson's disease.

Here, I describe a task for the study of reinforcement learning under conditions resembling a near-natural environment and test if the observed behaviour can be explained with the use of computational models. Furthermore, I have used the novel task and models to assess differences in the reinforcing effects of natural (saccharin) and drug (alcohol) rewards, and also to study reinforcement learning in genetically modified mice with selective mutations in the serotonin system.

The task is based on probabilistic reversal learning using the IntelliCage, which permits to observe the behaviour of female mice remaining in a group, under a natural circadian cycle and without food or water deprivation. Most importantly, the task requires minimal direct interaction with the experimenter. Mice in the IntelliCage were offered access to bottles with saccharin or ethanol solution, in separate experimental groups. The availability of the reward was probabilistic as the doors guarding access to a substance were programmed to open with 30% or 90% probability each time the animal approaches them. A preference of the greater reward probability is the evidence of reinforcement learning, and it was robustly observed for animals with access to saccharin as a reward. The observed choices were fitted with computational models based on the *Q-learning* algorithm, which I have extended to account for the effects of time intervals between choices and potential limitations imposed by memory. The best fit to the observed behaviour was achieved by models that assumed distinct learning rates for the rewarded and non-rewarded outcomes and also included an update of the expected value of the option not taken.

I find that despite differences in apparent reinforcement strength between alcohol and saccharine solutions, there were no obvious changes of learning mechanisms. Conversely, irrespective of reward type, there was a robust effect of the time interval between choices on the probability of repeating the same choice again. Furthermore, complete inactivation of 5-HT₇ receptors or ablation of NMDA receptors on serotonergic neurons had only limited effects on reinforcement learning, though the latter weakened the tendency of repeating the previous choice at longer time intervals.

In summary, I present a new reinforcement learning task and extended computational models that approximate animal behaviour in the task. The methodology is potentially useful for further studies in the field of neuropsychopharmacology, to model the effects of neuropsychiatric diseases on executive functions and the potential efficacy of therapies in restoring normal behaviour.

2. Wprowadzenie

2.1 Zjawisko uczenia ze wzmocnieniem

Uczenie się stanowi proces oparty na doświadczeniu, skutkujący trwałą zmianą w zachowaniu lub procesach umysłowych. Uczenie się różni się od zachowania instynktownego, czyli typowego dla gatunku i genetycznie zaprogramowanego, możliwością elastycznego przystosowania do zmiennych sytuacji i nowego środowiska. Do opisu procesu uczenia się, zarówno u ludzi i innych zwierząt, wykorzystywano głównie ujęcie behawiorystyczne (Berridge, 2000). W ujęciu behawiorystycznym zachowanie opisuje się w kategoriach bodźców i reakcji, a więc obserwowalnego działania (Berridge, 2000).

U podstaw współczesnej wiedzy o mechanizmach uczenia się leżą badania nad warunkowaniem bodźców, poczynając od przykładu związanego ze ślinieniem się psa przy jedzeniu. Jeśli podanie pokarmu wielokrotnie poprzedza bodziec, jak dźwięk dzwonka czy ton z kamertonu, zwierzę zaczyna ślinić się już podczas prezentacji takiego bodźca, co zaobserwował w swoich badaniach nad fizjologią trawienia Iwan Pawłow (cytowane za (Konorski, 1949)). Zdolność do kojarzenia odruchowych reakcji z nowymi bodźcami określona jest jako warunkowanie klasyczne. Jeśli bodziec bezwarunkowy (pokarm) poprzedza bodziec neutralny (dźwięk) to w wyniku warunkowania już sam dźwięk wywoła reakcję bezwarunkową (ślinienie). Znacznych odkryć w zakresie uczenia się dostarczyły badania Jerzego Konorskiego, który jako pierwszy proponował molekularne wyjaśnienie takiego zachowania, w kategoriach plastyczności synaptycznej (cytowane za (Bijoch i in., 2020)). Kluczowy dla tego uczenia się jest układ bodźców w czasie, który powinien być przylegający, czyli bodziec obojętny i bezwarunkowy powinny następować krótko po sobie. Dokładny zakres odstępu zależy od typu działania, np. przy reakcji mrugnięcia do sekundy, w reakcji awersji do smaku nawet do kilku godzin. Do ważnych przykładów warunkowania klasycznego należy awersja pokarmowa, fobie, czy zespół stresu pourazowego. Natomiast wiedza o zagadnieniach tego uczenia może posłużyć do zastosowań terapeutycznych, np. do przeciwwarunkowania, czyli wygaszania reakcji lękowych, poprzez uczenie spokojnego reagowania na bodziec warunkowy (Zimbardo i in., 2017).

Warunkowanie klasyczne odbywa się poprzez odruch, który jest działaniem biernym i mimowolnym. Z kolei do sytuacji gdy zwierzęta działają dla uzyskania jakiegoś celu odnosi się warunkowanie sprawcze. Stanowi to formę uczenia się w której reakcja zwierzęcia zmienia się na skutek konsekwencji tego działania, czyli bodźców następujących po działaniu, nagród i kar. Obrazuje to sformułowane przez Edwarda Thorndike'a prawo efektu stanowiące w uproszczeniu, że podczas uczenia się reakcje prowadzące do przyjemnych skutków będą powtarzane (Thorndike, 1927). W przedstawionym przez Thorndike'a przykładzie, kot znajdujący się w skrzynce problemowej rozpoznaje, metodą prób i błędów, w jaki sposób otworzyć drzwi od-

dzielające go od pokarmu (Thorndike, 1898). Bodziec, który następuje po reakcji i zwiększa jej siłę zostałby określony w podejściu behawiorystycznym, promowanym przez Burrhusa Frederica Skinnera, jako czynnik wzmacniający (ang. *reinforcer*) (Skinner, 1950). Wzmocnienie pozytywne odnosi się do sytuacji gdy ponowne wystąpienie reakcji staje się bardziej prawdopodobne, np. mysz w celu uzyskania pokarmu będzie wsadzać nos do jednego z portów klatki do warunkowania instrumentalnego. Natomiast wzmocnienie negatywne dotyczy wzmocnienia reakcji w celu usunięcia awersyjnego bodźca. Przeciwnie do wzmacniania, karanie powinno osłabiać zachowanie po którym następuje awersyjna konsekwencja (kara) i podobnie do wzmocnienia także w przypadku karania można mówić o karaniu pozytywnym oraz negatywnym. Do ważnych aspektów warunkowania należy częstotliwość otrzymywania wzmocnienia. We wzmacnianiu ciągle nagradza się każdą poprawną reakcję, w sporadycznym jedynie wybrane. W tym przypadku rozkłady wzmacniania można określić przez liczbę reakcji (rozkład stosunkowy) lub przez upływ czasu (rozkład interwałowy) oraz przez przewidywalność, jako stałe i zmienne proporcje lub odstępy czasowe. Rozkład stosunkowy o zmiennej, wzrastającej proporcji wykorzystywany jest w badaniach nad procesami motywacyjnymi poprzez odpowiedź na pytanie ile reakcji maksymalnie wykona zwierzę dla uzyskania nagrody.

Natomiast przykłady aktywności wykraczającej poza podejście radykalnego behawioryzmu skłaniają do opisu zjawiska także w koncepcjach psychologii poznawczej, w której uczenie przyjmuje postać zmian w procesach umysłowych nie zawsze manifestujących się w zachowaniu (Zimbardo i in., 2017). Przykładowo, mapa poznawcza odnosi się do umysłowej reprezentacji układu przestrzennego środowiska, którą według Edwarda Tolmana, posłużyły się szczury do odnajdywania dróg do pokarmu w różnych wersjach labiryntów (Tolman, 1948). Wskazuje to na występowanie umysłowej reprezentacji przestrzeni, w której znajduje się zwierzę oraz może stanowić podstawę do założenia, przyjętego w niniejszych badaniach, o zdolności myszy do wypracowania wewnętrznej strategii czy sposobu rozwiązywania zadania.

Uczenie ze wzmocnieniem odnosi się więc do uczenia sprawczego warunkowanego czynnikiem wzmacniającym, nagrodą. Intuicyjnie, nagrodę pierwotną stanowiłby pokarm dla zwierzęcia głodnego czy woda dla spragnionego, ale także skojarzone z pierwotnymi czynnikami wzmacniającymi bodźce wtórne, nie zaspokajające podstawowych potrzeb, jak zainteresowanie, pochwała czy uśmiech (Berridge, 2000; Zimbardo i in., 2017). Przetwarzanie sygnału o nagrodzie w układzie nerwowym, w przeciwieństwie do bodźców wzrokowych, czuciowych czy bólowych, jest złożone, rozproszone, oraz nie do końca poznane (Berridge, 1996; Kelley i Berridge, 2002; Wise, 2002; Wolfram Schultz, 2015).

2.2 Nagradzające działanie etanolu i sacharyny

Za nagradzające działanie słodkiego smaku, pokarmu lub etanolu odpowiada szereg częściowo zbieżnych mechanizmów. Kluczową rolę w działaniu nagród odgrywają neurony dopa-

minowe brzuszego śródmózgowia, które wysyłają projekcje do zwojów podstawy i kory czołowej. W centralnym układzie nerwowym występuje 5 głównych szlaków dopaminowych (Stahl, 2013). Szlak mezolimbiczny (ang. *mesolimbic pathway*) prowadzi od (ang. *ventral tegmental area*, VTA) do jądra półleżącego (ang. *nucleus accumbens*, NAc) znajdującego się w brzusznej prążkowiu i jego znaczenie odnotowano dla przetwarzania motywacji, przyjemności oraz nagrody, dlatego zyskał miano kluczowego elementu „układu nagrody” (Kelley i Berridge, 2002; Stahl, 2013; Wise i McDevitt, 2018). Szlak mezokortykalny prowadzi także od VTA, jednak wysyła aksony do obszarów kory przedczołowej. Wskazuje się, że odgałęzienia tego szlaku do grzbietowo-bocznej kory przedczołowej odgrywają istotną rolę w funkcjach poznawczych i wykonawczych, podczas gdy odgałęzienia tego szlaku do brzuszno-przyśrodkowych części kory przedczołowej są zaangażowane w funkcje afektywne (Stahl, 2013). Ze względu na działanie układu nagrody, niektórzy badacze łączą dwie powyższe drogi określając powstały szlak jako mezokortykolimbiczny (ang. *mesocorticolimbic*) (Le Moal i Simon, 1991; Lüscher i Ungless, 2006). Występuje także szlak nigrostriatalny, prowadzący od istoty czarnej (ang. *substantia nigra*) do jąder podstawy (w tym prążkowia), którego główną funkcją jest kontrola poruszania, szlak guzowo-lejkowy (ang. *tuberoinfundibular*) biegnący od podwzgórza do przedniego płata przysadki mózgowej oraz, najpóźniej opisany, piąty szlak dopaminowy, unerwiający wzgórze i powstający z wielu miejsc, w tym istoty szarej okółowodociągowej (ang. *periaqueductal gray*), brzuszego śródmózgowia, jąder podwzgórza i bocznego jądra przyramionowego (ang. *lateral parabrachial nucleus*) (Stahl, 2013).

W badaniach behawioralnych jako nagrodę naturalną działającą wyłącznie poprzez odczucie słodkiego smaku, stosuje się słodziki. Najczęściej wykorzystywany słodzik sacharyna, oznaczana w produktach spożywczych jako E954 (Polskie Towarzystwo Diabetologiczne, 2020), nie jest metabolizowana i zostaje wydalona w postaci niezmienionej (Sweatman i Renwick, 1979; Magnuson i in., 2016). Nagrodę pokarmową można „lubić” (ang. *liking*), ze względu na smak oraz, niezależnie, można jej „chcieć” (ang. *wanting*), a więc działać, żeby ją uzyskać (Berridge, 1996). Dla obu tych składowych nagrody naturalnej postuluje się występowanie osobnych obwodów neuronalnych. W kontekście odczucia smaku, sacharyna, podobnie jak inne substancje o słodkim smaku, działa przez pobudzenie receptorów w jamie ustnej. Informacja o smaku pokarmu odbierana jest od komórek receptorowych w kubkach smakowych, a następnie transmitowana zwojami czuciowymi do znajdującego się w pniu mózgu jądra pasma samotnego (ang. *nucleus tractus solitarius*), które stanowi miejsce integrujące informacje o stanie fizjologicznym układu pokarmowego oraz kontrolujące działania takie jak połykanie, lizanie czy gryzienie (Simon i in., 2006; Matsumoto, 2013). Już na tym poziomie może być kodowana informacja nie tylko o smaku pokarmu ale także o jego „smaczności” (ang. *palatability*), zależącej także od stanu fizjologicznego organizmu (Berridge, 1996). Informacja o pokarmie jest następnie transmitowana poprzez jądra mostu i wzgórze do pierwszorzędowej kory smakowej. Połączenia z pierwszorzędowej kory smakowej biegną następnie do ciał migdałowatych a stamtąd do bocznego podwzgórza i neuronów dopaminowych śródmózgowia (Simon i in., 2006). Na-

tomiaś „chcenie” nagrody naturalnej związane jest z dopaminowym układem mezokortyko-
limbicznym (Morales i Berridge, 2020). Słodki smak działa nagradzająco, natomiast należy
pamiętać, że za sięganie po pokarm odpowiada złożony układ regulacji pobierania pokarmu.

Substancje uzależniające są chemicznie niejednorodną grupą o odmiennych celach mo-
lekularnych. Cechą wspólną wszystkich substancji uzależniających jest to, że zwiększają one
stężenie dopaminy w docelowych strukturach szlaku mezokortykolimbicznego. Przyjmuje się,
że uwalnianie dopaminy z projekcji tego szlaku odgrywa kluczową rolę w wywoływaniu za-
chowania kompulsywnego obserwowanego przy uzależnieniu (Wise i Bozarth, 1987; Di Chiara
i Imperato, 1988; Lüscher i Ungless, 2006).

Najnowsze badania wskazują, że mechanizm działania etanolu jest złożony, ze względu
na występowanie zarówno specyficznych jak i pośrednich miejsc wiązania tego alkoholu, włą-
czając podjednostki kanałów jonowych, wewnątrzkomórkowe białka sygnalizacyjne, czynniki
transkrypcyjne, czynniki wzrostu, białka zaangażowane w epigenetyczną regulację ekspresji
genów, a nawet lipidy błonowe (Abraham i in., 2017). Klasyczne i uproszczone ujęcie mecha-
nizmu działania etanolu polega na wzmacnianiu hamowania synaps GABA (kwas γ -
aminomasłowy) i zmniejszeniu pobudzenia synaps glutaminianowych, co zostało zbadane
głównie na przykładzie formacji hipokampa. Etanol zwiększa uwalnianie GABA m.in. poprzez
blokowanie presynaptycznych receptorów GABA_B, a także stymulację postsynaptycznych recep-
torów GABA_A (Weiner i Valenzuela, 2006; Stahl, 2013). Etanol może hamować uwalnianie glu-
taminianu poprzez działanie na presynaptyczne metabotropowe receptory glutaminianu
(mGluR), presynaptyczne kanały wapniowe wrażliwe na napięcie (VSCC), postsynaptyczne
receptory NMDA (dla kwasu N-metylo-D-asparaginowy) oraz postsynaptyczne receptory
mGluR (Stahl, 2013). Wskazano także zależność wzmacniającego działania alkoholu od układu
opiodowego, co wspiera chociażby fakt stosowania w terapii uzależnienia naltreksonu, czyli
antagonisty receptorów μ -opiodowych (Stahl, 2013).

Mechanizm działania etanolu na układ nagrody stanowił przedmiot badań Amber Al-
hadeff, Nitsan Goldstein i wsp. W tych badaniach dożołądkowe podanie zarówno etanolu (5%
lub 10%) jak i tłuszczu (10%) zwiększyło aktywność neuronów dopaminowych VTA już w
pierwszej próbie, wskazując, że obie te substancje aktywują układ nagrody (Alhadeff i in.,
2019). Różnice między efektami tych substancji zanotowano natomiast w jądrze półleżącym
(NAc), którego aktywność zwiększyła się pod wpływem etanolu, ale nie tłuszczu oraz w grzbie-
towym prążkowiu, z dokładnie przeciwną obserwacją (Alhadeff i in., 2019). Według autorów,
taki wynik sugeruje, że neurony VTA aktywowane przez etanol transmitują informację do NAc,
natomiast neurony VTA aktywowane przez pokarm w postaci tłuszczu do obszarów poza NAc
(Alhadeff i in., 2019). Oznacza to dodatkowo, że zarówno etanol jak i pokarm działają na układ
nagrody bez udziału bodźców zmysłowych.

Etanol dostarcza energii metabolicznej oraz może wpływać na pobieranie pokarmu i
wagę ciała, chociaż badania nie dostarczyły jednoznacznej odpowiedzi czy powoduje zwiększe-
nie czy zmniejszenie sięgania po jedzenie. Dotychczasowe badania mechanizmów neuronalnych

leżących u podłoża regulacji pobierania pokarmu dotyczyły osobno jedzenia związanego z utrzymaniem masy ciała oraz funkcjami metabolicznymi („homeostatyczne”) oraz sięgania po pokarm spowodowanego przyjemnością, odczuciem nagrody („hedonistyczne”) (Rossi i Stuber, 2018). W swojej pracy z 2018r. Mark Rossi i Garret Stuber wskazali, że obwody związane z oboma tymi funkcjami oddziałują ze sobą oraz dostarczyli koncepcję integracji tych mechanizmów na podstawie dowodów anatomicznych i funkcjonalnych (Rossi i Stuber, 2018). Jedzenie „homeostatyczne” związane jest z homeostazą energetyczną, przetwarzaniem sygnałów o głodzie i sytości, a kluczowym, pierwszym, elementem odpowiadającego mu obwodu jest jądro łukowate (ang. *arcuate nucleus*), znajdujące się w podwzgórzu w pobliżu komory trzeciej, w którego komórkach ekspresji ulegają receptory m.in. dla leptyny, greliny i insuliny, czyli hormonów wydzielanych z układu pokarmowego (Rossi i Stuber, 2018). Jądro łukowate oraz pobliskie jądro przykomorowe, w koncepcji autorów, tworzy grupę neuronów „przykomorowych”, które integrują informacje od układu pokarmowego i modułują cele „pośrednie”. Do neuronów „pośrednich”, zlokalizowanych w strukturach podkorowych, będą należeć w tej koncepcji: boczne podwzgórze (ang. *lateral hypothalamic area*), jądro okołoramieniowe (ang. *parabrachial nucleus*), uzdeczka (ang. *habenula*), jądro centralne ciała migdałowatego (ang. *central nucleus of the amygdala*) oraz jądro łożyskowe prążka końcowego (ang. *bed nucleus of the stria terminalis*). Z grupy „pośredniej” informacja będzie transmitowana do neuronów „monoaminergicznym”, czyli układu dopaminowego oraz serotoninowego, które związane są także m.in. z motywacją, oraz biorąc pod uwagę dalsze obwody, z funkcjami wykonawczymi (Rossi i Stuber, 2018). Na szczególną uwagę w tym przypadku zasługuje połączenie obszarów „pośrednich” i VTA oraz transmisja od VTA do kory przedczołowej (ang. *prefrontal cortex*, PFC) (Rossi i Stuber, 2018), ponieważ stanowią one część obwodów neuronalnych odpowiedzialnych za procesowanie informacji o nagrodzie (Wise, 2002; Wise i McDevitt, 2018).

W wspomnianych badaniach Amber Alhadeff, Nitsan Goldstein i współpracowników wykonano także serię doświadczeń w celu określenia w jaki sposób etanol odróżnia się od nagród naturalnych w regulacji pobierania pokarmu (Alhadeff i in., 2019). W pierwszym kroku badacze ustalili w jaki sposób etanol działa na jądro łukowate, które zawiera dwie populacje neuronów kluczowe dla regulacji pobierania pokarmu: komórki z ekspresją białka powiązanego z Agouti (ang. *Agouti-related protein*, AgRP) oraz proopiomelanokortyny (ang. *proopiomelanocortin*, POMC). W badaniach mierzono *in vivo* fluorescencję zależną od wapnia u myszy z ekspresją wskaźnika wapniowego (GCaMP6s) w neuronach AgRP i POMC (Alhadeff i in., 2019). Pobieranie pokarmu aktywuje neurony AgRP i hamuje neurony POMC, natomiast stymulacja tych grup neuronów powoduje odpowiednio wzrost i spadek pobierania pokarmu. Zarówno dożołądkowe podanie etanolu (5% albo 15% w 1 mL) jak i samo-podawanie etanolu (10%) zmniejszyło aktywność neuronów AgRP, natomiast nie wpłynęło na aktywność neuronów POMC. Brak zmiany aktywności neuronów POMC sugeruje, że mechanizm działania etanolu jest inny niż dla pokarmu. Dodatkowo, określono, że efekt podanego dożołądkowo etanolu na neurony AgRP, w przeciwieństwie do tłuszczu, nie był transmitowany poprzez nerw błędny. Co

więcej, optogenetyczna stymulacja neuronów AgRP nie spowodowała zwiększenia spożycia etanolu (10%), oczekiwanego w sytuacji gdyby alkohol był postrzegany jako substancja kaloryczna. Według autorów, nie było to spowodowane nieprzyjemnym smakiem etanolu ze względu na to, że zwierzęta pod wpływem analogicznej stymulacji były skłonne pić roztwór glukozy (8%) z awersyjną chininą (0,3 mM). Na tej podstawie autorzy określili, że etanol nie jest postrzegany jako pokarm (Alhadeff i in., 2019). W kontekście pobierania pokarmu obserwowanego behawioralnie, obniżenie pobierania pokarmu w wyniku podania etanolu zaobserwowano wyłącznie u myszy utrzymywanych na diecie (85-90% masy ciała) (Alhadeff i in., 2019). Natomiast warto wziąć pod uwagę, że chroniczna deprywacja pokarmowa zwiększa wartość nagradzającą substancji uzależniającej (Carr, 2007). Podsumowując, autorzy zasugerowali, że etanol i pokarm inaczej modulują obwody regulacji pobierania pokarmu (Alhadeff i in., 2019). Podobieństwa sugerują natomiast możliwość dalszego rozwoju badań, np. w kontekście otyłości traktowanej jako rezultat uzależnienia (Volkow i Wise, 2005).

Poprzez działanie na wiele celów molekularnych etanol powoduje szeroki zakres efektów (Heilig i Spanagel, 2021). Ogólnie, efekt etanolu określany jest jako dwufazowa progresja (ang. *biphasic progression*), którą charakteryzuje w pierwszej fazie stymulacja psychomotoryczna, obniżona lękowość i rozhamowanie, w kolejnej, wraz ze zwiększaniem dawki, działanie depresyjne na centralny układ nerwowy, objawiające się ataksją (zaburzeniem koordynacji ruchowej), progresywnym upośledzeniem poznawczym, przejściową amnezją następczą (tzw. „urwanie filmu”) aż do utraty przytomności, zależnie od indywidualnej wrażliwości. Chroniczne sięganie po alkohol może prowadzić do wykształcenia tolerancji i zespołu uzależnienia, który obejmuje eskalację przyjmowania środka, trudności z kontrolowaniem jego używania, w tym sięganie po niego pomimo szkodliwych następstw, przedkładanie przyjmowania nad inne zajęcia, zwiększoną tolerancję, a czasem także zespół abstynencyjny (Światowa Organizacja Zdrowia, 2009). Chroniczne sięganie po alkohol może zaburzać podejmowania decyzji i zwiększać impulsywność (Heilig i Spanagel, 2021). Aspekty zachowania szczególnie związane ze spożyciem alkoholu, zarówno u ludzi jak i gryzoni, dotyczą hamowania działania, awersji do opóźnienia (ang. *delay aversion*) czy perseweracji, rozumianej jako skłonność do powtarzania tego samego działania, interpretowanej tutaj jako pozostanie w skupieniu pomimo rozproszeń (Dick i in., 2010). Wspomniane aspekty naturalnie związane są z czasem podjęcia działania. Co istotne, nadmierne chroniczne spożycie etanolu prowadzi do zaburzeń poznawczych, głównie upośledzenia funkcji wykonawczych (procesów poznawczych odpowiedzialnych za właściwą organizację, wykonanie i utrzymanie działań ukierunkowanych na cel), pamięci epizodycznej i zdolności wzrokowo-przestrzennych (Stephens i Duka, 2008; Bernardin i in., 2014).

Mechanizmy nagradzającego działania sacharyny i alkoholu są złożone, jakkolwiek zbiegają się na układzie dopaminergicznym w brzuszny mózgowie. Niemniej, odnotowane różnice pozwalają na przypuszczenie, że ich efekt behawioralny związany ze wzmocnieniem może być różny.

2.3 Badania funkcji wykonawczych i elastyczności poznawczej z wykorzystaniem testu uczenia z odwróceniem

Elastyczności poznawcza to zdolność do szybkiego dostosowania zachowania w zmieniającym się środowisku. Upośledzenie tej zdolności obserwuje się w zaburzeniach psychiatrycznych i neurologicznych (Cools, 2015; Waltz, 2017). Do badań elastyczności poznawczej u ludzi, innych naczelnych i gryzoni stosowany jest m.in. zbiór paradygmatów określany jako uczenie z odwróceniem (przeuczanie się, ang. *reversal learning*) (Cools, 2015; Izquierdo i in., 2017). W prostym schemacie uczestnik badania w sytuacji wielokrotnego wyboru między dwoma opcjami (bodźcami wizualnymi czy położeniem przestrzennym obiektów), z których jedna związana jest z uzyskaniem nagrody a druga nie, uczy się rozpoznawać konsekwencje ich wyboru. Po osiągnięciu kryterium, wskazującego poziom wykonania zadania, przypisanie opcji i rezultatu zostaje zmienione na dokładnie przeciwne, a uczestnik ponownie uczy się aż do osiągnięcia kryterium.

Jest to naturalnie związane z uczeniem ze wzmocnieniem, a raczej uczenie ze wzmocnieniem stanowi integralny element zachowania obserwowanego w takich badaniach. Uczenie ze wzmocnieniem jest zjawiskiem powszechnym. Nawet aktywność ludzi w mediach społecznościowych podlega opisowi jakościowo i ilościowo zgodnemu z zasadami uczenia się ze wzmocnieniem (Lindström i in., 2021). W porównaniu do badań w modelu zwierzęcym istotną różnicę w testowaniu pacjentów czy wolontariuszy stanowi sposób wzmocnienia, który polega nie na dostarczeniu nagrody pokarmowej po każdym odpowiednim wyborze tylko na uzyskaniu informacji o wyniku działania, a więc nagrodzie wtórnej, mentalnej reprezentacji nagrody. W przypadku gryzoni stosuje w badaniach klatki do warunkowania instrumentalnego, z dźwigniami, portami do wsadzenia pyszczka, dotykowym ekranem wyświetlającym wskazówki wzrokowe lub labirynty (Izquierdo i in., 2017). Dodatkowo, stosowany u ludzi, ale możliwy też u gryzoni, schemat probabilistyczny, w którym opcje będą związane z zadaniem prawdopodobieństwem uzyskania nagrody (np. 80% i 20%) obniża tempo uczenia i redukuje wykorzystanie prostych strategii (Izquierdo i in., 2017). W niniejszych badaniach wykorzystano test uczenia z probabilistycznym odwróceniem zaadaptowany do dwóch sposobów testowania: z wykorzystaniem klatki IntelliCage oraz w sposób klasyczny z wykorzystaniem aparatu do warunkowania instrumentalnego.

Wyzwaniem pozostaje określenie neuronalnego podłoża oraz zgodności testu uczenia z odwróceniem z innymi testami elastyczności poznawczej (Cools, 2015; Nikiforuk, 2016; Izquierdo i in., 2017). Wykonanie zadania z odwróceniem wymaga zdolności do uczenia się na podstawie otrzymanych nagród (oraz ich braku lub pominięcia) poprzez wybór opcji, szacowania prawdopodobieństwa wystąpienia odwrócenia oraz zrozumienia założeń zadania lub przestrzeni możliwości, za które mogą odpowiadać osobne obwody neuronalne (Clarke i Roberts, 2011; Izquierdo i in., 2017). Często wykorzystywany test przełączania uwagi czy inaczej test

przerzutności między modalnościami zmysłowymi (ang. *attentional set shifting*) polega na zdolności do uczenia się z odwróceniem ale także na zdolności do przeniesienia uwagi z jednej cechy bodźca na inną (Cools, 2015; Nikiforuk, 2016). W początkowych etapach testu przełączania uwagi u gryzoni nagrodę przewiduje np. jeden z zaproponowanych zapachów natomiast po deterministycznych odwróceniach, w kluczowym etapie testu, nagrodę nieoczekiwanie zaczyna reprezentować inny aspekt/cecha bodźca np. rodzaj materiału pod którym znajdować się będzie porcja pożywienia (Birrell i Brown, 2000). Wskaźnikiem elastyczności poznawczej będzie wtedy względna liczba prób do zadanego kryterium lub liczba wykonanych przez zwierzę błędów.

Pomimo, że zadanie z probabilistycznym odwróceniem służy głównie określeniu zmian w elastyczności poznawczej, to umożliwia także wnioskowanie o zakresie wrażliwości na informację zwrotną. Przykładowo w badaniu Michaela Franka i współpracowników, zastosowano zadanie rozróżnienia wizualnego między dwoma alternatywami, polegające na zdolności do wyboru opcji związanej z wyższym prawdopodobieństwem uzyskania nagrody lub unikania wyboru opcji o mniejszej pewności nagrody (Frank i in., 2004). Zarówno w wersji probabilistycznej jak i deterministycznej tego zadania pacjenci cierpiący na chorobę Parkinsona, z odstawionym lekiem, opierali się znamienne bardziej na unikaniu wyboru gorszej opcji, czyli dokonywali wyboru na podstawie głównie negatywnej informacji zwrotnej, natomiast pacjenci pod wpływem leku na podstawie pozytywnej informacji zwrotnej (Frank i in., 2004). Ten wynik nawiązuje bezpośrednio do neuronalnego podłoża procesu uczenia się ze wzmocnieniem, ponieważ z przetwarzaniem błędu przewidywania związana jest proporcjonalna zmiana aktywności komórek dopaminowych, natomiast poziom dopaminy jest obniżony w chorobie Parkinsona. Nowsze badania, autorstwa Johna Grogana i współpracowników, z użyciem zarówno zmodyfikowanej jak i oryginalnej wersji tego zadania, nie potwierdziły jednak wskazanego wyniku (Grogan i in., 2017). Sposobem na określenie wrażliwości na informację zwrotną może być także analiza sekwencji wyborów z zastosowaniem rozwiązań modelowania obliczeniowego. W materiałach dodatkowych publikacji Johna Grogana i współpracowników, w celu analizy wrażliwości na informację zwrotną, zastosowano model określony w niniejszej rozprawie jako „wartościowo zależny” (Tabela 2–1). Na tej podstawie określono, że waga zdarzenia pozytywnego (por. α_+ , Tabela 2–1) jest większa od wielkości kroku dla zdarzenia niekorzystnego (por. α_- , Tabela 2–1) (Grogan i in., 2017). Niestety, nie porównano wartości parametrów zależnie od wpływu leku i badanej grupy, dlatego interpretacja wyników modelowania w zależności od stanu chorobowego nie była możliwa. Natomiast w podobnym badaniu, Robba Rutledge’a i wsp., porównano wagę zdarzenia pozytywnego (α_+) i negatywnego (α_-) grupy pacjentów z włączonym leczeniem, bez niego oraz dobraną według wieku grupę kontrolną. Jedyna określona różnica dotyczyła wagi zdarzenia pozytywnego, dla pacjentów cierpiących na chorobę Parkinsona z włączonym leczeniem, która była wyższa w porównaniu do pozostałych grup (Rutledge i in., 2009). Chciałabym tutaj skomentować, że wyników badania Robba Rutledge’a i wsp. nie można bezpośrednio porównać z innymi wynikami modelowania, ponieważ zaprezentowane wartości wielkości kroku (α), możliwe do interpretacji także jako wrażliwość na infor-

mację zwrotną, przekroczyły w niektórych przypadkach 1, czyli wartość maksymalną przyjmowaną w standardowym użyciu modelowania na podstawie *Q-learning*, co nie zostało wytłumaczone przez autorów. Niemniej, wyniki należy interpretować w oparciu o charakterystykę choroby. W chorobie Parkinsona zmiany poziomu dopaminy dotyczą głównie szlaku nigrostriatalnego i obszarów korowych, natomiast leczenie oparte o zwiększenie poziomu dopaminy wpływa także na względnie nieuszkodzony obwód łączący VTA i brzuszne prążkowie, zwiększając także w nim poziom dopaminy. Z tego powodu włączenie leczenia może obniżać sprawność pacjentów w badaniach dotyczących procesów motywacyjnych, nagrody i uczenia się, co natomiast nie musi być prawdziwe dla badań elastyczności poznawczej. Na tej podstawie widoczne jest, że wzorzec zmian aktywności mózgu pod wpływem choroby i leków celujących w układu dopaminowy może zależeć od rodzaju testu, a raczej badanego aspektu funkcji poznawczych (Meder i in., 2019).

Wykorzystanie zadania z probabilistycznym odwróceniem oferuje możliwość wnioskowania o zmianach w funkcjach wykonawczych i elastyczności poznawczej. Włączenie modelowania obliczeniowego może natomiast rozszerzać możliwości badawcze czyli dostarczać narzędzi do badania zagadnień rozwijającej się psychiatrii obliczeniowej. Stosowanie modeli obliczeniowych pozwala poza analizą ilościową wrażliwości na informację zwrotną także m.in. na badanie nieoczywistych cech zachowania poprzez wprowadzenie nowych zmiennych. Śledzenie zmian wartości parametrów w zależności od leku lub stanu chorobowego może dostarczyć wiedzy o zmianach leżących u podstaw zmian chorobowych i objawów związanych z przyjmowanymi lekami. Z tego powodu modelowanie obliczeniowe coraz częściej włącza się do badań biomedycznych.

2.4 Neuronalne podłoże uczenia ze wzmocnieniem

W kontekście neurobiologicznych podstaw działania nagrody, kluczowym odkryciem było wskazanie wzorca zmian aktywności komórek dopaminowych, opisane w pracach Wolframa Schulza i współpracowników (W Schultz i in., 1997; Hollerman i Schultz, 1998; W. Schultz, 2017). W pracy Jeffrey'a Hollermana i Wolframa Schulza opublikowanej w 1998 opisano wyniki wykonanego przez makaki testu, polegającego na serii wyborów jednego z pary wyświetlanych obrazków, którego skutkiem było otrzymanie nagrody w postaci soku (Hollerman i Schultz, 1998). Zadaniu towarzyszył pomiar elektrofizjologiczny aktywności neuronów części zbitiej istoty czarnej i brzuszego pola nakrywki. Zwiększenie aktywności neuronów dopaminowych (rozpoznanych po właściwościach elektrofizjologicznych) tych obszarów widoczne było w czasie przyznania nagrody, jednak w trakcie uczenia się malało oraz przesunęło się z czasu przyznania nagrody do czasu najwcześniejszej wskazówki przewidującej nagrodę. Drugą ważną obserwacją było obniżenie mierzonej aktywności, od poziomu podstawowego, w czasie oczekiwania na nagrodę w sytuacji jej braku, lub gdy czas do jej przyznania zo-

stał wydłużony. Neurony reagowały zawsze w sytuacji nieoczekiwanej nagrody i w próbach gdy latencja do przyznania nagrody została skrócona względem oczekiwanej. Badacze zinterpretowali wyniki jako proporcjonalny związek aktywności neuronów dopaminowych z błędem przewidywania, czyli rozbieżnością między oczekiwaną i uzyskaną nagrodą, zarówno w kontekście pozytywnym (nagroda została przyznana pomimo niepewności) oraz negatywnym (nagroda nie została przyznana pomimo oczekiwania) (W Schultz i in., 1997; Hollerman i Schultz, 1998; W. Schultz, 2017).

W zadaniu zastosowanym w wspomnianej pracy Jeffrey'a Hollermana i Wolframa Schultza przypisanie uzyskania nagrody do obrazka było deterministyczne. Natomiast związek wskazówki z szansą uzyskania nagrody, a więc wprowadzenie probabilistycznej natury testu, w eksperymencie opisanym w pracy Christophera Fiorillo, Philippe'a N. Toblera i Wolframa Schultza dostarczyło głębszego zrozumienia tego zjawiska i potwierdziło przypuszczenia zawarte w pracy wspomnianej powyżej (Fiorillo i in., 2003). Prezentowane makakom wskazówki wizualne były związane z prawdopodobieństwem otrzymania soku na poziomie 0, 0.25, 0.5, 0.75 i 1, po 2-sekundowym wyświetleniu wskazówki. Przeprowadzony pomiar elektrofizjologiczny neuronów dopaminowych brzuszego śródmózgowia wskazał, zgodnie z oczekiwaniami, że wielkość zmiany aktywności zależy liniowo odwrotnie proporcjonalnie do prawdopodobieństwem uzyskania nagrody w czasie prezentacji wskazówki oraz liniowo i proporcjonalnie w czasie oczekiwanego dostarczenia nagrody. Wskazuje to, że im pewniejsze było uzyskanie nagrody w próbie tym większa będzie odpowiedź neuronalna w czasie wskazówki i mniejsza w czasie uzyskania nagrody. Dodatkowo wykryto także inny wzorec aktywności w postaci stopniowego wzrostu aktywności między czasem wskazówki i oczekiwanej nagrody, w sytuacji gdy rezultat wyboru nie był pewien (określony jako „rampa”). Ten wzrost był największy dla wskazówki związanej z 0,5 szansą uzyskania nagrody, mniejszy dla wartości 0,25 i 0,75 oraz niezauważalny gdy rezultat wyboru był pewien. Wartość tej zmiany dotyczyła niepewności wyniku działania i tym samym informatywności w procesie uczenia się. Nie wykazano, żeby za dwa powyższe wzorce aktywności odpowiadały osobne subpopulacje komórek (Fiorillo i in., 2003). Według autorów związek zmiany aktywności badanego obwodu dopaminowego z niepewnością wyniku działania może tłumaczyć promowanie ryzykownego zachowania i apetytywnej właściwości niepewności w grach hazardowych (Fiorillo i in., 2003). Aktualnym problemem pozostaje wyjaśnienie w jaki sposób zmiany aktywności komórek dopaminowych, wydzielania dopaminy, oraz zróżnicowania obwodów neuronalnych związanych z układem dopaminowym, prowadzą do jednoczesnego zaangażowania dopaminy zarówno w uczenie się (błąd przewidywania) jak i procesy motywacyjne (wartość nagrody) (Berke, 2018; Mohebi i in., 2019).

Dalsze badania różnych zespołów wykazały, że z uczeniem z odwróceniem związane są układy: dopaminowy, serotoninowy i glutaminianowy (Izquierdo i in., 2017) oraz szczególne znaczenie mają połączenia kory okołooczdolowej z prążkowiem i jądrem migdałowatym (Kehagia i in., 2010; Clarke i Roberts, 2011). Przykładowo, w badaniach den Ouden i wsp. nad rolą wskazanych neuroprzekazników zostało wykorzystane zadanie z probabilistycznym od-

wróceniem (ang. *probabilistic reversal learning task*) (den Ouden i in., 2013). Dla grupy ponad 680 uczestników określono zależność polimorfizmu genu kodującego transporter serotoniny oraz genu kodującego transporter dopaminy i częstości powtórzenia wyboru po uzyskaniu nagrody (ang. *win-stay*), zmiany po przegranej (ang. *lose-shift*) oraz perseweracji, rozumianej jako przynajmniej dwukrotne powtórzenie niekorzystnego wyboru. Zaobserwowano, że polimorfizm genu transportera serotoniny był związany z częstością działania zmiany po przegranej, rozumianej także jako wrażliwość na negatywną informację zwrotną, polimorfizm genu kodującego transporter dopaminy z zwiększoną częstością błędów perseweracji w fazie odwrócenia, natomiast nie wpłynął ani na częstość powtórzenia wyboru po nagrodzie ani zmiany po przegranej (den Ouden i in., 2013). Wskazane wyniki wskazują na znaczenie serotoniny i dopaminy w procesie podejmowania decyzji, chociaż nie stanowią wyczerpującej odpowiedzi na pytanie o ich jednoznaczną rolę.

Szczególne znaczenie dla przetwarzania informacji o nagrodzie ma układ serotoninowy. Badania tego zagadnienia skoncentrowały się na neuronach zawierających serotoninę (5-hydroksytryptaminę, 5-HT) w jądrach szwu w śródmózgowiu. Przy interpretacji wyników badań należy wziąć pod uwagę, że jądra szwu stanowią strukturę heterogenną pod względem komórkowym, zawierając zarówno neurony serotoninowe jak i dopaminowe i GABAergiczne (Pollak Dorocic i in., 2014; Liu i in., 2020). Dlatego, dla określenia roli wybranej populacji, niezbędne jest zastosowanie odpowiednich metod chemogenetycznych, optogenetycznych, czy innych, z zakresu inżynierii genetycznej. W tym celu można posłużyć się znacznikiem molekularnym komórek serotoninowych takim jak enzym szlaku przetwarzania serotoniny hydroksylaza tryptofanu 2 (Tph2) czy też transporter serotoniny (SERT). Neurony serotoninowe jąder szwu otrzymują wejścia synaptyczne m.in. z kory przedczołowej, pnia mózgu, bocznej uzdeczki (połączenie jąder podstawnych z układem nagrody), ciała migdałowatego (w tym jądra łożyskowego prążka krańcowego związanego z zachowaniem lękowym), jąder podstawnych (prążkowie, gałki bladej, istoty czarnej i brzuszego pola nakrywki) i podwzgórza (związanego z regulacją homeostatyczną, w tym pobieraniem pokarmu, temperaturą ciała oraz cyklem snu i czuwania) (Pollak Dorocic i in., 2014). Główne włókno serotoninowe biegnie od jąder szwu do opuszki węchowej z rozgałęzieniami docierającymi do niemal całego mózgu (Awasthi i in., 2021). Największą gęstością charakteryzują się projekcje serotoninowe dochodzące do podwzgórza, wzgórza, ciała migdałowatego, jąder podstawy i kory mózgowej (Awasthi i in., 2021).

Jedną z pierwszych koncepcji dotyczącej roli serotoniny, nazywana hipotezą opozycji wobec dopaminy, stanowiła, że ze względu na asymetryczną zmianę aktywności neuronów dopaminowych względem pozytywnego i negatywnego błędu przewidywania, za przetwarzanie negatywnego błędu przewidywania mógłby być odpowiedzialny w większej mierze układ serotoninowy, ze względu na związek z przetwarzaniem wrażeń awersyjnych (Daw i in., 2002; Glimcher, 2011).

Natomiast bardziej złożoną zależność aktywności układu serotoninowego i wzmocnienia wskazują m.in. badania zespołu Zacharego Mainen'a. Przykładowo, w pracy Sary Matias i wsp.

wskazano, że neurony serotoninowe są odpowiedzialne za hamowanie perseweracji, rozumianej jako uporczywe nieadaptacyjne powtarzanie działania mimo rezultatu gorszego od oczekiwanego (Matias i in., 2017). Udowodniono to poprzez obserwację, że chemogenetyczne wyłączenie neuronów serotoninowych, w grzbietowych jądrach szwu, u myszy, doprowadziło do mniejszego spadku tempa antycypacyjnych liźnięć podajnika wody w sytuacji gdy bodziec warunkowy nieoczekiwanie przestał przewidywać podanie napoju (Matias i in., 2017). Upraszczając, myszy nadal intensywnie wykonywały próby uzyskania nagrody pomimo jej braku. Tutaj chciałabym skomentować potencjalną niejednoznaczność w interpretacjach wyników badań w odniesieniu do perseweracji. Znaczenie perseweracji określa się w kontekście adaptacji do wykonania zadania, dlatego częste wykonywanie powtórzeń danego działania może być związane z małą elastycznością poznawczą lub przeciwnie ze zdolnością do skupienia uwagi, a więc mniejszą impulsywnością. Wracając do pracy Sary Matias i wsp., w kolejnym etapie opisywanych badań zastosowano schemat uczenia się z odwróceniem, w którym przypisanie bodźca warunkowego (zapachu) i bezwarunkowego (uzyskania wody, jej braku lub awersyjnego poddmuchu powietrza) było zmieniane na dokładnie przeciwne, np. duża objętość wody na podmuch powietrza. Zmiany aktywności neuronów serotoninowych, w jądrach szwu, oraz dopaminowych, z części zbitej istoty czarnej i brzusznyemu polu nakrywki, mierzone jako zmiany wewnątrzkomórkowego stężenia wapnia z wykorzystaniem sensorów fluorescencyjnych, różniły się w tym teście (Matias i in., 2017). Neurony serotoninowe podobnie jak dopaminowe zwiększyły aktywność w sytuacji rezultatu lepszego od oczekiwanego (pozytywny błąd przewidywania), jednak w przeciwieństwie do dopaminowych wykazały wzrost aktywności także w sytuacji rezultatu gorszego od oczekiwanego (negatywny błąd przewidywania) oraz nie wykazały odpowiedzi na awersyjny podmuch powietrza, podczas gdy neurony dopaminowe obniżyły wtedy swoją aktywność. Chociaż kwestia przetwarzania bodźca awersyjnego wymaga dalszych badań to wskazaną zmianę aktywności serotoninowej można interpretować jako sygnał nieoczekiwanego zdarzenia. Autorzy wskazują także na wolniejsze tempo adaptacji, po odwróceniu, dla zmiany aktywności neuronów serotoninowych. Wskazane wolniejsze tempo adaptacji może tłumaczyć rolę serotoniny w hamowaniu zachowania perseweracyjnego (Matias i in., 2017). W badaniach Madaleny Fonseci i wsp. optogenetyczna stymulacja neuronów serotoninowych, w grzbietowych jądrach szwu, wydłużyła czas jaki mysz była w stanie czekać na bodziec warunkowy, mimo że sama stymulacja nie miała efektu apetytywnego ani awersyjnego (Fonseca i in., 2015). Wskazuje to dodatkowo, że możliwe wzmacniające efekty stymulacji tej struktury nie są zależne od serotoniny (Fonseca i in., 2015). Dalsze badania wykazały, że optogenetyczna aktywacja komórek serotoninowych, w grzbietowych jądrach szwu, wydłużyła czas oczekiwania wyłącznie gdy nagroda była dostępna z wysokim prawdopodobieństwem (75% względem 50 i 25%) oraz, że czas oczekiwania wzrastał wraz z niepewnością czasu dostarczenia nagrody (Miyazaki i in., 2018). W dyskusji autorzy tych badań odrzucają koncepcję wpływu serotoniny na percepcję czasu czy szybkość dewaluacji nagrody z czasem (ang. *temporal discounting*) tłumacząc uzyskane wyniki jako rolę serotoniny w rozstrzyganiu kompromisu między (negatyw-

nymi) doznaniem sensorycznym a (pozytywnym) subiektywnym przekonaniem (Miyazaki i in., 2018). Co ciekawe, autorzy zasugerowali na tej postawie, że w depresji efekt farmakologiczny może być niewystarczający bez subiektywnej pewności pozytywnych rezultatów oraz cierpliwości w jej nabywaniu, dzięki psychoterapii behawioralno-poznawczej (Miyazaki i in., 2018). Taka interpretacja opiera się jednak na stwierdzeniu, że depresja jest związana lub spowodowana mniejszym stężeniem serotoniny czy zmienioną aktywnością w układzie serotoninowym, czyli serotoninowej hipotezie depresji (ang. *serotonin theory of depression*), która nadal jest przedmiotem sporu. Układ serotoninowy ma znaczenie dla elastyczności poznawczej (Nilsson i in., 2019) i przetwarzania informacji o nagrodzie, w kontekście „lubienia” jako hedonii, „chcenia” jako podtrzymywania motywacji oraz uczenia ze wzmocnieniem (Liu i in., 2020). Natomiast widoczne znaczenie układu serotoninowego w uczeniu ze wzmocnieniem i ogólniej w funkcjonowaniu poznawczym jest złożone i wyraźniejsze dla wybranych aspektów zachowania takich jak kompulsywność (perseweracja) i impulsywność (Phillips i Robbins, 2020), związanych z czasem podjęcia działania.

W badaniach przedstawionych w niniejszej rozprawie określono znaczenie receptorów NMDA (receptorów dla kwasu N-metylo-D-asparaginowego, NMDAR) na komórkach serotoninowych w uczeniu z odwróceniem, poprzez wykorzystanie myszy transgeniczných. NMDAR to jonotropowe receptory glutaminianu charakteryzujące się wysoką przepuszczalnością wapnia oraz zależnym od napięcia blokowaniem przez jony magnezu (Zito i Scheuss, 2009). NMDA stymuluje uwalnianie serotoniny w jądrach szwu, jak wynika z badań *in vitro* (Becquet i in., 1993) oraz *in vivo* (Tao i Auerbach, 2000). NMDAR są wyrażane także w neuronach serotoninowych istoty szarej okołowodociągowej, a ich liczba może ulegać zmianie w chronicznym bólu (Terashima i in., 2012). Wykazano, że AMPA (kwas α -amino-3-hydroksy-5-metylo-4-izoksazolopropionowy) i NMDA zwiększają częstotliwość wyładowań neuronów serotoninowych w jądrach szwu oraz zasugerowano, że NMDA aktywuje lokalne neurony GABAergiczne, które hamują neurony serotoninowe poprzez receptory GABA_A i GABA_B (Gartside i in., 2007). Efekty behawioralne zablokowania NMDAR wybiórczo na komórkach serotoninowych, czyli rola NMDA-zależnej plastyczności w komórkach serotoninowych, jest słabo poznana. W linii myszy testowanej w niniejszej rozprawie, indukowanie mutacji w wieku dorosłym powoduje brak funkcjonalnych NMDAR na komórkach zawierających hydroksylazę tryptofanu 2, poprzez brak podjednostki NR1. Nie opisano wcześniej efektu takiej modyfikacji.

Co ważne, pod względem potencjalnych interwencji farmakologicznych, efekt związany z serotoniną jest zależny od rodzaju receptorów serotoninowych. Receptory 5-HT dzielą się na jonotropowe (5-HT₃) i metabotropowe (5-HT₁, 5-HT₂, 5-HT₄, 5-HT₅, 5-HT₆, 5-HT₇), z czego receptor 5-HT₇ został opisany jako ostatni i jego znaczenie dla elastyczności poznawczej nie zostało wyczerpująco zbadane. Receptory 5-HT₇ w ośrodkowym układzie nerwowym znajdują się we wzgórzu, podwzgórzu, hipokampie, korze mózgowej, jądrach migdałowych i jądrach szwu oraz komórkach Purkiniego w mózgdzku, z kolei w obwodowym w mięśniach gładkich naczyń krwionośnych i przewodzie pokarmowym (Gellynck i in., 2013). Dostępnych jest kilka

selektywnych antagonistów (SB-258719, SB-269970, SB- 656104 i JNJ-18038683) i agonistów (AS-19, LP-44, LP-12, LP-211 i E-55888) oraz nieselektywny antagonistę (DR-4004) i agonistę (5-HT_{1A/7}; 8-OH-DPAT) tego receptora (Nikiforuk, 2015; Hoyer, 2019). Selektywna blokada receptorów NMDA działa antydepresyjnie i w tym kontekście wskazuje się wpływ na czas snu REM (Bonaventure i in., 2012; Nikiforuk, 2015). Jednak w badaniach klinicznych JNJ-18038683, u osób cierpiących na depresję, uzyskane wyniki okazały się niejednoznaczne, ze względu na brak efektywności przewyższającej kontrolę placebo, zarówno dla tego związku jak i escitalopramu (Bonaventure i in., 2012). Cel dwóch innych badań klinicznych dla tego związku stanowi leczenie zaburzeń poznawczych towarzyszących chorobie afektywnej dwubiegunowej (NCT02466685 oraz NCT03633357; clinicaltrials.gov). Podobnie, znaczenie w lękowości, aktywność napadowej (padaczkowej), bólu neuropatycznym, potencjalnym leczeniu migreny oraz zdolnościach poznawczych wydaje się być bardziej złożony (Nikiforuk, 2015). Zaproponowano, że prokognitywny efekt blokady receptora 5-HT₇ może być widoczny przy zwiększeniu złożoności zadania i w negatywnych objawach schizofrenii (Nikiforuk, 2015). W przypadku uczenia się i pamięci, badania dostarczają niejednoznacznych wyników, zależnych od testu behawioralnego, zastosowanych substancji, dawek, czasu i dróg ich podania oraz rodzaju modelu zwierzęcego (Roberts i Hedlund, 2012).

Badania z wykorzystaniem ligandów receptora 5-HT₇ wskazały na znaczenie tego receptora w procesach poznawczych. Podanie SB-269970 (w dawce 0,3 i 1 mg/kg) odwraca zaburzenie elastyczności poznawczej, spowodowane chronicznym stresem unieruchomienia, u szczurów, co określono w teście przerzutności między modalnościami zmysłowymi (ang. *extra-dimensional set-shifting test*) (Nikiforuk, 2012). Także podanie tego związku w wyższej dawce (1 mg/kg) i łącznie z nieaktywną dawką escitalopramu (oba w dawce 0,3 mg/kg) skutkowało polepszonym wynikiem testu, u zwierząt niepoddanych procedurze wywołującej stres (Nikiforuk, 2012). W innych badaniach sprawdzono czy ligandy receptorów serotoninowych osłabiają deficyty poznawcze występujące w schizofrenii. Podawanie zwierzętom niekonkurencyjnego antagonisty receptora NMDA, jak fencyklidyna czy ketamina, naśladuje zaburzenia poznawcze występujące w tej chorobie (Neill i in., 2010; Nikiforuk i in., 2013). W badaniach Lakshmi Rajagopal i wsp. myszy C57BL/6J wykonywały test uczenia z odwróceniem, w którym określano odsetek poprawnych odpowiedzi, czyli naciśnięć odpowiedniej dźwigni w klatce do warunkowania instrumentalnego. Opisano, że podanie fencyklidyny (10 mg/kg) spowodowało deficyty w odsetku poprawnych odpowiedzi. Natomiast lurazydon (atypowy lek przeciwpsychotyczny z wysokim powinowactwem do receptora 5-HT₇, w dawce 3 mg/kg), tandospiron (częściowy agonista receptora 5-HT_{1A}, 5 mg/kg) i SB269970 (antagonista receptora 5-HT₇, 4 mg/kg), ale nie AS19 (agonista receptora 5-HT₇, 10mg/kg), częściowo odwrócił deficyt spowodowany fencyklidyną (Rajagopal i in., 2016). Łączne podanie WAY100635 (selektywny antagonistę receptora 5-HT_{1A}, 0.6 mg/kg) i lurazydonu (3 mg/kg) zapobiegło polepszeniu wykonania zadania spodziewanego dla lurazydonu (Rajagopal i in., 2016). Podobnie, w badaniach Agnieszki Nikiforuk i współpracowników podanie ketaminy (10mg/kg) pogorszyło wykonanie

testu przerzutności między modalnościami zmysłowymi, u szczurów, jednak SB-269970 (1 mg/kg), a także amisulpryd (lek antypsychotyczny z wysokim powinowactwem do receptora 5-HT₇, 3 mg/kg) redukowały ten efekt (Nikiforuk i in., 2013). Wyniki przedstawionych badań wskazują prokognitywny efekt antagonistów receptora 5-HT₇.

W niniejszych badaniach test z probabilistycznym odwróceniem został wykonany przez genetycznie modyfikowane myszy pozbawione tego receptora linii *Htr7*^{tm1Sut} (Hedlund i in., 2003). Badania twórców linii *Htr7*^{tm1Sut} oraz współpracowników, wskazują, że u tych zwierząt nie obserwuje się zmian w aktywności lokomotorycznej (test otwartego pola, test pręta obrotowego Rotarod), odczuwaniu bólu (test odsunięcia ogona, ang. *tail flick analgesia test*), uczeniu przestrzennym (labirynt Barnes, ang. *Barnes maze test*), ostrości widzenia (test złudzenia optycznego klifu, ang. *visual cliff test*), lękowości (jasnego/ciemnego pudełka, ang. *light/dark transfer test*) ani zdolności do uczenia ze wzmocnieniem (instrumentalnego samopodawania pokarmu, ang. *operant food conditioning test*) (Roberts i in., 2004). Zasugerowano natomiast, że zwierzęta pozbawione receptorów 5-HT₇ wykazują rzadszą reakcję zamierania (ang. *freezing*) w teście warunkowania lękowego (Roberts i in., 2004). Dalsze badania znaczenia tego receptora w lękowości dostarczyły jednak niejednoznacznych wyników. Działanie przeciwlękowe zaobserwowano w teście zakopywania kulek (ang. *marble burying test*), natomiast nie w dwóch innych modelach behawioralnych zaburzenia obsesyjno-kompulsywnego (Hedlund i Sutcliffe, 2007). Według autorów, receptor 5-HT₇ stanowi mediator hipotermii zależnej od serotoniny, a więc jest czynnikiem istotnym dla termoregulacji (Hedlund i in., 2003). Jednak w warunkach naturalnych zwierzęta utrzymują normalną temperaturę ciała (Hedlund i in., 2003). Brak receptora 5-HT₇ u wskazanej linii zwierząt wiąże się z działaniem antydepresyjnym, co wykazano w teście wymuszonego pływania (ang. *forced swim test*) oraz teście zawieszenia za ogon (ang. *tail suspension test*) (Hedlund i in., 2005). Wskazano zmiany we wzorcu snu dla opisywanej linii zwierząt, a dokładniej zaobserwowano rzadsze, ale dłuższe epizody snu REM (ang. *rapid eye movement*) oraz zmniejszenie czasu spędzanego w śnie REM, pomimo braku różnic w dobowym wzorcu snu, długości czasu czuwania i snu wolnofalowego, co według autorów może wyjaśniać obserwowany efekt antydepresyjny (Hedlund i in., 2005). W niniejszej rozprawie analizowano wzorzec podejmowania decyzji dla tej linii zwierząt.

Zgodnie z przedstawionymi koncepcjami roli serotoniny w procesowaniu informacji o nagrodzie, w tym o uczeniu związanym nagrodą dotyczącymi m.in. hamowania behawioralnego, kontroli impulsywności, hipotezy opozycji do dopaminy, kodowania kary czy zwiększonej cierpliwości możliwa jest także rola serotoniny w generowaniu przyjemności, „lubienia”, wzmacnianiu motywacji obserwowanej behawioralnie czy związek z procesami pamięciowymi związanymi z emocjami (Liu i in., 2020). Ze względu na niejednoznaczność doniesień uzasadnione jest prowadzenie badań nad poszczególnymi elementami układu serotoninowego. Do badań opisanych w niniejszej rozprawie włączono dwa zwierzęce modele transgeniczne: linię myszy NR1^{Tph2CreERT2}, dla których w wieku dorosłym zaindukowano blokadę NMDA-zależnej

transmisji glutamatergicznej do komórek serotoninowych oraz zwierzęta pozbawione funkcjonalnych receptorów 5-HT₇.

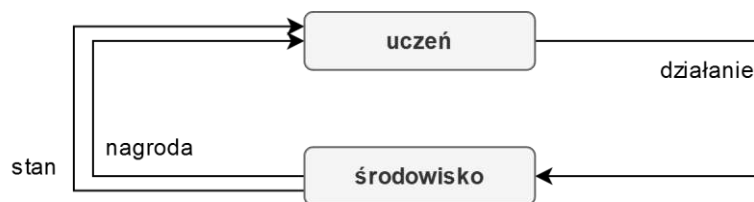
2.5 Modelowanie obliczeniowe uczenia ze wzmocnieniem

Do badań nad procesami podejmowania decyzji coraz częściej stosuje się analizę z wykorzystaniem modeli obliczeniowych, zarówno dla danych pochodzących z neuroobrazowania, elektrofizjologicznego pomiaru aktywności komórek tkanki nerwowej jak i testów behawioralnych. Wykorzystane w ten sposób modele obliczeniowe reprezentują szczegółowo ilościowe hipotezy dotyczące sposobu rozwiązywania problemu, które można poddać bezpośredniemu testowi eksperymentalnemu (Daw, 2011). Analiza z wykorzystaniem modeli polega na zrozumieniu obliczeń leżących u podłoża obserwowanego zachowania (procesu, czy aktywności obwodów układu nerwowego), a następnie wskazanie zmian zależnych od badanych czynników, np. zmian patofizjologicznych. Dlatego znajdują zastosowanie w rozwijającej się psychiatrii obliczeniowej (ang. *computational psychiatry*) (Maia i Frank, 2011).

Rozwój badań nad uczeniem ze wzmocnieniem rozpoczął się niezależnie w latach 50-tych w dziedzinie psychologii jako studium uczenia się u zwierząt, później wznowione w kontekście sztucznej inteligencji w latach 80-tych oraz w dziedzinie inżynierii dla rozwiązania problemu optymalnej kontroli poprzez programowanie dynamiczne, chociaż w tym kontekście nie używano określenia „uczenie” (Sutton i Barto, 1998). Interdyscyplinarna integracja dostępnych rozwiązań, w tym dodatkowo uczenia opartego o różnice czasowe (ang. *temporal-difference learning, TD*), nastąpiła w latach 80-tych (Sutton i Barto, 1998). Omówienie wspomnianych technik dostępne jest w innych opracowaniach (Sutton i Barto, 1998; Niv i in., 2005). Dla obserwowanego u zwierząt uczenia ze wzmocnieniem za podstawową zasadę uważa się wspomniane wcześniej prawo efektu zaproponowane przez Edwarda Thorndike’a (Thorndike, 1927; Sutton i Barto, 1998).

W ogólnym ujęciu uczenie ze wzmocnieniem jako forma uczenia maszynowego, odwzorująca także proces uczenia u zwierząt, polega na dynamicznych interakcjach ucznia ze środowiskiem w dyskretnych próbach. Celem ucznia jest przyjęcie strategii działania, która maksymalizuje sumę nagród do uzyskania od środowiska. Dzieje się to poprzez podejmowanie działania zgodnego z przyjętą strategią decyzyjną, po której uczeń otrzymuje od środowiska wzmocnienie. Otrzymane wzmocnienie, jako jedyna dostępna informacja, służy uczniowi do oceny jakości działania. Wykonanie akcji zmienia stan środowiska. Do istotnych cech środowiska poza dostarczaniem nagród oraz zmienianiem stanu w wyniku akcji ucznia należy także to, że generowane wzmocnienia i stany mogą być stochastyczne, a rozkłady prawdopodobieństwa nie są znane przez ucznia i niemożliwe do kontrolowania przez niego. Odróżnia to ucznia i środowisko. Uczeń wpływa na własne mechanizmy działania ale nie na mechanizmy działania środowiska. Ogólny zarys problemu uczenia ze wzmocnieniem składa się więc z sygnału reprezen-

tującego wybory wykonane przez ucznia (działanie), podstawę tych wyborów (stan) oraz definicję celu ucznia (nagroda) (Sutton i Barto, 1998) (Rycina 2-1).



Rycina 2-1 Interakcja między uczniem i środowiskiem w uczeniu ze wzmocnieniem (Sutton i Barto, 1998).

Najlepiej prześledzić to na przykładzie. Jeden z najpopularniejszych algorytmów uczenia ze wzmocnieniem został opisany przez Richarda Suttona i Andrew Barto jako rozwiązanie problemu „wielorękiego bandyty”. Określenie n-ręki bandyta odnosi się do gry hazardowej, w której pociągnięcie za którąkolwiek z n dźwigni skutkuje uzyskaniem nagrody lub nie, w sposób probabilistyczny. Z perspektywy gracza oznacza to, że w pojedynczej próbie może wykonać tylko jedno działanie, wybrać jedną z alternatywnych opcji. Po wykonaniu działania uzyskuje informacje o rezultacie działania, jednak bez wskazówki o tym, czy jego wybór był poprawny, lepszy, lub gorszy od pozostałych opcji. Graczowi zależy na uzyskaniu jak największej nagrody czy sumy nagród, w czym pomocne jest rozpoznanie, która opcja związana jest z wyższą szansą nagrody. Podczas decyzji pojawia się napięcie między eksplorowaniem dostępnych opcji i wykorzystywaniem tylko najlepszej z nich. Zakładamy, że gracz w każdej próbie szacuje wartość dostępnych opcji działania. Dlatego kluczowa w tym rozwiązaniu jest wartość oczekiwanej nagrody skojarzona z pojedynczym działaniem, krócej wartość działania (1), czyli

$$q(a) = E[R | A = a] \quad 1$$

gdzie R oznacza rezultat działania a ze zbioru A . Naturalnym sposobem określenia wartości działania jest uśrednienie liczby nagród uzyskanych dzięki niemu (2),

$$Q_{n+1}(a) = \frac{1}{n} \sum_{i=1}^n r_i = Q_n + \frac{1}{n} [r_{n+1} - Q_n] \quad 2$$

gdzie wartość $Q \in [0,1]$ określa wartość oczekiwanej nagrody dla działania a zależnie od rezultatu tego działania $r \in [0,1]$ w próbie n . Wzór ten wskazuje w jaki sposób wielkość Q jest zmieniana na podstawie wszystkich poprzednich nagród. Wskazany sposób sformułowania wartości oczekiwanej nagrody pozwala na jej wydajną iteracyjną aktualizację po kolejnych wyborach. Schemat wskazanego algorytmu przedstawia Rycina 2-2.

```

 $Q(a) \leftarrow 0$ 
 $N(a) \leftarrow 0$ 
for  $i \leftarrow 1$  to  $n$ 
     $A \leftarrow \arg \max_a Q(a)$ 
     $R \leftarrow \text{bandit}(A)$ 
     $N(A) \leftarrow N(A) + 1$ 
     $Q(A) \leftarrow Q(A) + \frac{1}{N(A)} [R - Q(A)]$ 

```

Rycina 2-2 Pseudokod prostego algorytmu do rozwiązania problemu n-rękiego bandyty (Sutton i Barto, 1998). Wyrażenie $\arg \max_a$ wskazuje a , dla którego następująca funkcja jest zmaksymalizowana, tj. wybór działania o największej wartości Q . Funkcja przypisująca rezultat R do działania ucznia A („bandit”), odpowiadająca działaniu środowiska, nie jest graczowi znana.

W opisywanym rozwiązaniu wartość oczekiwanej nagrody aktualizowana jest w oparciu o wartość średnią wszystkich poprzednich wyborów. Natomiast w środowisku niestacjonarnym, w którym zasada przyznawania nagród zmienia się, lepszym rozwiązaniem byłoby położenie większej wagi na decyzje bliższe w czasie względem bieżącej. Dlatego wprowadzono stały współczynnik α określający wielkość kroku (3), tak że

$$Q_{n+1} = Q_n + \alpha[R_n - Q_n], \quad \alpha \in \langle 0,1 \rangle \quad 3$$

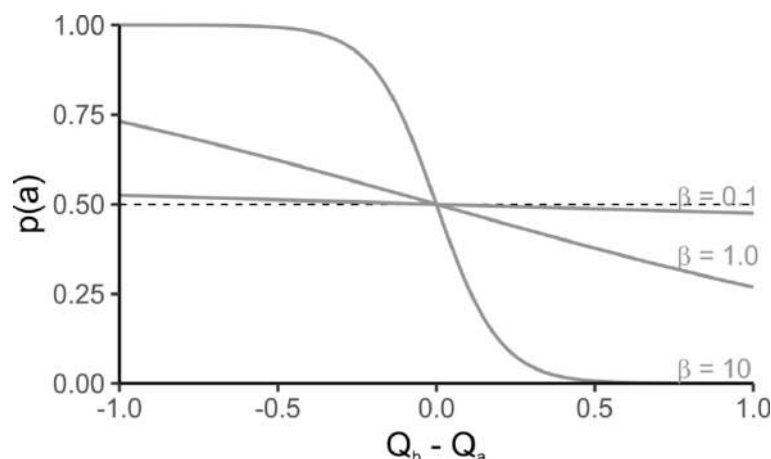
(Sutton i Barto, 1998), gdzie wartość $Q \in [0,1]$ określa wartość oczekiwanej nagrody zależnie od rezultatu tego działania $R \in [0,1]$ w próbie n . Wskazana wartość określana jest w badaniach jako parametr tempa uczenia (ang. *learning rate*), jednak nie odnosi się to bezpośrednio do miary zdolności poznawczych. Tak wyrażony współczynnik określa zakres poprzedzających wyborów, które mają znaczenie dla bieżącej decyzji i stanowi wykładniczą średnią zależącą od pozycji (ang. *exponential, recency-weighted average*) (Sutton i Barto, 1998), co zostało szerzej przedstawione w pracy przeglądowej Paula Glimchera (Glimcher, 2011). Optymalna wartość tego współczynnika, uzyskana przez dopasowanie do danych eksperymentalnych, ma znaczenie dla wnioskowania o zachowaniu. Wartości skrajne 0 i 1 oznaczają brak zdolności do szacowania na podstawie doświadczenia. Wartości zbliżone do 1 wskazują na efekt wyłącznie ostatniej próby, natomiast bliskie 0 na znikomą wagę nowych informacji.

Drugim kluczowym elementem pozostaje zasada wyboru działania. Podstawową zasadę stanowi wybór najlepszej możliwej opcji za każdym razem, inaczej strategia zachłanna. Funkcja *softmax* określa prawdopodobieństwo wyboru działania zależne od wartości oczekiwanej nagrody ze stałym współczynnikiem $\tau \in [0,1]$ nazywanym temperaturą (Sutton i Barto, 1998) (4). Alternatywnie można wykorzystać współczynnik $\beta \in [0, \infty]$, określany jako odwrócona temperatura. W przypadku dwóch możliwych opcji działania wskazaną zależność ilustrują po-

niższe równanie i wykres (Rycina 2-3). W danej chwili dostępna będzie opcja a i opcja alternatywna, oznaczona jako b . Dla każdej z nich określa się wartość oczekiwaną nagrody w bieżącym momencie i na tej podstawie określa prawdopodobieństwo wyboru jednej z nich (4).

$$p(A_n = a) = \frac{1}{e^{\beta(Q_{b,n} - Q_{a,n})} + 1}$$

4



Rycina 2-3 Prawdopodobieństwo wyboru danej opcji w zależności od różnicy w wartości oczekiwanej nagrody między dwoma opcjami i wartości współczynnika β . Prawdopodobieństwo wyboru opcji o większej wartości oczekiwanej nagrody wzrasta proporcjonalnie do różnicy między możliwymi opcjami oraz wartości współczynnika β . Dlatego alternatywnie określa się wartość β jako zdolność do rozróżnienia opcji, czułość na nagrodę czy „zachłanność” ucznia. Oznaczenie Q wskazuje wartość oczekiwaną nagrody dla dostępnych opcji.

Wskazuje to, że im większa różnica między wartością oczekiwaną nagrody dla dostępnych opcji tym pewniejszy staje się wybór działania z większą oczekiwaną nagrodą. Przykładowo, niską wartość współczynnika β można rozumieć także jako decydowanie bliskie losowości (np. (Wilson i Collins, 2019)), efekt przewagi eksploracji nad wykorzystaniem wiedzy (np. (Bai i in., 2014)), czy małą czułość ucznia na różnicę w wartościach oczekiwanej nagrody dla dostępnych opcji. Warto zwrócić uwagę, że w tak sformułowanym problemie nie włączono koncepcji zmiany stanu, która stanowi istotny element algorytmu „Q-learning” opisanego przez Christophera Watkina (Watkins i Dayan, 1992). Przedstawione rozwiązanie można natomiast potraktować jako jego uproszczoną wersję.

W badaniach w obszarze neuronauk stosowano modyfikacje oryginalnego algorytmu. Modele zastosowane w badaniach opisywanych w rozprawie różniły się głównie sposobem liczenia aktualizacji wartości oczekiwanej nagrody dla dostępnych opcji (Tabela 2–1).

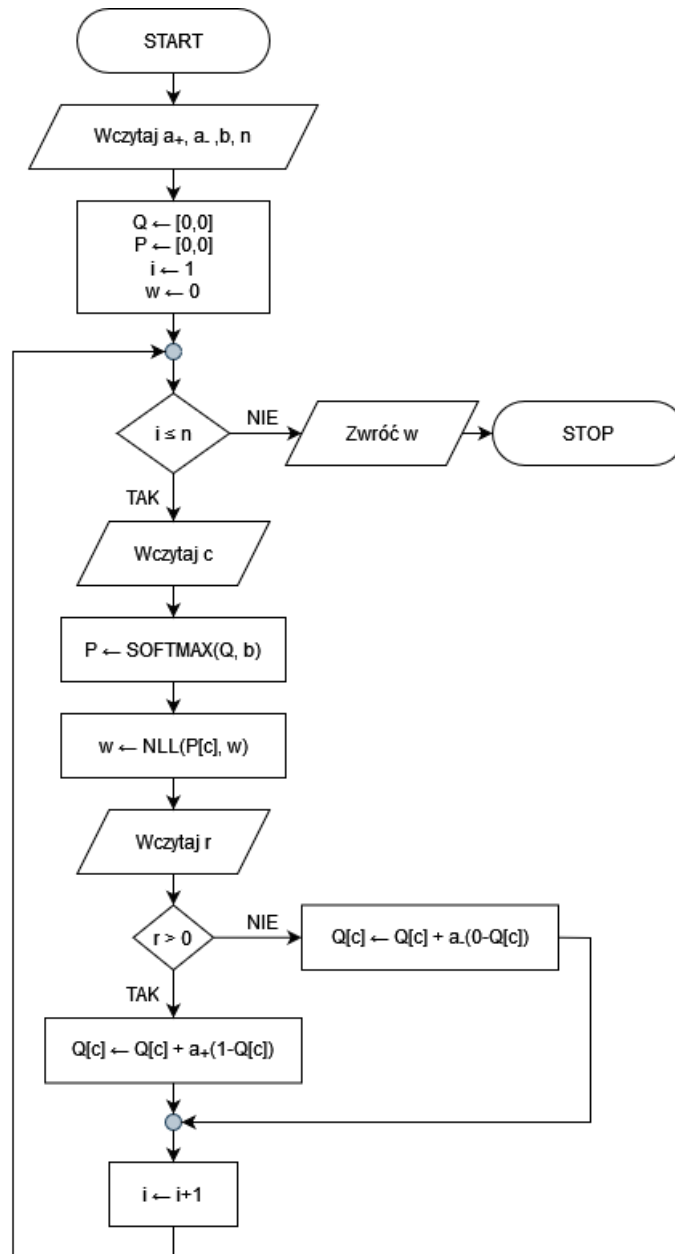
Tabela 2—1 Sposób aktualizacji wartości oczekiwanej nagrody w modelach zastosowanych w niniejszych badaniach

model	rezultat działania			
	nagroda*		brak nagrody*	
	opcja wybrana (a)	opcja alternatywna (b)	opcja wybrana (a)	opcja alternatywna (b)
podstawowy	$Q_a + a(1 - Q_a)^*$	-	$Q_a + a(0 - Q_a)$	-
uwagowy	$\alpha_n = \mu 1 - Q_n + (1 - \mu)\alpha_0$		$\alpha_n = \mu 0 - Q_n + (1 - \mu)\alpha_0$	
	$Q_a + \alpha_n(1 - Q_a)$	-	$Q_a + \alpha_n(0 - Q_a)$	-
wartościozależny	$Q_a + a_+(1 - Q_a)$	-	$Q_a + a_-(0 - Q_a)$	-
urojony	$Q_a + a(1 - Q_a)$	$Q_b + a(0 - Q_b)$	$Q_a + a(0 - Q_a)$	$Q_b + a(1 - Q_b)$
mieszany	$Q_a + a_+(1 - Q_a)$	$Q_b + a_+(0 - Q_b)$	$Q_a + a_-(0 - Q_a)$	$Q_b + a_-(1 - Q_b)$
incydentalny	$Q_a + a_\rho(1 - Q_a)$	$Q_b + a_\phi(1 - Q_b)$	$Q_a + a_\rho(0 - Q_a)$	$Q_b + a_\phi(0 - Q_b)$
podwójnie mieszany	$Q_a + a_{+\rho}(1 - Q_a)$	$Q_b + a_{-\phi}(0 - Q_b)$	$Q_a + a_{-\rho}(0 - Q_a)$	$Q_b + a_{+\phi}(1 - Q_b)$
podwójnie mieszany z rozpadem Q	$e^{-\tau_n S}(Q_a)^{**}$	$e^{-\tau_n S}(Q_b)$	$e^{-\tau_n S}(Q_a)$	$e^{-\tau_n S}(Q_b)$
	$Q_a + a_{+\rho}(1 - Q_a)$	$Q_b + a_{-\phi}(0 - Q_b)$	$Q_a + a_{-\rho}(0 - Q_a)$	$Q_b + a_{+\phi}(1 - Q_b)$
podwójnie mieszany z rozpadem β^{***}	$Q_a + a_{+\rho}(1 - Q_a)$	$Q_b + a_{-\phi}(0 - Q_b)$	$Q_a + a_{-\rho}(0 - Q_a)$	$Q_b + a_{+\phi}(1 - Q_b)$

Ciemniejszy odcień wskazuje dodatkowe obliczenia poprzedzające określenie wartości oczekiwanej nagrody. *Wartości 0 i 1 kodują rezultat działania (uzyskanie dostępu do nagrody-1, brak dostępu-0) i zostały wprowadzone dla przejrzystości.

Oznaczenie τ_n odnosi się do czasu, który minął od poprzedniej próby. *Modyfikacja dotyczy zasady wyboru działania. Sposób aktualizacji wartości oczekiwanej nie różni się od modelu „podwójnie mieszanego”.

Pierwsze często stosowane rozszerzenie stanowi rozdzielenie parametru α , tempa uczenia, w zależności od rezultatu wyboru (ang. *dual learning rate*) (Tabela 2—1). Ten model został nazwany w niniejszej rozprawie jako „wartościozależny”, ze względu na kluczowe znaczenie wartości uzyskanej nagrody, rezultatu działania. Dokładniej, jeśli próba zakończyła się przyznaniem nagrody to do aktualizacji oczekiwanej nagrody użyta zostanie wartość α_+ , jeśli nagroda nie została przyznana α_- , inaczej α_+ w przypadku dodatniego błędu przewidywania, α_- przy wartości ujemnej błędu przewidywania (Rycina 2-4). Jeśli wartość optymalna parametrów α_+ i α_- jest różna to wskazuje to na zróżnicowaną wagę położoną na zdarzenie pozytywne i negatywne (Daw i in., 2002; Harada, 2020), co można interpretować w kontekście wrażliwości na informację zwrotną.



Rycina 2-4 Schemat blokowy algorytmu modelu z podwójnym tempem uczenia przy dwóch możliwych opcjach działania, zastosowany w niniejszych badaniach. Wartość oczekiwanej nagrody (Q) dla opcji rzeczywiście wybranej (c) przez ucznia ulega aktualizacji zależnie od rezultatu działania (r). Jeśli nagroda została przyznana („tak”) to wartość oczekiwanej nagrody wzrasta zależnie od współczynnika $\alpha+$, jeśli nie („nie”) to maleje zależnie od wartości $\alpha-$. Dla przejrzystości wpisano w równanie wartość 1 w przypadku uzyskania nagrody oraz 0 przy jej braku. Decyzja w kolejnej próbie podejmowana jest poprzez określenie prawdopodobieństwa uzyskania nagrody na podstawie aktualnej wartości oczekiwanej obu działań poprzez funkcję *softmax*. Uzyskanie nagrody (r) stanowi element środowiska niemożliwy do kontrolowania przez ucznia. W tych badaniach wartość początkową dla oczekiwanej nagrody oraz prawdopodobieństwa wyboru obu opcji przyjęto jako 0. Po obliczeniu wartości dla wszystkich prób (n) osobnika uzyskiwana jest całkowita wartość *nll* określająca dopasowanie modelu do danych (w). Do funkcji niezbędne jest podanie zmiennych określających wolne parametry modelu ($a+$, $a-$, b), które stanowią wartości maksymalizujące dopasowanie modelu do danych (nie pokazano). Dodatkowym elementem zastosowanym w niniejszych badaniach było ograniczenie wartości prawdopodobieństwa działań (P) do wartości z zakresu od 0,001 do 0,999 jako zabezpieczenie przed skrajnie niskimi i wysokimi wartościami funkcji *nll* (nie pokazano).

Inne rozwiązanie stanowi model, w którym uczenie się napędzane jest niepewnością. W odnoszącym się do tej koncepcji rozwiązaniu dostarczonym przez Yu Bai, Kentaro Katahira i

Hideki Ohira tempo uczenia α w każdej próbie jest proporcjonalnie zależne od wartości błędu przewidywania (Bai i in., 2014). W tym modelu, nazwanym w niniejszych badaniach „uwagowym”, wartość parametru α zmienia się w każdej próbie w odniesieniu do wartości podstawowej α_0 zależnie od wartości bezwzględnej błędu przewidywania (Tabela 2—1). W zamyśle im większy błąd przewidywania tym więcej informacji powinno być dostarczonych i zapamiętanych z bieżącej próby (Bai i in., 2014). Odpowiada to także wspomnianemu wcześniej zakresowi zmiany aktywności komórek dopaminowych w odpowiedzi na wielkość błędu przewidywania (Fiorillo i in., 2003) oraz hipotezie postawionej przez Johna Pearce i Geoffreya Hall’a o tym, że bodziec będzie przetwarzany tylko wtedy gdy będzie potrzebny, czyli gdy jego następstwo nie będzie pewne (Pearce i Hall, 1980). W niniejszych badaniach zastosowano aktualizację wartości α bieżącej próby przed szacowaniem oczekiwanej nagrody. Wartość parametru μ determinuje zakres wpływu błędu przewidywania na wartość parametru α . Im niższa wartość tym mniejszy wpływ błędu przewidywania. Zgodnie z wynikami symulacji podanymi przez Yu Bai, i wsp. model ten powinien być lepszy, w porównaniu do modelu podstawowego, szczególnie w trudniejszej wersji zadania z probabilistycznym odwróceniem, gdy prawdopodobieństwa uzyskania nagrody dla dwóch dostępnych opcji są do siebie zbliżone np. 40 i 60% (Bai i in., 2014).

Kolejna ważna modyfikacja to możliwość aktualizacji oczekiwanej nagrody dla opcji wybranej i opcji alternatywnych jednocześnie. Ten model określono jako „urojony” (ang. *fictitious update*), ponieważ w sytuacji wyboru między dwiema opcjami jednoczesnej aktualizacji ulega wartość oczekiwanej nagrody dla opcji wybranej (a) i niewybranej (b) (Tabela 2—1) (Hampton i in., 2007; Gläscher i in., 2009). Zakłada się, że zmiana wartości oczekiwanej nagrody dla opcji niewybranej jest przeciwna względem opcji związanej z rzeczywistym wyborem oraz przyjmuje się osobne określenie błędu przewidywania dla każdej opcji. Wtedy wartość oczekiwanej nagrody wzrasta w sytuacji rzeczywistego uzyskania nagrody dla opcji wybranej oraz w sytuacji nieuzyskania nagrody dla opcji alternatywnej oraz maleje dla opcji, która nie dostarczyła nagrody oraz tej, która hipotetycznie nie dostarczyłaby nagrody. Wiąże się to z myśleniem o przypuszczalnych skutkach działania, które nie zostało podjęte (ang. *counterfactual thinking*), a więc tworzeniem umysłowej reprezentacji działania środowiska. Dlatego model „urojony” mógłby zostać uznany za prostą wersję rozwiązania, w których zakłada się przeszukiwanie modelu środowiska pod kątem najlepszego sposobu działania.

Istotnym ograniczeniem rozwiązania „urojonego” jest zrównanie wagi położonej na rzeczywisty i przewidywany rezultat działania, wyrażonej parametrem α . Intuicyjnie, działanie pewniejsze co do rezultatu (rzeczywiste) mogłoby być ważniejsze w określaniu wartości oczekiwanej nagrody od hipotetycznego (urojonego). Dla sprawdzenia tej hipotezy w modelu „mieszanym”, zastosowanym w niniejszych badaniach, wartość oczekiwanej nagrody szacowana jest dla opcji wybranej i alternatywnej z wykorzystaniem parametru α_+ w sytuacji uzyskania nagrody i α_- w przypadku jej braku (Cieślak i in., 2018). Z kolei w modelu „podwójnie mieszanym”, który został opracowany na potrzeby niniejszych badań, zastosowano 4 osobne współ-

czynniki α . W sytuacji uzyskania dostępu do nagrody wartość oczekiwanej nagrody wzrasta dla opcji wybranej z wagą α_{+p} i jednocześnie maleje dla opcji alternatywnej z wagą $\alpha_{-\phi}$, natomiast w sytuacji braku nagrody maleje dla opcji wybranej z wagą α_{-p} i wzrasta dla opcji alternatywnej z wagą $\alpha_{+\phi}$. Istotny aspekt zastosowania modelu „podwójnie mieszanego” stanowi jego redukowalność do założeń innych modeli, a dokładniej: przy parametrach $\alpha_{-\phi}$ i $\alpha_{+\phi}$ o wartości 0 i nieróżnych α_{+p} i α_{-p} będzie stanowił model „podstawowy”, dla parametrów $\alpha_{-\phi}$ i α_{+p} o wartości 0 model „wartościозależny”, dla równych wartości α model „urojony”, natomiast dla jednakowych wartości par α_{+p} i $\alpha_{-\phi}$ oraz α_{-p} i $\alpha_{+\phi}$ rozwiązanie „mieszane” (Tabela 2–1).

Dodatkowo w niniejszych badaniach wykorzystano modele kontrolne. W prostym modelu „wzmocnienie” (ang. *noisy win-stay lose-shift*) prawdopodobieństwo wyboru zdeterminowane jest wartością rezultatu poprzedniego działania zgodnie z zasadą „zostań po wygranej” oraz „zmień po przegranej”, wraz z możliwością działania przeciwnego, zakładanego jako losowe, określonego wartością wolnego parametru ε (Wilson i Collins, 2019) (5).

$$p(a, n) = \begin{cases} 1 - \frac{\varepsilon}{2} & (r_{n-1} = 1 \wedge c_{n-1} = a) \vee (r_{n-1} = 0 \wedge c_{n-1} \neq a) \\ \frac{\varepsilon}{2} & (r_{n-1} = 1 \wedge c_{n-1} \neq a) \vee (r_{n-1} = 0 \wedge c_{n-1} = a) \end{cases} \quad 5$$

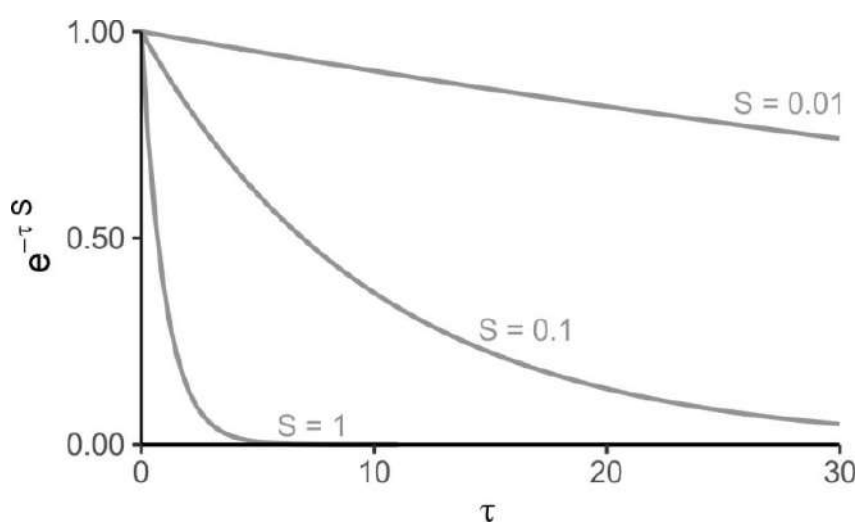
Wartość ε niższa od 1 wskazuje na częstsze od losowego wykonywanie korzystnego działania powtórzenia wyboru po wygranej oraz decyzji na korzyść opcji alternatywnej w przypadku braku nagrody. W równaniu c_{n-1} oznacza działanie poprzedzające bieżący wybór, natomiast r_{n-1} oznacza wartość nagrody otrzymanej po poprzedniej próbie, $r \in \{0,1\}$ i $\varepsilon \in (0.001, 1.999)$. Drugi model kontrolny opiera się na założeniu, że każda próba jest wykonywana w sposób losowy, czyli prawdopodobieństwo wyboru każdego z działań, w każdej próbie, przyjęto jako 0,5.

Wzajemny wpływ czasu decyzji i strategii ich podejmowania pozostaje jak dotąd opisanym w niewielkim stopniu aspektem badań uczenia ze wzmocnieniem, pomimo dyskusji efektu pamięci roboczej (Collins i Frank, 2012) oraz propozycji włączenia efektu zapomnienia (bez odniesienia do rzeczywistego czasu) (Wilson i Collins, 2019). Prawdopodobną przyczyną wydaje się trudność w opracowaniu schematu doświadczalnego, struktury testu, która nie wymuszałaby określonego czasu podjęcia działania. Natomiast w celu opisu zachowania zwierząt, obserwowanego w niniejszych badaniach, potrzebne było wprowadzenie modyfikacji modeli dotyczącej czasu podjęcia decyzji. Pierwszym pomysłem było wprowadzenie efektu zapomnienia. Inspiracji dostarczyło doświadczenie stanowiące współczesne odtworzenie badań wykonanych przez Hermanna Ebbinghausa, w których zależność zapamiętanej informacji w czasie określono jako funkcję potęgową (cytowane za (Murre i Dros, 2015)). W odniesieniu do tego możliwą do łatwego włączenia funkcję udało się znaleźć w pracy Piotra Woźniaka, Edwarda Gorzelańczyka i

Janusza Murakowskiego (Wozniak i in., 1995). Dokładniej, za efekt zapominania odpowiadałaby funkcja wykładnicza $e^{-\tau S}$, czyli rozpad zależny od czasu τ , w której tempo determinowane jest wartością S (ang. *storage*) w zakresie od 0 do 1. Oznacza to, że im więcej czasu minęło od poprzedniego działania tym wartość oczekiwanej nagrody związana z tym działaniem będzie mniejsza (6)(Rycina 2-5). W niniejszych badaniach założono jednoczesny i jednakowy rozpad wartości oczekiwanej nagrody dla opcji wybranej i alternatywnej do rzeczywistego wyboru, zgodnie z

$$Q_{n+1} = e^{-\tau_{n+1}S}(Q_n).$$

6



Rycina 2-5 Tempo rozpadu zależnie od wartości współczynnika S . Wartość funkcji rozpadu zależnie od czasu mijającego między kolejnymi próbami uzyskania nagrody (τ) zależnie od wartości współczynnika S , gdzie $S \in (0,1)$.

Odwzorowuje to sytuację, w której uczeń z czasem traci oczekiwanie względem wszystkich dostępnych opcji. Dla opisaney właściwości możliwa jest alternatywna interpretacja w postaci zjawiska znanego w ekonomii jako dyskontowanie czasowe (ang. *temporal discounting*) i impulsywności w psychologii (Long i Platt, 2005). Odnosi się to do funkcji wartości nagrody, która maleje w czasie. Dyskontowanie czasowe jest zazwyczaj badane poprzez oferowanie wyboru między mniejszą nagrodą teraz, a większą nagrodą w przyszłości. Przykładowo, abstynenci alkoholowi rzadziej wybierają większą, ale opóźnioną kwotę (Mitchell i in., 2005). Preferencja mniejszej nagrody, dostępnej szybciej, wyrażona tempem spadku wartości (ang. *temporal discounting rate*), może wskazywać na działanie impulsywne lub służyć adaptacji do środowiska, w którym opóźniona nagroda jest wysoce niepewna. Wysokie tempo spadku wartości notowane jest przy działaniach szkodliwych dla zdrowia, jak nadmierne jedzenie, przyjmowanie substancji uzależniających (Story i in., 2014) i w uzależnieniu (MacKillop i in., 2011). Wskazane zjawisko badane jest także u gryzoni z wykorzystaniem innych testów.

Dodatkowo, w niniejszych badaniach rozważone zostało alternatywne rozwiązanie polegające na wprowadzeniu niepewności wzrastającej z czasem jaki minął od poprzedniej decyzji. W tym celu wspomnianą funkcję rozpadu połączono z wartością parametru β (7),

$$\beta_n = e^{-\tau_{n+1}\delta}(\beta_0), \quad 7$$

gdzie τ wskazuje czas, który minął od poprzedniego wyboru. Oznacza to spadek domyślnej wartości β_0 , a więc zdolności do rozróżnienia wartości oczekiwanej nagrody, wraz z czasem jaki minął od poprzedniej próby uzyskania nagrody. Im więcej czasu minęło od poprzedniej próby tym trudniejsze będzie określenie dla ucznia, która z dostępnych opcji jest korzystniejsza. Funkcję rozkładu można zestawić z dowolną funkcją obliczania wartości oczekiwanej nagrody, np. modelem „podstawowym” czy „podwójnie mieszanym”.

Podsumowując, dopasowanie zadania eksperymentalnego i modelu obliczeniowego o zrozumiałych, możliwych do interpretacji, założeniach odpowiadających badanemu procesowi pozwala wnioskować o zachowaniu w dokładny, ilościowy sposób (Wilson i Collins, 2019). Modelowanie obliczeniowe dostarcza narzędzi do dalszej analizy zachowania obserwowanego pod wpływem czynników farmakologicznych różnych gatunków zwierząt i służy zrozumieniu terapeutycznych efektów leków w zaburzeniach neuropsychiatrycznych (Robbins i Cardinal, 2019). W koncepcji uczenia ze wzmocnieniem, opartej o warunkowanie klasyczne, uczenie się jest interpretowane jako testowanie hipotez albo tworzenie reprezentacji otoczenia, dzięki której zwierzę przewiduje rezultaty swoich wyborów na podstawie błędu przewidywania, czyli różnicy pomiędzy oczekiwanym i uzyskanym rezultatem. W niniejszych badaniach włączono zestaw modyfikacji opartych o algorytm *Q-learning* wraz z rozwiązaniami kontrolnymi oraz wnioskowano o zachowaniu na podstawie optymalnych wartości parametrów. Wnioskowanie na podstawie wartości optymalnych parametrów modelu stanowi jedno z najważniejszych narzędzi modelowania procesu uczenia ze wzmocnieniem.

3. Cele badawcze

- Cel badań stanowiło opracowanie modelu behawioralnego testu z probabilistycznym odwróceniem, w warunkach możliwie zbliżonych do naturalnych, wraz z analizą strategii podejmowania decyzji. W celu opisu strategii podejmowania decyzji, w badaniach zaimplementowano modele obliczeniowe uczenia ze wzmocnieniem oparte o algorytm *Q-learning* wraz z nowymi rozszerzeniami. Dlatego celem było także szczegółowe porównanie i wyłonienie najlepszego modelu.
- Kolejny cel odnosił się do określenia potencjalnych zmian w sposobie wykonania testu z probabilistycznym odwróceniem pod wpływem substancji uzależniającej tj. etanolu, w niskim stężeniu, w porównaniu do nagrody naturalnej.
- Niniejsze badania dotyczyły także roli receptorów związanych z transmisją sygnału w układzie serotoninowym w uczeniu ze wzmocnieniem. Dokładniej, podjęto próbę określenia wpływu braku receptorów NMDA, zlokalizowanych wybiórczo na komórkach serotoninowych, oraz receptorów 5-HT₇, na sposób wykonania testu probabilistycznego uczenia z odwróceniem.
- Przedmiotem badań było także uzyskanie opisu zależności strategii podejmowania decyzji od czasu spontanicznego działania, w teście naśladującym warunki naturalne dowolnego sięgania po nagrodę.

4. Metody

4.1 Zwierzęta

Doświadczenia zostały wykonane na myszach (1) szczepu C57BL/6J, (2) zmodyfikowanych genetycznie myszy z selektywną inaktywacją receptorów NMDA w komórkach serotoninowych linii NR1^{Tph2CreERT2} (Niewoehner i in., 2007; Weber i in., 2011) oraz (3) myszach pozbawionych funkcjonalnych receptorów 5-HT₇ linii Htr7^{tm1Sut} (Hedlund i in., 2003). Procedury wykonywane w ramach doświadczeń były zatwierdzone przez II Lokalną Komisję Bioetyczną w Krakowie (Numery zgód: 1000/2012, 1159/2015 i 2/2020). Zwierzęta przebywały w warunkach temperatury 22±2°C, wilgotności 35–60%, cyklu okołodobowego światło-ciemność po 12 godzin w klatkach po 2–5 osobników. Zwierzęta miały wolny dostęp do wody i paszy (Special Diets Services). Jeśli w opisie doświadczenia nie wskazano inaczej, warunki eksperymentalne były zgodne z warunkami hodowli.

4.1.1 Myszy pozbawione receptorów NMDA na komórkach serotoninowych

Myszy transgeniczne NR1^{Tph2CreERT2} nie posiadają funkcjonalnych receptorów NMDA na komórkach serotoninowych. Mutacja indukowana jest za pomocą systemu Cre/loxP. Związanie tamoksifenu i receptora estrogenowego CreERT2 powoduje translokację rekombinazy Cre do jądra komórkowego. Domena wiążąca ligand tego receptora jest czynnikiem regulującym działanie rekombinazy Cre. W wyniku działania rekombinazy Cre następuje wycięcie zintegrowanego z chromosomem genu flankowanego przez sekwencje loxP (Feil i in., 1996). Linia myszy NR1^{Tph2CreERT2} została wytworzona poprzez skrzyżowanie myszy TphCreERT2 (Weber, 2009) z myszami NR1^{flox/flox} (Niewoehner i in., 2007). Dla tej linii zwierząt genem zawierającym fragment oflankowany sekwencjami loxP był *Grin1*, czyli gen kodujący podjednostkę NR1 receptora NMDA, bez której nie powstają funkcjonalne receptory. Umieszczenie sekwencji CreERT2 pod kontrolą promotora genu *Tph2* (hydroksylazy tryptofanu 2, Tph2) skutkuje ograniczeniem mutacji do komórek Tph2 pozytywnych, czyli neuronów serotoninowych. Do badań włączono grupy zwierząt z mutacją [Tg/0; flox/flox] oraz kontrolne [0/0; flox/flox], po podaniu tamoksifenu. Zwierzęta były hodowane w schemacie kongenicznym i krzyżowane wstecz więcej niż 6 pokoleń do C57BL/6N.

4.1.2 Indukcja mutacji

Tamoksifen (Sigma) został rozpuszczony w oleju słonecznikowym (100 mg tamoksifenu na 5 mL oleju) i przefiltrowany przez błonę nitrocelulozową z porami o średnicy 0,22 µm. Myszy otrzymywały mieszaninę w postaci zastrzyku dootrzewnego w objętości odpowiadającej

5 µl na gram wagi ciała (w dniu rozpoczęcia podawania) codziennie przez 5 kolejnych dni (Jastrzebska i in., 2016). W dniu rozpoczęcia podawania zwierzęta były w wieku od 8 do 10 tygodni. Od zakończenia podawania tamoksifenu do testowania behawioralnego miały nie mniej niż dwa tygodnie.

4.1.3 Myszy pozbawione receptorów 5-HT₇

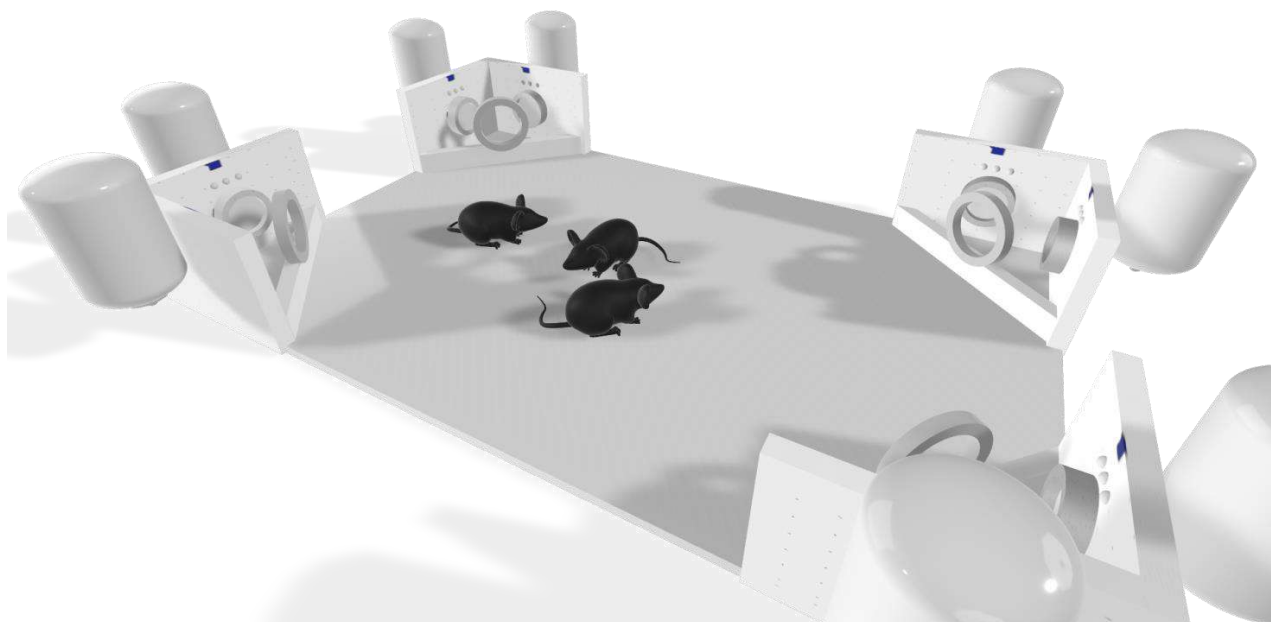
Genetycznie zmodyfikowane myszy *Htr7*^{tm1Sut} pozbawione są receptorów serotoninowych 5-HT₇ (www.informatics.jax.org/allele/MGI:2450732). W tej linii wstawienie genu oporności na neomycynę do eksonu II genu receptora 5-HT₇ powoduje powstaje skróconego, niefunkcjonalnego, białka receptora 5-HT₇ (Hedlund i in., 2003). W badaniach wykorzystano wyłącznie osobniki homozygotyczne z mutacją [ko/ko] i kontrolne [wt/wt]. Zwierzęta były hodowane w schemacie kongenicznym i krzyżowane wstecz więcej niż 6 pokoleń do C57BL/6N.

4.2 Doświadczenia

4.2.1 Test probabilistycznego odwrócenia w klatce IntelliCage

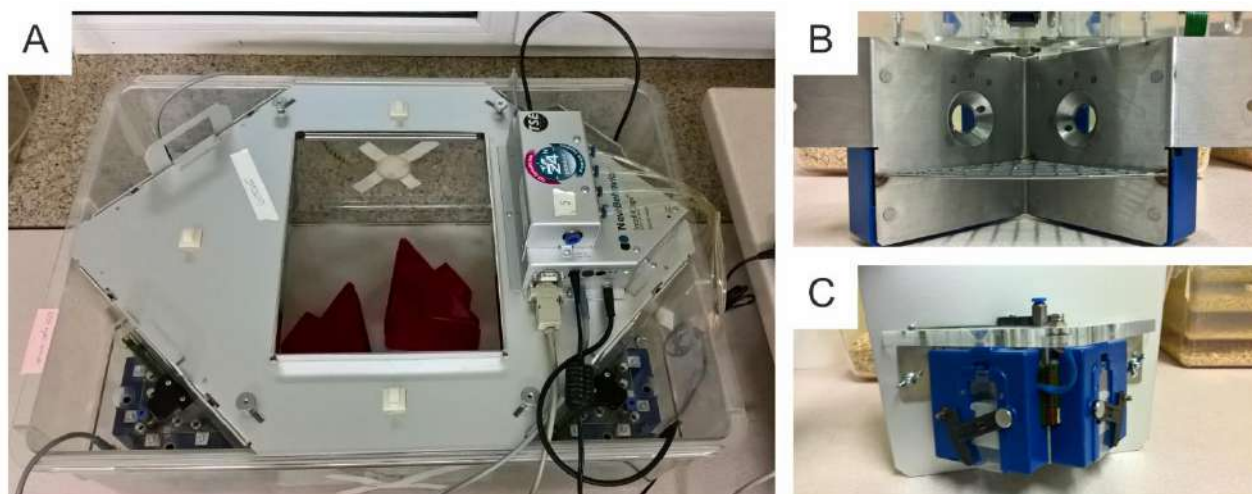
Aparat IntelliCage pozwala na automatyczne rejestrowanie zachowania grupy myszy podczas wykonywania testu. Podczas zastosowanego w niniejszych badaniach testu z probabilistycznym odwróceniem grupa 14 samic mogła sięgać po substancje znajdujące się w narożnikach aparatu. Zależnie od doświadczenia, substancją o właściwościach nagradzających była woda ze słodzikiem (0,1% w./obj. sacharyny), woda z etanolem (4% w./obj.) lub mieszanina obu

Aparat IntelliCage (New Behavior) stanowi prostokątna plastikowa przezroczysta klatka o rozmiarach 55 x 37,5 x 20,5 cm z metalową pokrywą oraz czterema komorami, nazywanymi narożnikami, znajdującymi się w rogach klatki (Rycina 4-1, Rycina 4-2A). Od strony wnętrza klatki narożnik składa się z okrągłego otworu, przez który mysz wchodzi do wewnętrznej komory. Komora narożnika zawiera parę okrągłych drzwiczek, przez które mysz może uzyskać bezpośredni dostęp do nasady butelki z napojem (Rycina 4-2B). Po zewnętrznej stronie narożnika dostępne dla eksperymentatora są dwa otwory na butelki skierowane leżkiem do dołu (Rycina 4-2C). Narożnik wyposażony jest w diody oraz system detekcji wykorzystujący chipy (tzw. transpondery, UNO PICO ID, AnimaLab) wszczepiane podskórnie myszom przed rozpoczęciem doświadczenia.



Rycina 4-1 Wizualizacja wnętrza aparatu IntelliCage, wersja uproszczona. Widoczna jest powierzchnia dolna klatki oraz cztery narożniki. W każdym narożniku znajdują się wewnętrzne drzwi prowadzące do lejka butelki po obu stronach komory wewnętrznej. Wykonanie wizyty przez myszy wiązało się z przekroczeniem otworu prowadzącego do komory narożnika, w którym mieściła się pojedyncza mysz. Zestaw trzech diod LED znajdował się nad każdą parą wewnętrznych drzwiczek. Schemat poglądowy bez zachowania dokładnych wymiarów.

Obecność transpondera w pobliżu detektora, w komorze narożnika, wraz z sygnałem z czujnika temperatury, warunkuje rejestrację wizyty zwierzęcia w narożniku. Notowane jest także przekroczenie drzwi wewnętrznych oraz liczba liźnięć butelki. Dno aparatu pokryte jest ściółką, a na niej dostępne są dla zwierząt cztery plastikowe elementy, które myszy wykorzystują jako gniazda. Przez cały czas trwania doświadczenia zwierzęta mają nielimitowany dostęp do karmy, umieszczonej we wgłębieniu górnej pokrywy klatki. Woda dostępna jest przez cały okres trwania doświadczenia w co najmniej w dwóch narożnikach aparatu jednocześnie, zgodnie ze schematem doświadczenia.



Rycina 4-2 Aparat IntelliCage. Zdjęcie A przedstawia aparat IntelliCage od góry. Dla przejrzystości pokazano aparat bez butelek, górnej pokrywy i ściółki. W aparacie umieszczone są cztery plastikowe elementy służące myszom jako gniazda. Zdjęcia B wskazuje wewnętrzną komorę narożnika od strony wnętrza klatki, C zewnętrzną część narożnika.

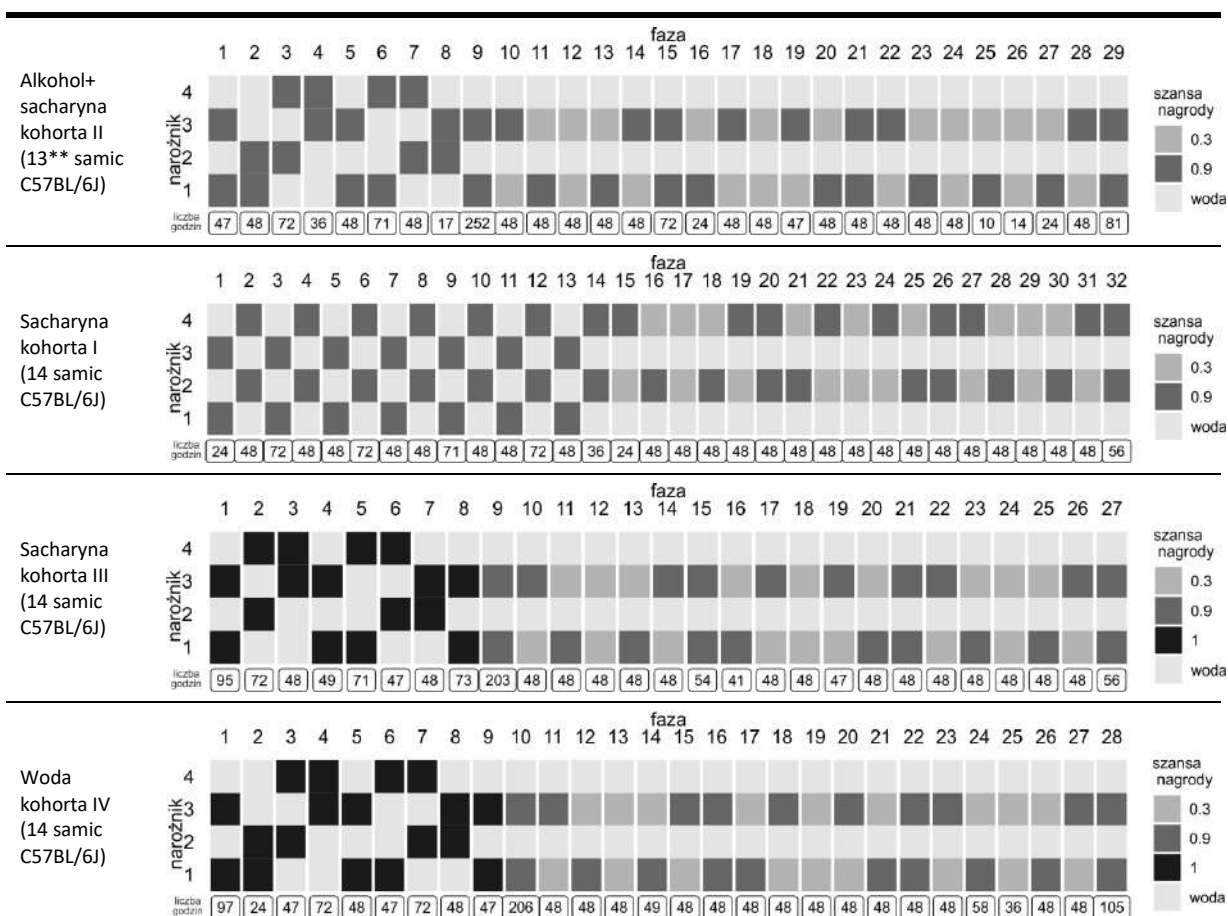
Doświadczenie składa się z etapu adaptacji i właściwego testu, które różnią się od siebie położeniem butelek z nagrodą i jej dostępnością. Podczas pierwszych 4–7 dni od wprowadzenia zwierząt do aparatu we wszystkich narożnikach znajdowała się woda co umożliwiało przystosowanie się myszy do korzystania z aparatu. Następnie, w etapie adaptacji, substancja o właściwościach nagradzających (nagroda) dostępna była w dwóch narożnikach. Prawdopodobieństwo jej uzyskania, czyli otwarcia wewnętrznych drzwi narożnika, w etapie adaptacji wynosiło 90 lub 100% zależnie od schematu doświadczenia (Tabela 4–1, Tabela 4–2).

Tabela 4–1 Zestawienie doświadczeń ze schematem odwróceń włączającym fazy pośrednie, o równym prawdopodobieństwie uzyskania nagrody.

grupa*	schemat doświadczalny																																
Alkohol kohorta I (14 samic C57BL/6J)	faza																																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	4																																
	3																																
	2																																
1																																	
liczba godzin	24	48	72	48	46	72	48	48	71	48	48	72	47	36	24	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	56
Alkohol kohorta II (14 samic C57BL/6J)	faza																																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25							
	4																																
	3																																
	2																																
1																																	
liczba godzin	46	46	72	48	48	71	48	48	252	48	48	48	48	48	72	24	48	48	47	48	48	48	48	48	48	10							
		szansa nagrody																															
		0.3																															
		0.9																															
		woda																															

grupa*

schemat doświadczalny



* Wszystkie osobniki były w wieku ~8 tygodni. **Jeden osobnik zgubił transponder i został wykluczony z doświadczenia.

W etapie adaptacji położenie butelek z nagrodą było zmieniane co ok. 48 godzin podczas fazy jasnej, żeby zapobiec wykształceniu przez myszy indywidualnej preferencji narożnika. W trakcie właściwego testu położenie butelek pozostawało niezmiennie natomiast do dwóch narożników z nagrodą przypisane było prawdopodobieństwo uzyskania dostępu do butelek, wynoszące 30 i 90%. Przypisanie wartości prawdopodobieństwa uzyskania nagrody do narożników było odwracane co 48 godzin przez co najmniej 12 dni, w połowie ciemnej fazy cyklu dobowego zwierząt. W każdym etapie doświadczenia w dwóch narożnikach dostępna była woda. Natomiast wyłącznie wizycie zwierzęcia w narożniku z nagrodą towarzyszył sygnał świetlny z diod zlokalizowanych nad wewnętrznymi drzwiami. Ten sygnał trwał 2 sekundy, niezależnie od przyznania czy braku dostępu do butelek. Wewnętrzne drzwi zamykały się po 10 sekundach od pierwszego liźnięcia, niezależnie od rodzaju substancji. Dane uzyskano z wykorzystaniem oprogramowania do obsługi systemu IntelliCage dostarczonego przez producenta i przetworzono z wykorzystaniem napisanego do tego celu pakietu języka R “icager” (www.github.com/jmjablons/icager). Liczba godzin trwania fazy została ustalona na podstawie informacji o czasie wizyt zwierząt znajdujących się w klatce.

Tabela 4–2 Zestawienie doświadczeń ze skróconym schematem, bez faz pośrednich o równym prawdopodobieństwie uzyskania nagrody.

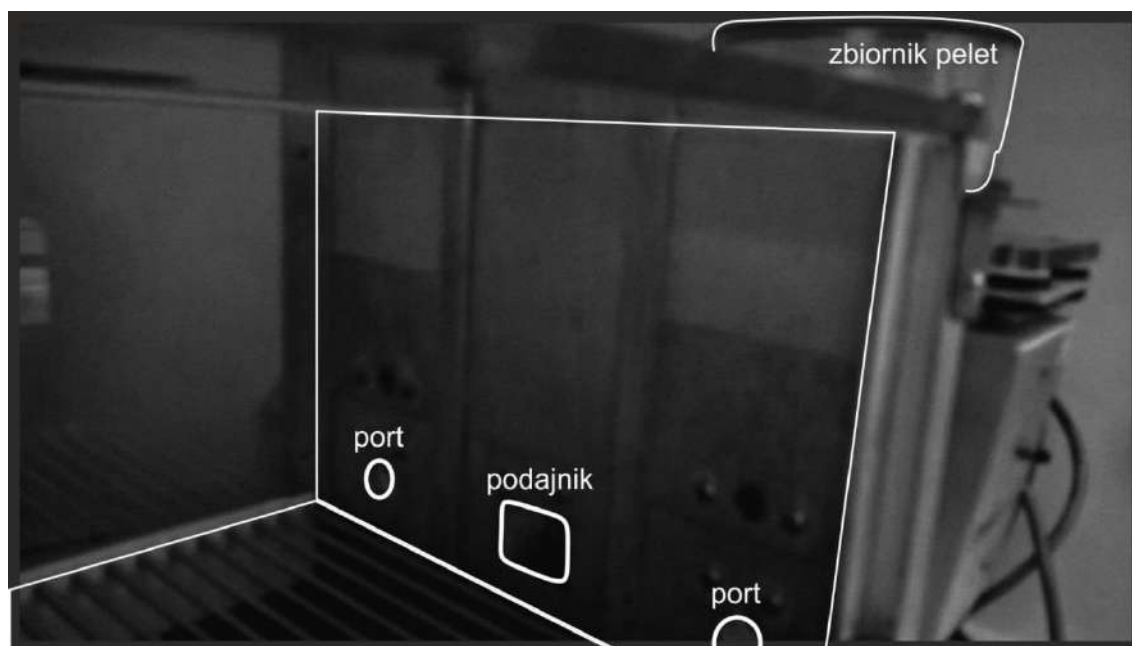
grupa	skład grupy (mutant/ kontrola)	wiek (min– max, dni)	schemat doświadczalny																
Htr7 kohorta I	5/9 samic	70–136	narożnik	faza															
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
				4	3	2	1												
			liczba godzin	42	51	45	54	57	48	48	48	48	48	48	48	48	48	48	
Htr7 kohorta II	7/7 samic	94–115	narożnik	faza															
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
				4	3	2	1												
			liczba godzin	47	47	49	47	39	48	48	48	48	72	48	48	48	48	48	
NR1 ^{Tph2CreE} RT2** kohorta I	6/8 samic	100– 127	narożnik	faza															
				1	2	3	4	5	6	7	8	9	10	11	12	13	14		
				4	3	2	1												
			liczba godzin	48	46	50	35	48	48	48	48	48	48	48	48	48	48		
NR1 ^{Tph2CreE} RT2 kohorta III	7/7* samic	87–112	narożnik	faza															
				1	2	3	4	5	6	7	8	9	10	11	12	13	14		
				4	3	2	1												
			liczba godzin	48	48	51	35	48	48	48	48	48	48	48	48	48	48		
C57BL/6J* *	14 samic	~56–70	narożnik	faza															
				1	2	3	4	5	6	7	8	9	10	11	12	13	14		
				4	3	2	1												
			liczba godzin	69	50	48	36	48	47	48	35	37	48	48	48	48	48		
			szansa nagrody																
			0.3																
			0.9																
			woda																

*Wśród nich dwie myszy krzyżówki wstecznej szczepu NR1flox **Sygnał świetlny towarzyszący wizytom w narożnikach z butelkami z nagrodą był koloru niebieskiego, zamiast domyślnie żółtego.

W trakcie doświadczeń pojawiały się problemy techniczne z rejestracją liźnięć butelek. Detektor liźnięć wykrywa zamknięcie przez mysz obwodu elektrycznego tworzonego przez noskę butelki i metalową kratkę w podłodze narożnika. Czulość rejestracji zmienia się wskutek zabrudzenia kratki lub noska butelki oraz jest trudna do właściwej kalibracji. Z tego względu miara liczby liźnięć jest niedokładna i dlatego w analizie danych posłużono się głównie miarami związanymi z wejściami do narożników.

4.2.2 Test probabilistycznego odwrócenia w klatce do warunkowania instrumentalnego

Doświadczenie zostało wykonane metodą opisaną w pracy Przemysława Cieślaka i współpracowników (Cieślak i in., 2018). W zarysie, myszy wykonywały w aparacie do warunkowania instrumentalnego test polegający na wyborze jednego z dwóch portów. Wybór skutkował otrzymaniem nagrody lub jej brakiem. Nagrodą w doświadczeniu było jedzenie w postaci chrupek (14 lub 20 mg, BioServ). W celu zmotywowania zwierząt do wykonania testu, przez cały czas doświadczenia ograniczano dostęp do pożywienia w klatkach domowych. Waga zwierząt była utrzymywana na poziomie niższym o 15% względem wartości w dniu poprzedzającym rozpoczęcie doświadczenia.



Rycina 4-3 Wnętrze klatki do warunkowania instrumentalnego. Zdjęcie z boku. Wskazano dwa otwory (porty) z tylnym światłem po obu stronach podajnika chrupek pokarmowych oraz zarys podłogi i ścianek klatki określających przestrzeń, w której znajdowała się mysz w trakcie testu. Pojemnik na chrupki pokarmowe oraz mechanizm automatycznego ich podawania znajdowały się na zewnątrz miejsca wykonania testu przez mysz.

Do doświadczenia zostały wykorzystane aparaty do warunkowania instrumentalnego (ENV-307W-CT, Med Associates). W aparacie znajdowały się: podajnik chrupek, dwa otwory

(porty) ze światłem, po obu stronach podajnika oraz lampka i wentylacja po przeciwnej stronie klatki (Rycina 4-3).

Całe doświadczenie składało się z czterech faz przygotowawczych oraz właściwego testu (Tabela 4–3). W pierwszych trzech fazach tylko jednemu z portów towarzyszył sygnał świetlny (port aktywny) i wetknięcie pyszczka wyłącznie do tego portu skutkowało otrzymaniem nagrody. W pierwszej fazie aktywny był lewy port, w drugiej prawy, w kolejnej aktywność portu była losowana w każdej próbie. W kolejnej fazie przygotowawczej wprowadzono niepewność związaną z pominięciem nagrody. Polegało to na braku nagrody z 50% szansą po wyborze aktywnego portu i było sygnalizowane światłem emitowanym po przeciwnej stronie aparatu. Sesje powtarzane były aż do spełnienia przyjętego z góry kryterium fazy (Tabela 4–3), nie częściej niż jedna sesja na dzień. Właściwy test składał się z jednogodzinnych sesji powtarzanych przez 14 kolejnych dni. Podczas właściwego testu oba porty aparatu były aktywne, jednak przy rozpoczęciu sesji przypisywane im było losowo prawdopodobieństwo uzyskania nagrody 20 i 80% odpowiednio dla jednego i drugiego otworu. W 60 próbie (połowie sesji) przypisanie szansy uzyskania nagrody do portów było odwracane. Czas od rozpoczęcia próby do wetknięcia pyszczka do portu (latencja wyboru) był nieograniczony. Z kolei czas od dostarczenia nagrody do kolejnej próby wynosił 5s. Nienagradzane wybory łączyły się z włączeniem światła w klatce na czas 5s. Doświadczenie wykonywano w jasnej fazie cyklu dobowego myszy.

Tabela 4–3 Fazy testu z probabilistycznym odwróceniem w klatce Skinnera

	przygotowanie - lewa strona	przygotowanie– prawa strona	przygotowanie– losowa strona	przygotowanie - uczenie pominięć (ang. <i>omission training</i>)	test probabil- istycznego od- wrócenia (ang. <i>probabilistic reversal learning task</i>)
kryterium zakoń- czenia pojedynczej sesji	40 prób lub 40 minut	40 prób lub 40 minut	60 prób lub 60 minut	120 prób lub 60 minut	120 prób lub 60 minut
kryterium zakoń- czenia fazy	40 nagród	40 nagród	60 nagród	ND	14 sesji

4.2.3 Test preferencji sacharyny

Test preferencji słodkiego napoju służy do określenia zmiany wrażliwości na nagrodę (np. (Papp i in., 1996)). Test wykonano metodą analogiczną do zawartej w pracy Kamili Jastrzębskiej i współpracowników (Jastrzebska i in., 2016). Do wykonania testu wykorzystano klatki (20 x 26 x 13,5 cm), w których myszy przebywały indywidualnie przez 24 godziny. W tym czasie mogły korzystać z dwóch butelek o objętości 25 ml. W jednej butelce dostępna była woda, w drugiej roztwór sacharyny (0,1% w./obj., Sigma). W celu zniwelowania indywidualnej

preferencji strony położenie butelek zostało zamienione po 12 godzinach (połowie czasu) od rozpoczęcia testu. Po zakończeniu testu określono objętości wypitych płynów.

4.3 Podsumowanie doświadczeń

Do przygotowania niniejszej rozprawy posłużyły dane z trzech doświadczeń: wykonanych przy zastosowaniu klatki IntelliCage, (1) ze schematem odwróceń włączającym fazy pośrednie, o równym prawdopodobieństwie uzyskania nagrody (Tabela 4–1), (2) ze skróconym schematem, bez faz pośrednich, (Tabela 4–2), (3) z zastosowaniem klatki do warunkowania instrumentalnego oraz pochodzące z testu preferencji sacharyny (Tabela 4–4). Pojedynczą grupę myszy oznaczono w tabelach (Tabela 4–1, Tabela 4–2, Tabela 4–4) jako numer kohorty.

Tabela 4–4 Zestawienie doświadczeń z zastosowaniem klatki do warunkowania instrumentalnego oraz pochodzących z innych testów

grupa	skład grupy (mutant/kontrola)	wiek (min–max dni)	schemat doświadczenia
NR1 ^{Tph2CreERT2} kohorta I	9/8 samce	116–123	Test probabilistycznego odwrócenia w klatce do warunkowania instrumentalnego (14 sesji po 120 prób z odwróceniem przypisania szansy nagrody w 60 próbie)
NR1 ^{Tph2CreERT2} kohorta II	6/3 samce	106–109	Test probabilistycznego odwrócenia w klatce do warunkowania instrumentalnego (j.w)
NR1 ^{Tph2CreERT2} kohorta I	6/8 samce	107–141	test preferencji sacharyny

4.4 Analiza statystyczna

W analizie statystycznej kryterium istotności wyniku przyjęto na poziomie 0,05. Do porównania mediany jednej grupy z wartością hipotetyczną wykorzystano test Wilcoxona. Do porównań mediany dwóch niezależnych grup wykorzystano test Manna-Whitneya-Wilcoxona. Do porównań median między dwoma zależnymi grupami wykorzystano test Wilcoxona z powtórzonymi pomiarami. Do porównania mediany między więcej niż dwoma grupami wykorzystano test Kruskala-Wallisa i post-test Dunna. Wykonano także analizę wariancji (ANOVA) dla danych z pomiarami wielokrotnymi, co zaznaczono w tekście. Wszystkie analizy przeprowadzono z wykorzystaniem środowiska R (R Core Team, 2020). Wykorzystano także pakiet *rstatix* (Kassambara, 2020).

Do analizy wyników doświadczeń z wykorzystaniem klatki IntelliCage włączono analizę szansy powtórzenia poprzedniego wyboru w oparciu o regresję logistyczną. Szansę powtórzenia poprzedniego wyboru szacowano na podstawie czynników: wystąpienia nagrody, czasu

decyzji w minutach oraz lokalizacji narożnika (8). Na tej podstawie wyznaczono także prawdopodobieństwo tego działania (9). Wartość współczynników β wskazała na zakres wpływu odpowiadających im czynników na obserwowane zachowanie. Wartości współczynników nie stanowiły odniesienia do parametru β i β_0 w modelach obliczeniowych opartych o *Q-learning*. Analizę wykonano dla danych każdego osobnika indywidualnie z wykorzystaniem funkcji *glm* pakietu *stats*. W równaniu (8) i (9) oznaczenie p odpowiada prawdopodobieństwu działania „zostań”, czyli wyboru zgodnego z poprzednim, $r \in \{0,1\}$ wystąpieniu lub braku nagrody w poprzedniej próbie, τ wskazuje wartość interwału decyzji w minutach, $s \in \{1,2\}$ oznacza indeks narożnika w poprzedniej próbie, jednoznaczny z jego lokalizacją tj. niższy i wyższy indeks. Interwał decyzji oznacza czas od zakończenia próby uzyskania nagrody do rozpoczęcia kolejnej. Wskazaną analizę przeprowadzono wyłącznie na próbach uzyskania nagrody wykonanych w etapie odwróceń.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 r + \beta_2 \tau + \beta_3 s \quad 8$$

$$p = \frac{e^{\beta_0 + \beta_1 r + \beta_2 \tau + \beta_3 s}}{1 + e^{\beta_0 + \beta_1 r + \beta_2 \tau + \beta_3 s}} \quad 9$$

4.5 Modelowanie procesu uczenia ze wzmocnieniem

Zastosowana w tych badaniach analiza z wykorzystaniem modelowania obliczeniowego polegała na testowaniu hipotez o procesie uczenia na podstawie dopasowania do danych eksperymentalnych oraz wartości parametrów, podobnie do wcześniejszych prac (Noworyta-Sokolowska i in., 2019; Jabłońska i in., 2021). Testowane modele uczenia się zostały napisane w języku C++, a następnie, za pomocą narzędzi statystycznych, szacowano ich jakość i dopasowanie do danych, w środowisku R. Każdy model posiadał wolne parametry, których wartość znajdowana była tak, żeby zmaksymalizować dopasowanie modelu do danych, czyli uzyskać jak najniższą wartość *nll* (10). Wartości parametrów posłużyły do wnioskowania o zachowaniu zwierząt. Wartości interwałów dłuższe niż 660 minut zostały ograniczone do tej wartości ze względu na techniczne kwestie obliczeniowe. Wartość e^{-660} , mogąca pojawić się dla modeli z rozszerzeniem o zależność od czasu decyzji, jest bliska najniższej możliwej do przechowania wartości ($\sim 2,225 \times 10^{-308}$). Modelowanie przeprowadzono na podstawie prób uzyskania nagrody z całego etapu odwróceń. Przy modelowaniu danych z doświadczenia z wykorzystaniem klatki do warunkowania instrumentalnego przyjęto wartość Q (4, Tabela 2—1) równą zero na początku każdej sesji.

4.5.1 Określenie dopasowania i jakości modelu

Do określenia dopasowania modelu do danych wykorzystano miarę AIC (*Akaike Information Criterion*) oraz w_i (*Akaike weight*) (Akaike, 1974; Burnham i Anderson, 2004). W równaniach k oznacza liczbę wolnych parametrów modelu (10), n numer próby uzyskania nagrody (11), N całkowitą liczbę prób uzyskania nagrody osobnika (11), M zbiór modeli (12). Wartość AIC_{min} stanowi wartość referencyjną, określoną dla każdego osobnika indywidualnie. Wartość AIC_{min} może odwoływać się do wybranego modelu referencyjnego lub do modelu, który w zestawieniu z pozostałymi rozwiązaniami dostarczył najniższej wartości AIC. W równaniu (11) wartość prawdopodobieństwa p dla opcji wybranej a w próbie n określona jest poprzez funkcję *softmax* jak wskazano wcześniej (4).

$$AIC = 2(nll) + 2k \quad 10$$

$$nll = \sum_{n=1}^N -\log(p(a, n)) \quad 11$$

Wartość AIC obliczono dla każdego modelu z wykorzystaniem optymalnych wartości parametrów określonych indywidualnie dla każdego zwierzęcia. Wartości prawdopodobieństwa (4) ograniczono do zakresu [0.001, 0.999]. Porównywanie wartości AIC różnych modeli przeprowadzono indywidualnie dla każdego osobnika na podstawie miary wagi Akaike (12, 13).

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{m=1}^M \exp\left(-\frac{1}{2}\Delta_m\right)} \quad 12$$

$$\Delta_i = AIC_i - AIC_{min} \quad 13$$

Jako dodatkowa miara dopasowania modelu posłużył ułamek poprawnych przewidywań określony jako liczba prób, dla których przewidywana za pomocą modelu wartość prawdopodobieństwa wykonanego działania przekroczyła 50%. Przewidywanie wykonano dla danych każdego osobnika indywidualnie z zastosowaniem zestawu optymalnych wartości parametrów określonych dla tego osobnika.

Dodatkowo, dla określenia czy dopasowanie modelu wynikło z efektu wzmocnienia czy dodatkowych czynników weryfikowano uzyskane wartości parametrów związanych z rzeczywistym wzmocnieniem, tj. parametrów α i β w modelu „podstawowym” i „urojonym”, α_0 i β w „uwagowym”, α_p i β w „incydentalnym”, α_+ i β w „wartościозależnym” i „mieszanym”, α_{+p} i β

w „podwójnie mieszanym” i „podwójnie mieszanym z rozpadem Q” oraz α_{+p} i β_0 w modelu „podwójnie mieszanym z rozpadem β ”.

4.5.2 Wolne parametry

Optymalne wartości wolnych parametrów modelu zostały znalezione poprzez przeszukiwanie k -wymiarowej (k jako liczba wolnych parametrów modelu) przestrzeni wartości nll (11) quasinewtonowską, zoptymalizowaną pod względem wykorzystania pamięci, metodą Broyed-Fletcher-Goldfarb-Shanno “L-BFGS-B” dostępną poprzez funkcję “optim” pakietu “stats” (wersja 4.0.2). Optimalny zbiór wolnych parametrów to zbiór poprzez który uzyskuje się najniższą wartości nll modelu.

Tabela 4–5 Wartości parametrów modeli

parametr	wartości początkowe optymalizacji	zakres	model
$\alpha, \alpha_0, \alpha_+, \alpha_-, \alpha_p, \alpha_\phi, \alpha_{+p}, \alpha_{+\phi}, \alpha_{-p}, \alpha_{-\phi}$	{0.05, 0.5, 0.95}	[0–1]	podstawowy wartościowo niezależny urojony uwagowy incydentalny podwójnie mieszanym podwójnie mieszanym + rozpad Q podwójnie mieszanym + rozpad β
β, β_0	{1.25, 2.5, 3.75}	[0–50]	podstawowy wartościowo niezależny urojony uwagowy incydentalny podwójnie mieszanym podwójnie mieszanym + rozpad Q podwójnie mieszanym + rozpad β
δ, μ, Σ	{0.05, 0.5, 0.95}	[0–1]	uwagowy podwójnie mieszanym + rozpad Q podwójnie mieszanym + rozpad β
ε	-	[0,1–1,999]	wzmocnienie

Optymalne wartości parametrów zostały wyznaczone dla każdego osobnika indywidualnie, wraz z przekazaniem do funkcji wartości granicznych parametrów. Dla zapewnienia najlepszego wyniku przeszukiwanie rozpoczynano od dokładnie 3 wartości dla każdego parametru, czyli np. dla modelu z trzema parametrami iloczyn kartezjański tych wartości określa $5^3 = 125$ punktów startowych (Tabela 4–5). Liczba parametrów danego modelu (k) została podana wcześniej (Tabela 2–1). Wartości startowe parametrów optymalizacji ustalono na podstawie badań wstępnych. Optymalną wartość parametru ε (model „wzmocnienie”) określono poprzez przeszukiwanie całego zakresu 1900 wartości {0,1; 0,101; 0,102; ...; 1,999}. Wartości optymal-

nych parametrów dla grupy osobników przedstawiono jako mediana i 95% przedział ufności. Przedział ufności określono za pomocą podstawowej metody „bootstrap”, z 10000-krotnym próbkowaniem, z użyciem pakietu „boot” w środowisku R (R Core Team, 2020).

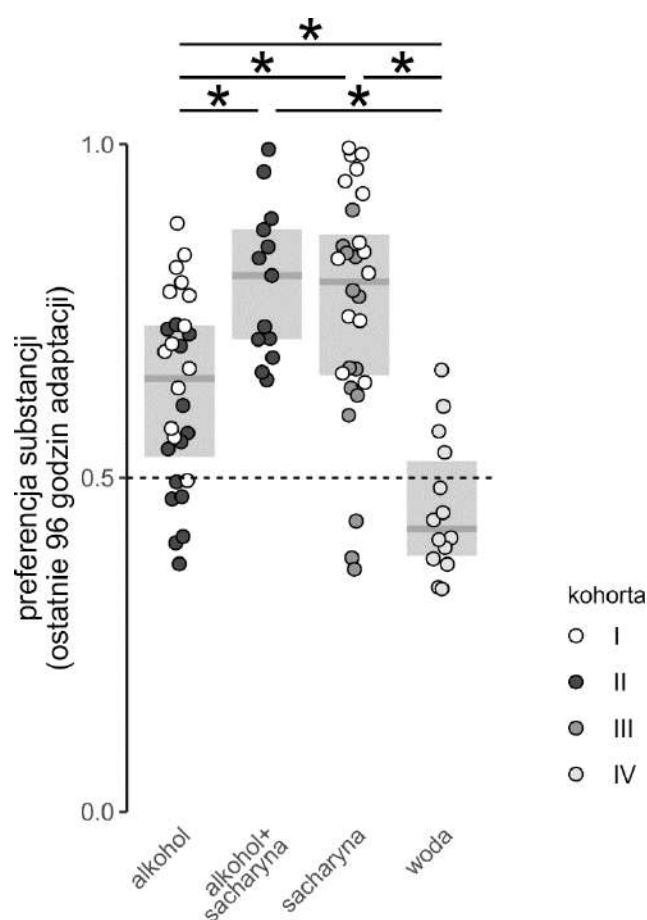
5. Wyniki

W przeprowadzonych doświadczeniach analizowano strategię podejmowania decyzji w zadaniu wymagającym zdolności do uczenia ze wzmocnieniem oraz zależność między strategią a czasem podjęcia działania. W pierwszej części przedstawiono analizę zachowania myszy C57BL/6 w teście probabilistycznych odwróceń, z wykorzystaniem klatki IntelliCage, w zależności od oferowanej nagrody: etanolu i sacharyny. W kolejnej części opisano wpływ braku receptorów serotoninowych 5-HT₇ i selektywnej inaktywacji receptorów NMDA na komórkach serotoninowych na zachowanie zwierząt obserwowane w uproszczonej wersji testu. W ostatnim doświadczeniu efekt selektywnej inaktywacji receptorów NMDA na komórkach serotoninowych testowano z wykorzystaniem klatek do warunkowania instrumentalnego, w schemacie bliższym klasycznemu. Analiza każdego doświadczenia polegała na przedstawieniu obserwowanego zachowania oraz scharakteryzowaniu strategii podejmowania decyzji z wykorzystaniem modelowania obliczeniowego.

5.1 Analiza uczenia ze wzmocnieniem zależnie od rodzaju nagrody w teście probabilistycznego odwrócenia w klatkach IntelliCage

Celem tej części pracy było scharakteryzowanie strategii podejmowania decyzji oraz określenie wzajemnej zależności strategii, czasu jaki upłynął od poprzedniego wyboru i oferowanej nagrody. W teście wykonanym w IntelliCage samice myszy linii C57BL/6J otrzymywały jako nagrodę: roztwór 0,1% (w./obj.) sacharyny (grupa sacharynowa; „sacharyna”), 4% roztwór (w./obj.) etanolu (grupa alkoholowa; „alkohol”) lub mieszaninę powyższych (grupa z mieszaniną; „alkohol+sacharyna”), które określono dalej jako substancje. Każda grupa zwierząt miała dostęp wyłącznie do jednego rodzaju substancji oraz do wody. Włączono także grupę kontrolną, w której substancja została zastąpiona wodą. Łącznie 6 grup zostało przetestowanych osobno w 4 kohortach (oznaczonych chronologicznie jako I – IV). W kohortach I i II test prowadzony był jednocześnie w dwóch aparatach z zastosowaniem jednakowego schematu doświadczenia (Tabela 4—1). Analizy tej części wykonano na danych pochodzących z doświadczenia przeprowadzonego przez Łukasza Szumca i Jana Rodrigueza Parkitnę (Zakład Neurofarmakologii Molekularnej IF PAN), za zgodą autorów. Wyniki zostały opublikowane (Jabłońska i in., 2021).

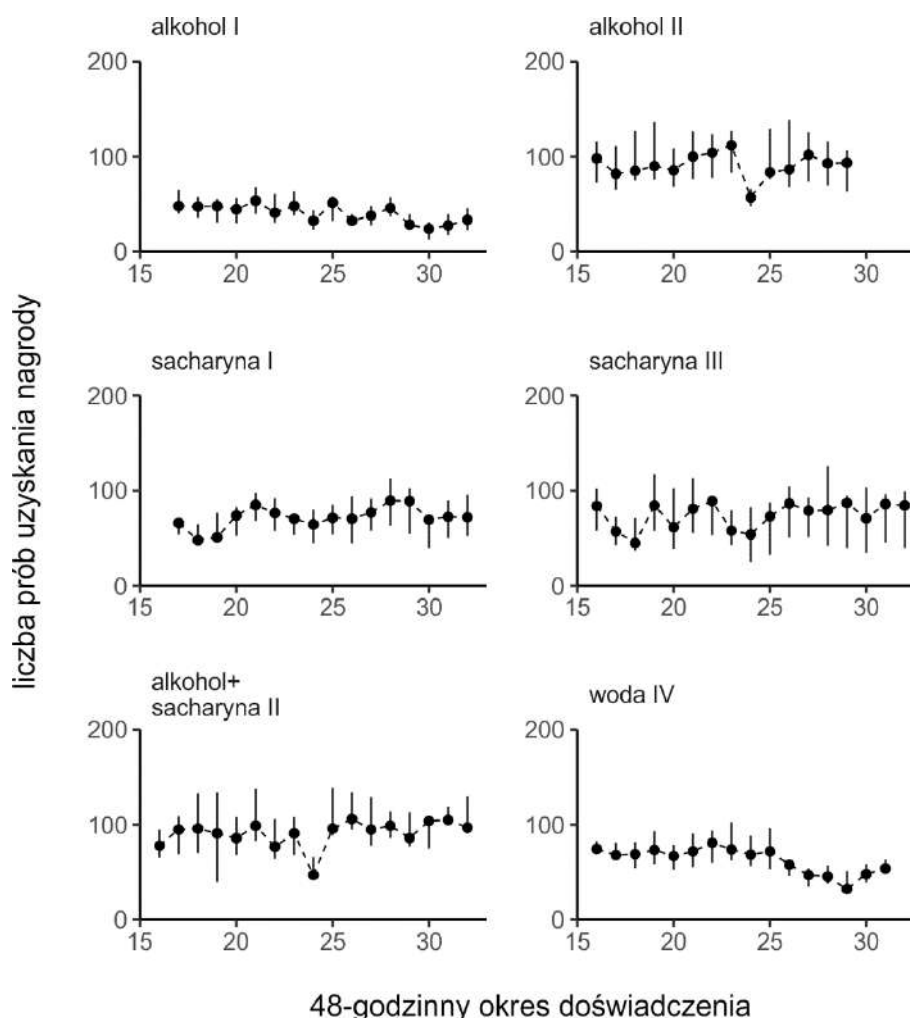
W pierwszym kroku sprawdzono czy zastosowane substancje działają wzmacniająco na zachowanie, czyli czy stanowią nagrodę dla zwierząt. W tym doświadczeniu, aby uzyskać dostęp do substancji, myszy kierowały się do jednego z czterech dostępnych narożników aparatu. W dwóch z nich dostępne były butelki z substancją, w pozostałych z wodą. Preferencję substancji określono jako liczbę liźnięć butelek z nagrodą względem wszystkich liźnięć, wykonanych podczas 96 ostatnich godzin etapu adaptacji (Tabela 4–1). Preferencja alkoholu była niższa w porównaniu do preferencji sacharyny i mieszaniny substancji, ale wyższa względem wody (test Kruskalla-Wallisa $p < 0,001$) (Rycina 5-1). Wskazana wartość była większa od 0,5 w grupie alkoholowej (test Wilcoxona z wartością 0.5, $p < 0,001$), w grupie z mieszaniną substancji ($p < 0,001$) i sacharynowej ($p < 0,001$), ale nie w grupie wodnej ($p = 0,135$). Wskazało to nagradzające właściwości wszystkich testowanych substancji.



Rycina 5-1 Właściwości nagradzające substancji. Preferencja substancji jako proporcja liczby „liźnięć” butelek z substancją względem wszystkich butelek (z substancją i wodą), w okresie ostatnich 96 godzin etapu adaptacji. Na wykresach pudełkowych długość prostokąta reprezentuje 50% środkowych obserwacji, pionowa linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu oznaczenie kohorty. Przerywana linia wskazuje próg wystąpienia preferencji. Znakiem „*” oznaczono znaczenie (test Kruskala-Wallisa $p < 0,05$ i test Dunna $p < 0,05$). Liczebność grup to „alkohol”: 28, „alkohol+sacharyna”: 13, „sacharyna”: 28 i „woda”: 14 osobników.

W trakcie doświadczenia myszy kierowały się do narożników klatki w celu uzyskania napoju. Z wizytami wyłącznie w narożnikach z nagrodą związany był sygnał świetlny pojawiający się na 2 sekundy nad wewnętrznymi drzwiczkami. Wizyty w narożnikach z nagrodą prze-

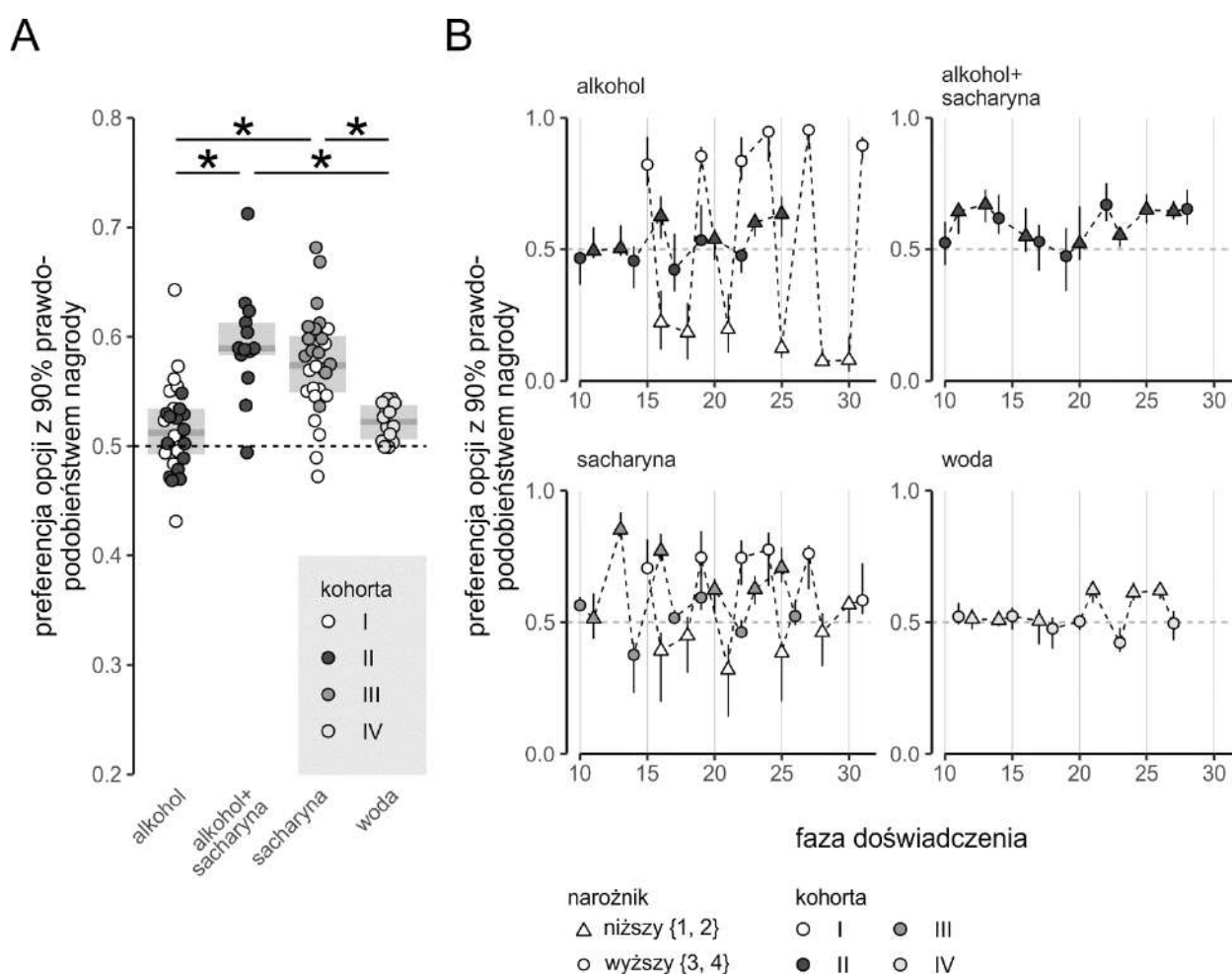
kraczące ten czas określono jako próby uzyskania nagrody. Liczba prób uzyskania nagrody utrzymywała się na stałym poziomie dla kolejnych 48-godzinnych okresów doświadczenia (Rycina 5-2). W grupie alkoholowej II zanotowano nieprawidłowości w danych od fazy 25 (okres 29, etap odwróceń) dlatego do dalszych analiz posłużyły wyłącznie dane uzyskane do 10 godziny tej fazy włącznie.



Rycina 5-2 Liczba prób uzyskania nagrody w etapie odwróceń. Liczba prób uzyskania nagrody dla kolejnych 48-godzinnych okresów doświadczenia. Punkt wskazuje medianę, zakresy 50% środkowych obserwacji. Numer odpowiada oznaczeniu kohorty. Dla przejrzystości pokazano wyłącznie okresy odpowiadające etapowi testu (odwróceń). Zmniejszona aktywność widoczna dla okresu 24 (faza 19) grupy alkoholowej II była spowodowana kilkunastogodzinnym brakiem rejestracji wizyt. Liczebność grup to alkohol I: 14, alkohol II: 14, alkohol+sacharyna: 13, sacharyna I: 14, sacharyna III: 14 i kontrolna: 14 osobników.

Gdyby przyjąć, że mysz dąży do maksymalizacji uzyskanych nagród i minimalizacji wysiłku to optymalną strategią byłby wybór opcji związanej z wyższym prawdopodobieństwem uzyskania nagrody. Przypisanie prawdopodobieństwa otwarcia wewnętrznych drzwiczek, prowadzących do butelek znajdujących się w narożnikach klatki (uzyskania dostępu do nagrody), określone jako faza, było odwracane co 48 godzin, w środku ciemnej fazy aktywności okołodobowej myszy. W tym doświadczeniu etap odwróceń charakteryzowały fazy o jednakowym prawdopodobieństwie uzyskania nagrody ($90\% \times 90\%$ lub $30\% \times 30\%$) przeplatane z fazami

o zróżnicowanej szansie nagrody w dwóch narożnikach klatki ($90\% \times 30\%$ oraz $30\% \times 90\%$) (Tabela 4–1). Otwarcie drzwiczek prowadzących do butelek z wodą następowało bez opóźnienia każdorazowo po wykryciu zwierzęcia w narożniku. W grupie kontrolnej dwa narożniki służyły dostępowi do „nagrody” (wody, z zadaniem prawdopodobieństwem) i dwa jako dostęp do wody (zdarzenie pewne), analogicznie do pozostałych grup. Częstość wyboru opcji związanej z większym prawdopodobieństwem nagrody określono jako liczbę prób uzyskania nagrody w narożniku o 90% prawdopodobieństwie uzyskania nagrody, względem liczby wszystkich prób w obu nagradzanych narożnikach, w fazach ze zróżnicowanym prawdopodobieństwem uzyskania nagrody ($30\% \times 90\%$ oraz $90\% \times 30\%$). Tak wyznaczona preferencja opcji o większym prawdopodobieństwie nagrody różniła się między nagrodami i była niższa dla grupy kontrolnej jak i alkoholowej w porównaniu do grupy sacharynowej i z mieszaniną substancji (test Kruskala-Wallisa $p < 0,001$) (Rycina 5-3A). W grupach kohorty I mediana wartości preferencji przekroczyła 50% wyłącznie gdy stroną o większym prawdopodobieństwie nagrody był narożnik o wyższym oznaczeniu (Rycina 5-3B), co sugerowało wykształcenie indywidualnej preferencji lokalizacji narożnika w tych grupach.



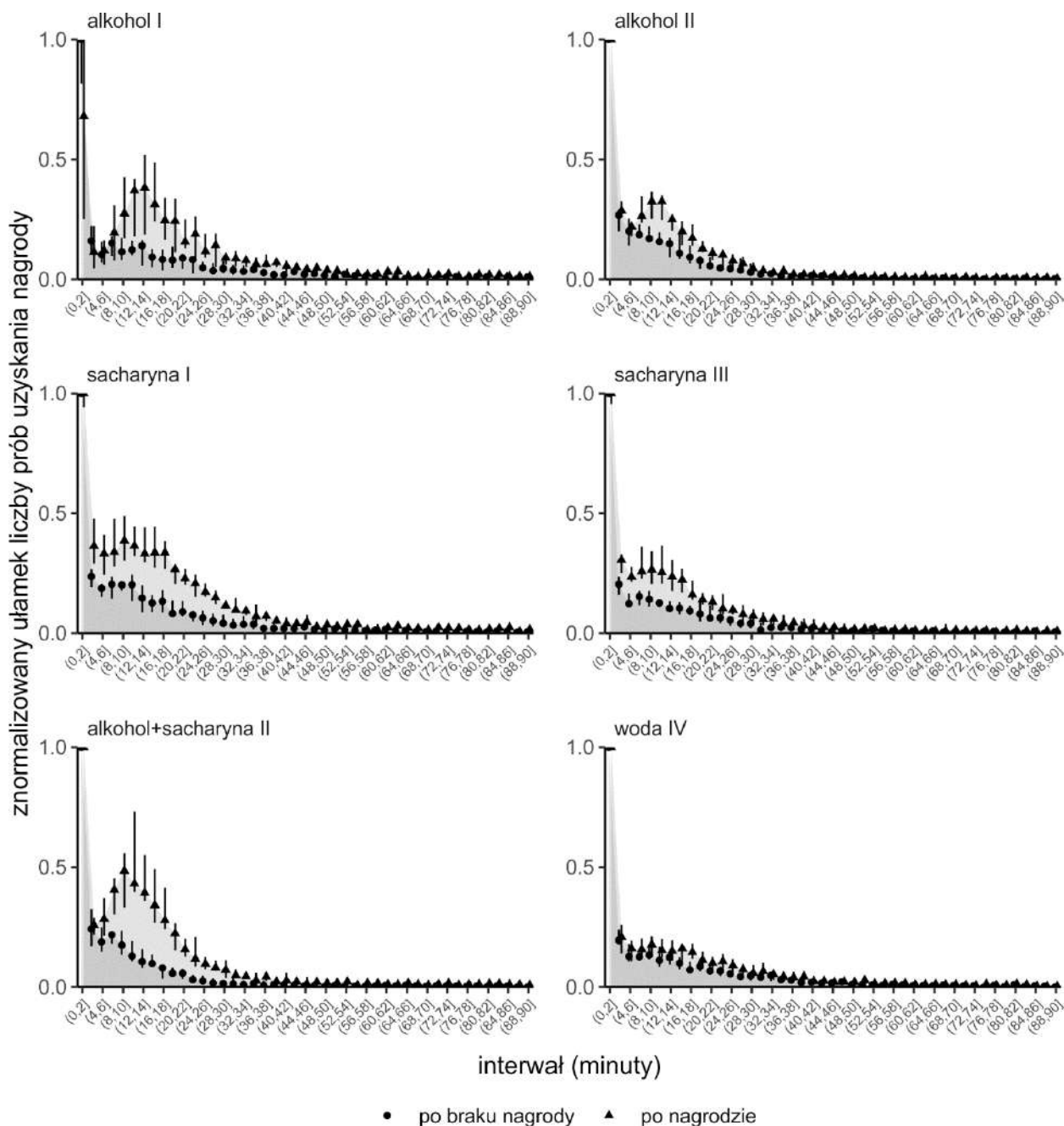
Rycina 5-3 Preferencja opcji z większym prawdopodobieństwem nagrody. Preferencja określona jako liczba prób uzyskania nagrody w narożniku o 90% prawdopodobieństwie uzyskania nagrody względem wszystkich prób, dla faz etapu odwróceń o zróżnicowanym prawdopodobieństwie uzyskania nagrody ($30\% \times 90\%$ i $90\% \times 30\%$). Przerywana linia wskazuje przyjętą

granice losowości. Liczebność grup to: „alkohol”: 28, „alkohol+sacharyna”: 13, „sacharyna”: 28 i „woda”: 14 osobników. (A) Na wykresie długość prostokąta reprezentuje 50% środkowych obserwacji, pionowa linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu kohortę. Znakiem „*” oznaczono znamienność (test Kruskala-Wallisa $p < 0,05$ i test Dunna $p < 0,05$). (B) Preferencja określona dla kolejnych faz etapu odwróceń o zróżnicowanym prawdopodobieństwie uzyskania nagrody ($30\% \times 90\%$ i $90\% \times 30\%$). Na wykresie punkt wskazuje medianę, zakresy 50% środkowych obserwacji, odcień punktu kohortę, kształt oznaczenie narożnika klatki (niższy odpowiada numerowi 1 lub 2, wyższy 3 lub 4, zależnie od kohorty).

Skrajna preferencja narożnika może oznaczać wybór niezależny od szansy uzyskania nagrody, a więc wskazywać na brak uczenia się albo persewerację. Dlatego sprawdzono czy liczba wyborów narożnika zależy od prawdopodobieństwa uzyskania w nim nagrody, dla wskazanych grup. W obu grupach liczba faz o zróżnicowanym prawdopodobieństwie uzyskania nagrody była równa (po 6 faz $30\% \times 90\%$ i $90\% \times 30\%$). Całkowita liczba prób uzyskania nagrody w narożniku o wyższym indeksie była większa gdy prawdopodobieństwo uzyskania nagrody w nim wynosiło 90% w porównaniu do 30%, w grupie sacharynowej I (sparowany test Wilcoxona $p = 0,00525$), ale nie w grupie alkoholowej I ($p = 0,0906$), co wskazało na silną indywidualną preferencję w grupie alkoholowej I.

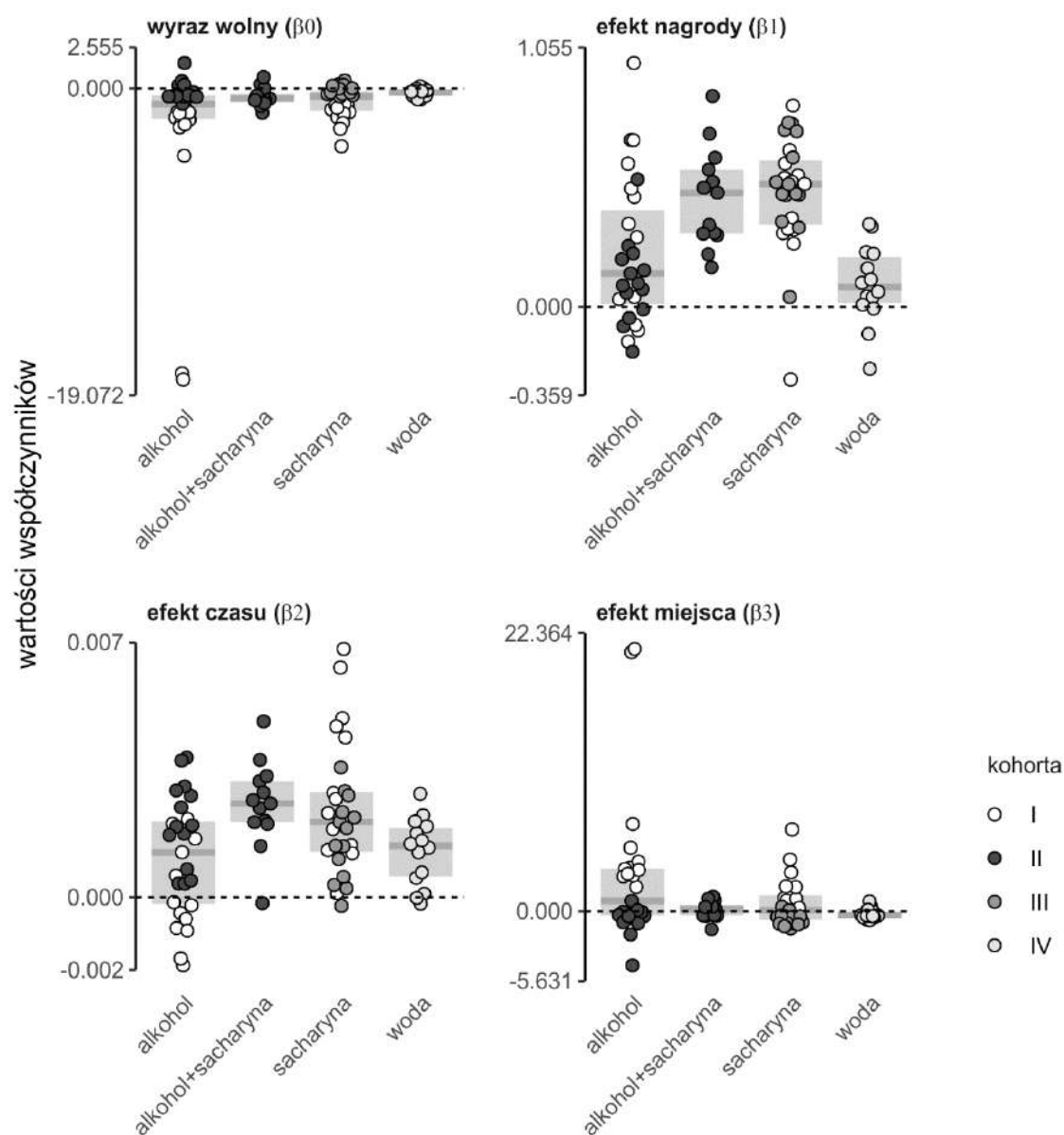
W doświadczeniu możliwe było zaobserwowanie zależności strategii podejmowania decyzji i czasu działania. Czas decyzji, określony jako czas mijający od zakończenia próby uzyskania nagrody do rozpoczęcia kolejnej próby przedstawia Rycinę 5-4. Dla prób zakończonych brakiem dostępu do nagrody dominował interwał krótszy niż dla prób zakończonych uzyskaniem nagrody. W obu przypadkach czas decyzji najczęściej mieścił się w zakresie do 2 minut. W sytuacji uzyskania nagrody w poprzedniej próbie w grupach sacharynowych częstość interwału do kolejnej próby stopniowo malała wraz z wydłużaniem interwału. Natomiast w grupach alkoholowych i z mieszaniną substancji, w tej samej sytuacji, widoczna była druga wartość szczytowa przesunięta do zakresu kilkunastu minut.

Z powyższych analiz wynika, że wybór mógł być zależny od przynajmniej trzech czynników: otrzymania dostępu do nagrody w poprzedniej próbie (efekt nagrody), indywidualnej preferencji lokalizacji narożnika (efekt miejsca) i czasu decyzji (efekt czasu). Do oszacowania zakresu wpływu wskazanych czynników na decyzję o powtórzeniu wyboru posłużyła regresja logistyczna z predyktorami w postaci: rezultatu poprzedniego działania (nagroda rozumiana jako uzyskanie dostępu do butelki lub jego brak), numeru narożnika określającego pozycję w klatce (niższy lub wyższy) w poprzednim działaniu oraz długości interwału w minutach. Wpływ czynników określono na podstawie prób uzyskania nagrody w całym etapie odwróceń. Z analizy wykluczono osobnika z grupy alkoholowej i osobnika z grupy sacharynowej, które wykonały za mało wyborów jednej z opcji do przeprowadzenia tej analizy (2/135 i 1/649, odpowiednio). Wpływ lokalizacji narożnika określono jako wartość bezwzględną współczynnika dla narożnika o wyższym indeksie.



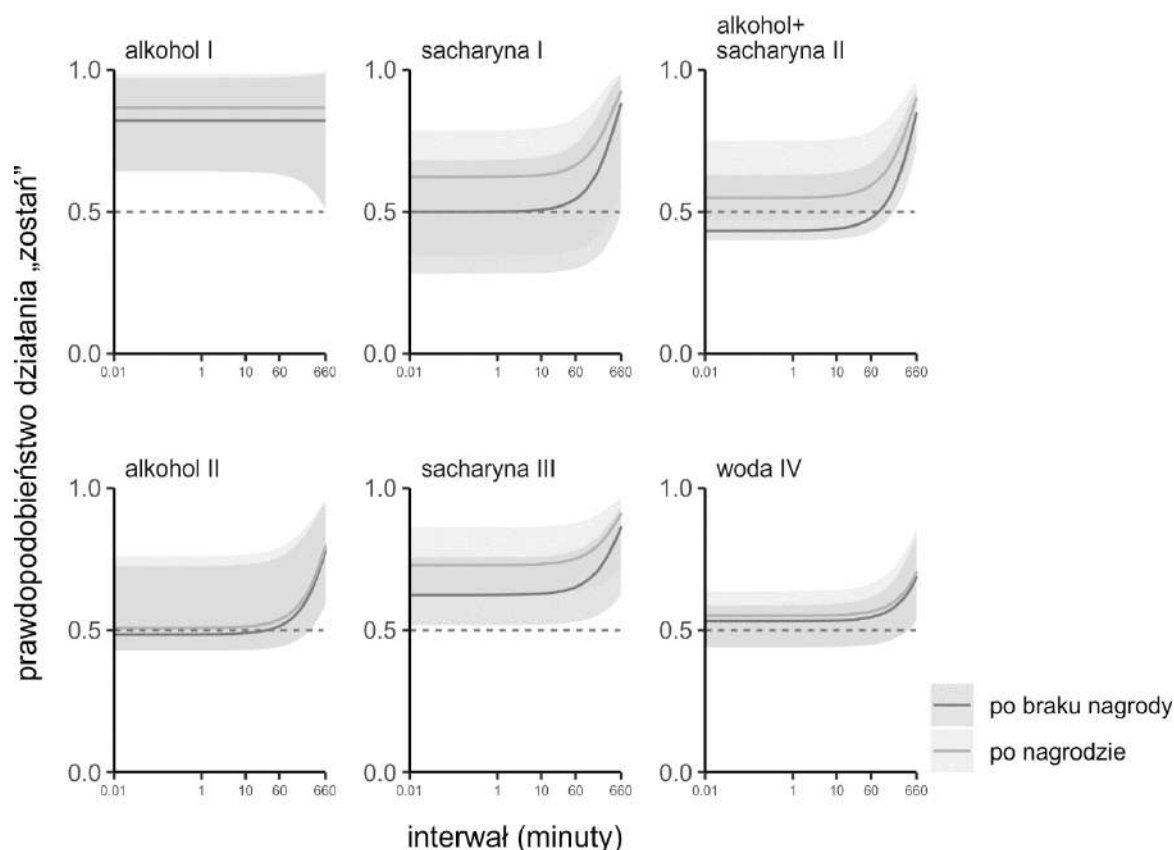
Rycina 5-4 Znormalizowana częstość interwału do kolejnego działania. Ułamek liczby prób uzyskania nagrody w etapie odwróceń znormalizowany względem referencyjnej wartości maksymalnej liczby prób uzyskania nagrody. Przedział z największą liczbą prób uzyskania nagrody wyznaczono indywidualnie dla każdego osobnika. Liczba prób uzyskania nagrody dla tego zakresu służyła jako wartość referencyjna dla pozostałych zakresów. Punkt wskazuje wartość medianę, zakresy 50% środkowych pomiarów. Interwał określa czas mijający do kolejnej próby uzyskania nagrody. Dla przejrzystości zakres interwału przedstawiono do wartości 90 minut i pole pod krzywą zaprezentowano w ciemniejszym odcieniu. Liczebność grup to alkohol: 28, alkohol+sacharyna: 13, sacharyna: 28 i kontrolna: 14 osobników.

Ujemna wartość wyrazu wolnego wskazuje na domyślnie zwiększoną szansę działania „zmień” (Rycina 5-5). Natomiast wartości współczynników powyżej 0 wskazują na zwiększone prawdopodobieństwo powtórzenia działania. Widoczne było zwiększenie szansy powtórzenia wyboru po otrzymaniu nagrody (efekt nagrody), ale także zależnie od lokalizacji narożnika (efekt miejsca) oraz wraz z wydłużaniem interwału do wyboru (efekt czasu) (Rycina 5-5).



Rycina 5-5 Współczynniki regresji logistycznej. Na wykresach pudełkowych długość prostokąta reprezentuje 50% środkowych obserwacji, pionowa linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu oznaczenie kohorty. Przerywana linia wskazuje kierunek zmiany (wartość powyżej wskazuje zwiększoną szansę działania „zostań”). Liczebność grup to „alkohol”: 28, „alkohol+sacharyna”: 13, „sacharyna”: 28 i „woda”: 14 osobników.

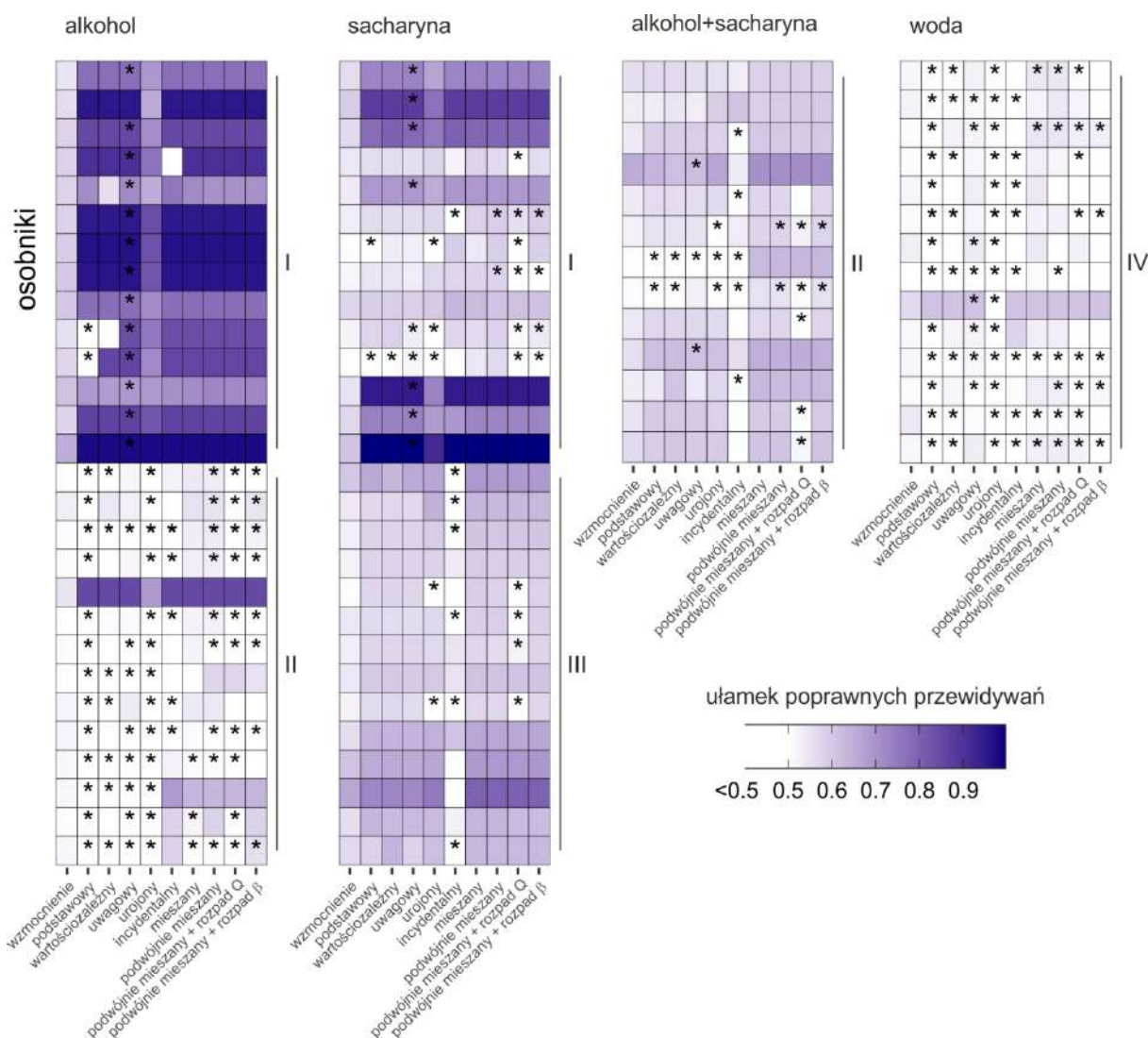
Dlatego wartość prawdopodobieństwa działania „zostań” wzrastała wraz z wartością interwału decyzji, zależnie od rezultatu działania oraz indeksu narożnika (Rycina 5-6).



Rycina 5-6 Prawdopodobieństwo działania „zostań” na podstawie regresji logistycznej. Prawdopodobieństwo działania „zostań” w zależności od interwału, szacowane na podstawie współczynników uzyskanych z regresji logistycznej. Efekt miejsca włączono jako $|\beta_3(1)|$. Linia ciągła wskazuje medianę, zakresy 50% środkowych obserwacji. Linia przerywana wskazuje wartość 0,5 przyjętą jako próg losowości. Wartość interwału została wyrażona w skali logarytmicznej. Liczebność grup to alkohol I: 13, alkohol II: 14, alkohol+sacharyna: 13, sacharyna I: 13, sacharyna III: 14 i woda: 14 osobników.

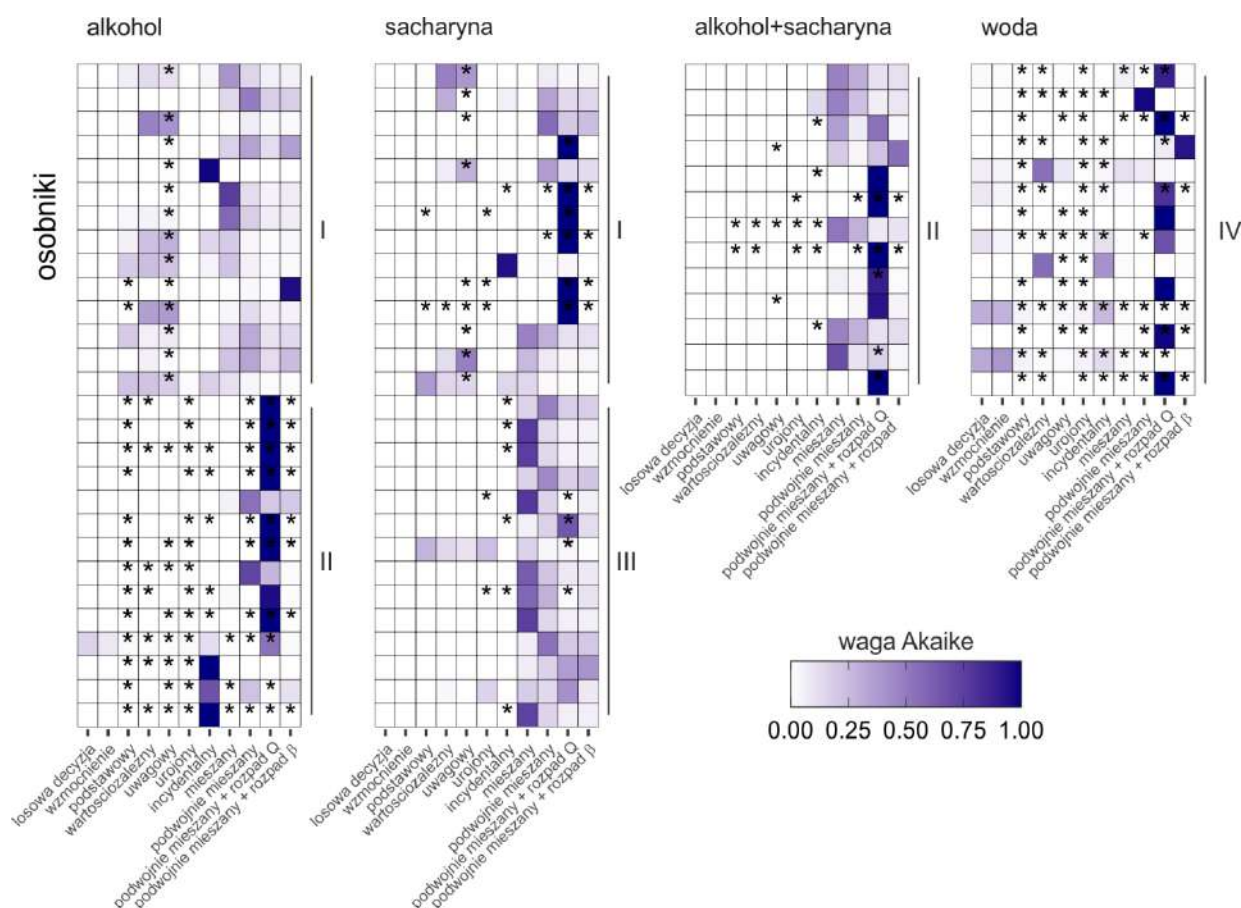
Wykazano różnice między substancjami dla współczynników β_0 (test Kruskala-Wallisa $p = 0,0328$), β_1 ($p < 0,001$), β_2 ($p = 0,00705$) ale nie β_3 ($p = 0,121$). Efekt nagrody, wyrażony wartością β_1 , był niższy dla grupy alkoholowej w porównaniu do grupy z mieszaniną substancji (test Dunna, $p = 0,0251$) i sacharynowej ($p = 0,00109$) oraz dla grupy kontrolnej w porównaniu do grupy z mieszaniną substancji ($p = 0,0033$) i sacharynowej ($p < 0,001$). Wskazało to silniejsze wzmacniające działanie roztworu sacharyny i mieszaniny substancji w porównaniu do alkoholu i wody. Efekt długości interwału, opisany wartością współczynnika β_2 , mieścił się w zakresie od $-0,0018$ do $0,0068$, a więc wpływ interwału na decyzję był wyraźny dla interwałów dłuższych od kilkudziesięciu minut (Rycina 5-6). Wartość β_2 była niższa dla grupy alkoholowej w porównaniu do grupy z mieszaniną substancji ($p = 0,0212$). Dlatego grupa alkoholowa różniła się od względem efektu czasu od grupy z mieszaniną substancji. W kilkunastu przypadkach wartość współczynnika dla interwału (8/27 osobników grupy alkoholowej, 1/13 grupy z mieszaniną substancji, 1/27 grupy sacharynowej i 2/14 grupy kontrolnej) i nagrody (7/27 osobników grupy alkoholowej, 1/27 grupy sacharynowej i 3/14 grupy kontrolnej) nie była dodatnia sugerując odbiegający wzorzec zachowania u tych osobników (Rycina 5-5).

Na podstawie przedstawionych wyników widoczne było wystąpienie uczenia ze wzmocnieniem. W celu szczegółowego określenia strategii podejmowania decyzji, dalszy opis zachowania poprowadzono z wykorzystaniem modeli obliczeniowych określających prawdopodobieństwo działania dla kolejnych prób uzyskania nagrody. Modele różniły się między sobą założeniami względem przebiegu zjawiska oraz złożonością i liczbą parametrów (Tabela 2–1). Optymalny zestaw parametrów jak i dopasowanie każdego modelu zostały wyznaczone indywidualnie dla każdego osobnika na podstawie prób uzyskania nagrody z całego etapu odwróceń. Określenia „uzyskanie nagrody” i „nagroda” odnoszą się do uzyskania dostępu do nagrody, a więc otwarcia wewnętrznych drzwiczek narożnika. W pierwszym kroku sprawdzono który model relatywnie dobrze odwzorowywał zachowanie na podstawie ułamka poprawnie przewidzianych wyborów (Rycina 5-7). Wartość ułamka poprawnych przewidywań odnosi się do liczby prób uzyskania nagrody, które zostały przewidziane z prawdopodobieństwem powyżej 50%. Ułamek poprawnych przewidywań o wartości równej i poniżej 0,5 wskazuje na zdolność do tłumaczenia zachowania nie lepszą niż założenie losowego wyboru. Taką sytuację opisano dla 47% (39/83) osobników w doświadczeniu dla modelu „incydentalnego”, 42,2% w modelu „urojonym”, 38,6% w modelu „podstawowym”, 33,7% w modelu „podwójnie mieszanym z rozpadem Q” i w pozostałych przypadkach dla mniej niż 1/3 badanych zwierząt (Rycina 5-7). Należy wziąć pod uwagę, że wartość ułamka poprawnych przewidywań była mniejsza bądź równa 50% dla ponad połowy przypadków zastosowania modeli dla grupy kontrolnej (75/140) i alkoholowej II (73/140) i nie więcej niż 14,6% (mieszanina substancji) w pozostałych grupach, co wskazuje gorsze dopasowanie modeli do danych grupy alkoholowej II (Rycina 5-7).



Rycina 5-7 Ułamek poprawnych przewidywań modeli zastosowanych w badaniach. Odcień pola odpowiada wartości ułamka poprawnych przewidywań uzyskanych dla danego modelu. Rząd określa wartości pojedynczego osobnika, kolumna wartości danego modelu. Wartości niższe i równe 0,5 oznaczono jednakowo. Znak „*” określa przypadki dla których parametry związane z rzeczywistym wzmocnieniem wyniosły zero. Wyniki przygotowano na podstawie danych z okresu odwróceń. Po prawej stronie wykresów oznaczono przynależność zwierząt do kohort.

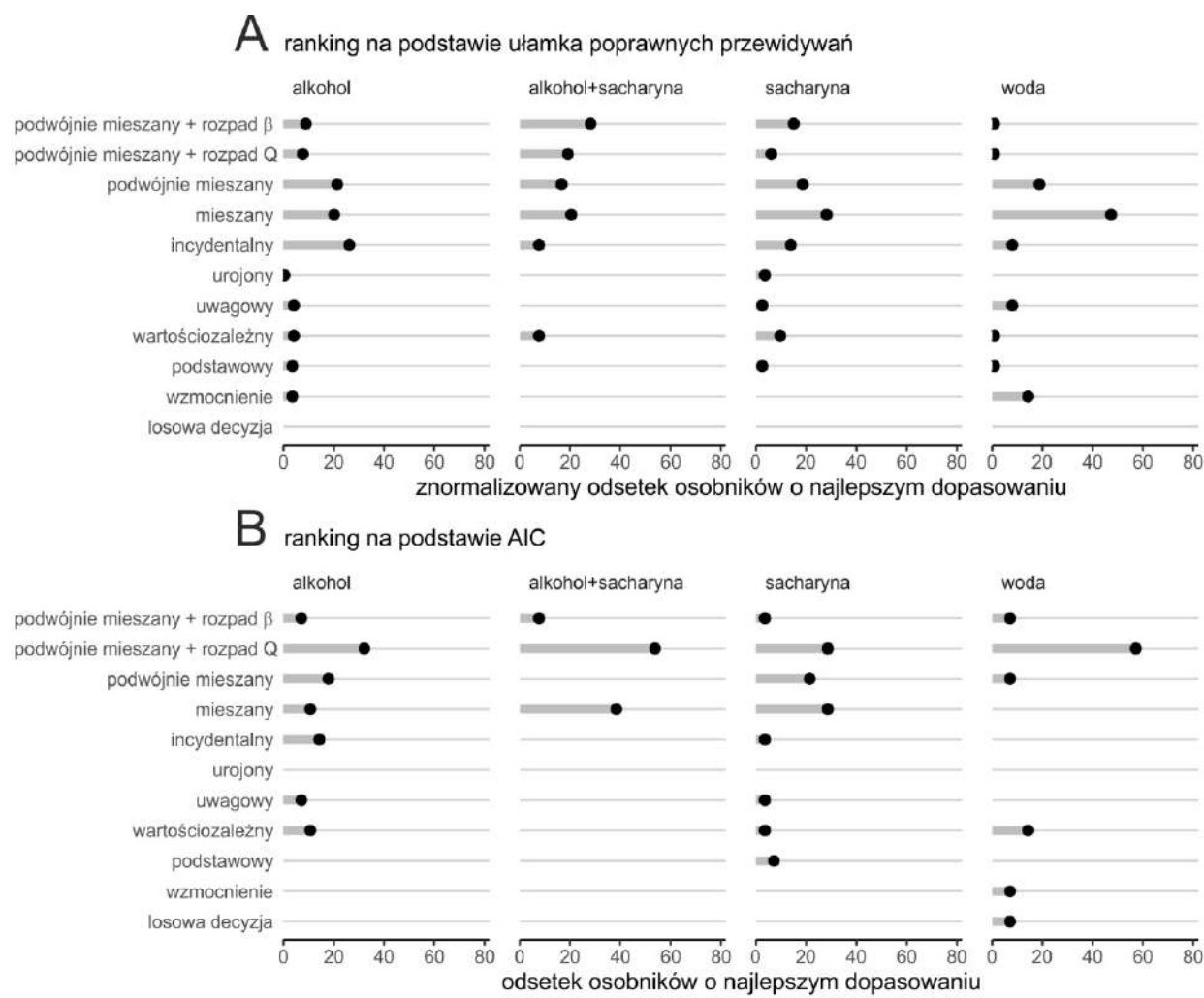
W odróżnieniu do przedstawionego ułamka poprawnych przewidywań miara AIC obliczana jest w oparciu o wartość prawdopodobieństwa każdej decyzji bez zadanego kryterium 50%. Przykładowo, model przewidujący wybory z prawdopodobieństwem 60% oraz model poprawnie przewidujący tyle samo wyborów ale z prawdopodobieństwem na poziomie 80% uzyskają taki sam odsetek poprawnych przewidywań, natomiast drugie rozwiązanie uzyska niższą wartość AIC. Dla jasnego zobrazowania wyników można posłużyć się miarą wagi Akaike (w), opartą na różnicy wartości AIC bieżącego modelu oraz modelu o najniższej wartości. Wartość w wskazuje prawdopodobieństwo najlepszego dopasowania modelu, względem pozostałych rozwiązań. Przykładowo, jeśli wszystkie 11 modeli charakteryzowałoby podobne dopasowanie do danych to w_i każdego z nich wyniosłoby $1/11$. Nie określono jednoznacznie najlepszego modelu spośród zaproponowanych (Rycina 5-8).



Rycina 5-8 Prawdopodobieństwo warunkowe modelu na podstawie wagi Akaike. Odcień pola odpowiada wartości w modelu. Rząd określa wartości pojedynczego osobnika (wartości sumują się do 1), kolumna wartości danego modelu. Znak „*” określa przypadki dla których parametry związane z rzeczywistym wzmooczeniem wyniosły zero. Wyniki przygotowano na podstawie danych z okresu odwróceń. Po prawej stronie wykresów wskazano przypisanie osobników do kohort.

Dla każdej substancji określono odsetek osobników o najlepszym dopasowaniu dla poszczególnych modeli (Rycina 5-9). Dla części zwierząt w doświadczeniu (32/83) ułamek poprawnych przewidywań był jednakowy dla dwóch lub więcej modeli, więc zastosowano normalizację w postaci przypisania wag. Oznacza to, że jeśli dla danego osobnika liczba poprawnych przewidywań była jednakowa dla np. 5 modeli to waga wyniku każdego z nich wyniesie 1/5, natomiast jeśli jeden model uzyskał najwyższą wartość wagi wyniesie 1. Dlatego określony ranking modeli odpowiada, w tym przypadku, nie liczbie osobników tylko sumie wag. W grupie alkoholowej dwoma najlepiej dopasowanymi modelami były: „incydentalny” i „podwójnie mieszany”, w grupie sacharynowej: „mieszany” i „podwójnie mieszany”, w grupie z mieszaniną substancji: „podwójnie mieszany z rozpadem β ” i „mieszany” (Rycina 5-9A). Natomiast na podstawie wartości wagi Akaike najlepiej dopasowanym modelem był „podwójnie mieszany z rozpadem Q ” w grupie alkoholowej (dla 32,1% osobników) i w grupie z mieszaniną substancji (53,8%) oraz modele „mieszany” (28,6%) i „podwójnie mieszany z rozpadem β ” (28,6%) w grupie sacharynowej (Rycina 5-9B). Tak utworzone rankingi wskazały na lepsze dopasowanie modeli zakładających określenie wartości niepodjętego działania (urojoną aktualizację) oraz

rozdzielenie wagi kroku (parametr α) w zależności od rezultatu działania, odnoszące się do wrażliwości na informacji zwrotną.



Rycina 5–9 Ranking modeli na podstawie odsetka grupy z największym dopasowaniem. Czarny punkt wskazuje wartość odsetka lub znormalizowanego odsetka grupy o największym dopasowaniu danego modelu na podstawie ułamka poprawnych przewidywań (A) oraz wartości AIC (B). Pokazano wartości dla wszystkich zastosowanych modeli. Dla wyników na podstawie ułamka poprawnych przewidywań (A) zastosowano normalizację w postaci przypisania wagi zależnej od liczby modeli o jednakowej wartości ułamka poprawnych przewidywań dla osobnika. Oznacza to, że jeśli dwa modele uzyskały jednakową wartość to waga każdego z nich wynosi 1/2. Liczebność grup to: „alkohol”: 28, „alkohol+sacharyna”: 13, „sacharyna”: 28 i „woda”: 14 osobników.

W kolejnym kroku wnioskowanie o zachowaniu zwierząt przeprowadzono na podstawie wartości parametrów modeli (Tabela 5–1). Najpierw analizowano modele z najlepszym dopasowaniem na podstawie wskazanych rankingów.

Tabela 5–1 Optymalne wartości parametrów modeli

model	para- ra- metr	alkohol I	alkohol II	alkohol+ sacharyna	sacharyna I	sacharyna III	woda
wzmoc- nienie	ε	0,837 (0,825–0,868)*	1 (0,988–1,03)	0,92 (0,906–0,957)	0,903 (0,857–0,95)	0,913 (0,881–0,961)	1,01 (0,993–1,05)

model	para- ra- metr	alkohol I	alkohol II	alkohol+ sacharyna	sacharyna I	sacharyna III	woda
podsta- wowy	α	0,00701 (0,00262–0,0105)	0 (0–0)	0,11 (0,047– 0,146)	0,0101 (0– 0,0136)	0,0481 (0,00662– 0,0731)	0 (0–0)
	β	7,81 (6,77–9,16)	0 (0–0)	0,682 (0,408– 0,797)	3,66 (1,55–6,7)	2,03 (1,48–3,1)	0 (0–0)
wartości- zależny	α_-	0,00358 (0– 0,00716)	0 (0–0)	0,18 (0,0502– 0,33)	0,00336 (0– 0,00609)	0,0314 (0– 0,0535)	0 (0–0)
	α_+	0,0112 (0,00176– 0,0161)	0,165 (0–0,331)	0,00886 (0– 0,0151)	0,0302 (0– 0,0465)	0,0147 (0– 0,0219)	0 (0–0)
	β	5,61 (4,42–6,34)	0,0809 (0– 0,162)	1,2 (0–1,84)	3,62 (3,14–6,59)	2,51 (0,0444– 3,78)	0 (0–0)
uwagowy	α_0	0 (0–0)	$1,23 \times 10^{-6}$ (0– $2,45 \times 10^{-6}$)	0,0347 (0– 0,0682)	0 (0–0)	0,0216 (0– 0,0324)	0 (0–0)
	β	9,01 (8,33–10)	0,0593 (0– 0,114)	1,14 (0,851– 1,66)	3,3 (0–6,04)	2,03 (1,23–2,77)	0,304 (0,102– 0,539)
	μ	0,00716 (0,00247– 0,00989)	0,131 (0–0,262)	0,137 (0– 0,217)	0,0103 (0– 0,0171)	0 (0–0)	0,0903 (0– 0,171)
urojony	α	0,00233 (0,00132– 0,00263)	0 (0–0)	0,0528 (0,0246– 0,0922)	0,00276 (0– 0,00367)	0,0191 (0,00809– 0,0333)	0 (0–0)
	β	14,9 (7,15–20)	0 (0–0)	0,8 (0,56–1,01)	4,22 (0–7,8)	1,69 (0–2,42)	0 (0–0)
mieszany	α_-	0 (0–0)	1 (1–1)	0 (0–0)	0 (0–0)	0 (0–0)	0,83 (0,66– 1,28)
	α_+	0,0424 (0,00854– 0,0657)	0,0981 (0,0382– 0,165)	0,0877 (0,0711–0,146)	0,066 (0–0,106)	0,043 (0,0122– 0,0566)	0,106 (0,0549– 0,212)
	β	3,31 (2,94–3,82)	0,142 (0–0,223)	0,899 (0,551– 1)	1,82 (1,25–3,21)	1,37 (0,988– 1,59)	0,158 (0,0918– 0,206)
incyden- talny	α_ϕ	0 (0–0)	0,115 (0–0,231)	0,0241 (0– 0,0481)	0 (0–0)	0,113 (0–0,22)	0,515 (0,291– 0,995)
	α_p	0,01 (0–0,0155)	0,00411 (0– 0,00821)	0,00596 (0– 0,0119)	0,00838 (0,00264– 0,0109)	0,00466 (0– 0,00931)	0 (0–0)
	β	4,78 (4,2–5,26)	6,96 (6,42–10,2)	7,17 (0–10,6)	3,93 (3,26–4,27)	7,77 (1,39–9,99)	26,8 (3,7–48)
podwójnie mieszany	$\alpha_- \phi$	0,00644 (0– 0,00881)	0,576 (0,303– 1,14)	0,103 (0,0484– 0,187)	0,0038 (0– 0,00628)	0,0699 (0,039– 0,0924)	0,000356 (0– 0,000711)
	$\alpha_- p$	0 (0–0)	1 (1–1)	0 (0–0)	0 (0–0)	0 (0–0)	0,381 (0– 0,762)
	$\alpha_+ \phi$	0 (0–0)	0,908 (0,817– 1,82)	0 (0–0)	0,00117 (0– 0,00235)	0 (0–0)	0,712 (0,424– 1,42)
	$\alpha_+ p$	0,0197 (0– 0,0234)	0 (0–0)	0,00928 (0– 0,0145)	0,0191 (0,000903– 0,0285)	0,0123 (0,00253– 0,0198)	0,00167 (0– 0,00334)
	β	4,32 (3,4–5,04)	0,207 (0–0,276)	2,27 (1,61– 3,56)	3,63 (2,56–4,8)	3,18 (1,78–4,36)	0,575 (0– 0,971)
podwójnie mieszany z rozpadem Q	$\alpha_- \phi$	0,00508 (0,00236– 0,00678)	0,366 (0–0,674)	0,138 (0–0,26)	0,00277 (0– 0,00554)	0,103 (0,012– 0,135)	0,755 (0,515– 1)
	$\alpha_- p$	0 (0–0)	0,125 (0–0,25)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)
	$\alpha_+ \phi$	0 (0–0)	1 (1–1)	0 (0–0)	0,00669 (0– 0,0124)	0 (0–0)	1 (1–1)
	$\alpha_+ p$	0,0198 (0– 0,0242)	0 (0–0)	0,00449 (0– 0,00898)	0,0195 (0,00191– 0,0391)	0,0116 (0,001– 0,0213)	0 (0–0)
	β	4,43 (3,55–5,23)	0,985 (0–1,17)	2,04 (1,69– 2,28)	2,89 (0,801– 3,86)	2,17 (1,32–2,86)	0,847 (0–1,09)
	S	0 (0–0)	0,151 (0,0822– 0,186)	0,00174 (0– 0,00347)	0,000422 (0– 0,000844)	0 (0–0)	0,0486 (0– 0,069)
podwójnie mieszany z rozpadem β	$\alpha_- \phi$	0,00743 (0– 0,0113)	0,0245 (0– 0,0447)	0,105 (0,0545– 0,192)	0,0106 (0– 0,0192)	0,07 (0,0358– 0,0963)	0 (0–0)
	$\alpha_- p$	0,000405 (0– 0,000811)	$0,5 (3,23 \times 10^{-5}–1)$	0 (0–0)	$4,99 \times 10^{-6}$ (0– $9,99 \times 10^{-6}$)	0 (0–0)	0 (0–0)
	$\alpha_+ \phi$	0 (0–0)	0,277 (0–0,554)	0 (0–0)	0,00469 (0– 0,00849)	0 (0–0)	0,00223 (0– 0,00446)
	$\alpha_+ p$	0,0192 (0– 0,0246)	0,000173 (0– 0,000345)	0,0112 (0– 0,0191)	0,0182 (0– 0,0316)	0,0112 (0,000423– 0,0171)	0,27 (0–0,54)

model	para- ra- metr	alkohol I	alkohol II	alkohol+ sacharyna	sacharyna I	sacharyna III	woda
	β	4,47 (3,63–5,09)	3,18 (0,258–5,83)	2,49 (1,46–4,02)	3,47 (1,89–5,6)	2,87 (1,66–3,56)	4,42 (0–8,11)
	δ	$4,18 \times 10^{-5}$ (0– $8,37 \times 10^{-5}$)	0,1 (0–0,199)	0,000473 (0–0,000946)	0 (0–0)	0 (0–0)	1 (1–1)

*Wartości przedstawiono jako mediana i 95% przedział ufności. Liczebność wyniosła 13 osobników w grupie „alkohol+sacharyna” i po osobników 14 w pozostałych grupach.

W modelu „mieszanym” szacowanie wartości oczekiwanej nagrody odbywa się z wykorzystaniem α_+ dla opcji wybranej (wzrost wartości oczekiwanej nagrody) jak i niewybranej (spadek wartości oczekiwanej nagrody) w sytuacji uzyskania nagrody. Z kolei w przypadku nieuzyskania nagrody w analogiczny sposób na wielkość kroku wskazuje wartość α_- . Wartość α_- nie różniła się znacząco od α_+ dla grupy alkoholowej ($p = 0,973$), sacharynowej ($p = 0,0993$) ani z mieszaniną substancji ($p = 0,168$).

W rozwiązaniu „podwójnie mieszanym”, stanowiącym rozszerzenie modelu „mieszanego”, wartość oczekiwanej nagrody wzrasta przy uzyskaniu nagrody dla opcji wybranej (z wielkością kroku α_{+p}) i maleje w sytuacji alternatywnej do niej ($\alpha_{-\phi}$) oraz maleje w sytuacji braku nagrody dla opcji wybranej (α_{-p}) i wzrasta dla opcji alternatywnej ($\alpha_{+\phi}$). Wartości parametrów dotyczących pozytywnego wzmocnienia, czyli α_{+p} i β , wyniosły 0 dla większości osobników grupy alkoholowej II (9/14), sugerując, że za dopasowanie modelu do danych odpowiadały inne czynniki poza wzmocnieniem.

W celu określenia wrażliwości zwierząt na rezultat działania oraz znaczenie zakładanego i rzeczywistego rezultatu, dla modelu „podwójnie mieszanego”, zestawiono ze sobą wartości wielkości kroku wskazanych sytuacji. Analiza na podstawie wartości parametrów umożliwiła wnioskowanie o efekcie wartości działania oraz efekcie urojonego wyboru, w sytuacji uzyskania i braku nagrody (Tabela 5–2). Efekt wartości działania, określony jako różnica między α_{+p} i α_{-p} , był znaczący wyłącznie w grupie sacharynowej (Tabela 5–2). Wyższa wartość α_{+p} w porównaniu do α_{-p} wskazała na większą wielkość kroku dla zdarzenia związanego z rzeczywistym uzyskaniem dostępu do nagrody, w porównaniu do rzeczywistego braku nagrody, dla tej grupy.

Tabela 5–2 Różnice w parametrach α w modelu „podwójnie mieszanym”

	parametr referencyjny (a)	parametr badany (b)	Substancja	Sparowany test Wilcoxo- na p	Osobników o a>b	Osobników o b>a
Efekt wartości	α_{+p}	α_{-p}	alkohol	0,954	19/28	9/28
			sacharyna	0,0144	24/28	4/28
			alkohol+sacharyna	0,675	9/13	4/13
Efekt przewidywania w sytuacji nagrody	α_{+p}	$\alpha_{+\phi}$	alkohol	0,793	20/28	8/28
			sacharyna	0,00406	25/28	3/28
			alkohol+sacharyna	0,497	10/13	3/13
Efekt przewidywania w sytuacji braku nagrody	α_{-p}	$\alpha_{-\phi}$	alkohol	0,605	8/28	15/28
			sacharyna	0,0247	7/28	21/28
			alkohol+sacharyna	0,305	3/13	10/13

Warto zestawzić to z wynikiem dla modelu „wartościowego”, w którym wielkość kroku dla zdarzenia uzyskania nagrody (α_+) i braku nagrody (α_-) może być różna, natomiast nie założono aktualizacji na podstawie skutków niepodjętych działań. W modelu „wartościowym” różnica między α_- i α_+ była znamionna wyłącznie dla grupy z mieszaniną substancji (sparowany test Wilcoxa $p = 0,0234$), dla której wartość α_- była większa od α_+ dla większości (8/13) osobników, co wskazało na większą wrażliwość na brak nagrody niż dostęp do niej, dla tej grupy zwierząt. Niezgodność wyników dla tych modeli wskazuje, że dodanie urojonej aktualizacji, z rozdzieleniem na sytuację rzeczywistego i oczekiwanego skutku działania, w modelu „podwójnie mieszanym”, dostarcza dokładniejszego wyniku. Ponadto, w modelu „podwójnie mieszanym” efekt urojonego wyboru dla uzyskania nagrody, traktowany jako różnica między α_{+p} i $\alpha_{+\phi}$, był znamionny także w grupie sacharynowej (Tabela 5–2), wskazując na większe znaczenie rzeczywistego niż przypuszczalnego rezultatu. Podobnie, efekt dla urojonego ($\alpha_{-\phi}$), w porównaniu do rzeczywistego (α_{-p}), wyboru w sytuacji braku nagrody był znamionny także wyłącznie w grupie sacharynowej (Tabela 5–2).

Wskazuje to na większą wrażliwość zwierząt nagradzanych sacharyną na uzyskanie dostępu do nagrody w porównaniu do braku nagrody oraz większą wrażliwość na rzeczywisty dostęp do nagrody w porównaniu do przewidywanego. Opisany wynik pokazuje także większą wrażliwość na oczekiwany brak nagrody w porównaniu do rzeczywistego braku. Alternatywnie, uzyskany wynik można tłumaczyć także większą wrażliwością na rezultat działania w sytuacji dostępu do nagrody dla obu dostępnych opcji, a więc dominacją pary parametrów α_{+p} i $\alpha_{-\phi}$ nad α_{-p} i $\alpha_{+\phi}$. W tej analizie widoczne było zróżnicowanie wewnątrzgrupowe, dlatego w celu wnioskowania, posłużono się tendencjami określonymi dla większości grupy.

W modelu „podwójnie mieszanym”, „mieszanym” oraz „urojonym” założono, że jeśli podczas próby wybrana opcja dostarczyła nagrody to alternatywna opcja nie wiąże się z uzyskaniem nagrody. W sytuacji uzyskania dostępu do nagrody oznaczało to wzrost wartości oczekiwanej nagrody dla opcji wybranej i spadek dla opcji niewybranej. Alternatywnie, nagroda mogłaby być przyznawana przez środowisko jednocześnie w obu narożnikach, co wiązałoby się z wzrostem wartości oczekiwanej nagrody dla obu opcji, podczas próby zakończonej nagrodą, oraz spadku tej wartości dla obu opcji, w próbie zakończonej brakiem nagrody. Dlatego wprowadzono kontrolny model „incydentalny”, w którym wartość α_p odwołuje się do sytuacji rzeczywistego wyboru oraz α_ϕ do urojonego, w pojedynczej próbie. Optymalna wartość parametru α i β dla kilku osobników każdej z grup równała się górnemu lub dolnemu limitowi optymalizacji wskazując na trudności w znalezieniu wartości optymalnych parametrów (dla 9/28 alkoholowej, 6/13 z mieszaniną substancji, 8/28 sacharynowej i 9/14 kontrolnej). Ten model charakteryzowało także gorsze dopasowanie do danych ze względu na wysoki odsetek osobników o wartości ułamka poprawnych przewidywań równej i poniżej 0,5, a więc dopasowaniu gorszym niż dla losowej decyzji.

W kolejnym kroku do modelowania włączono zależność czasu i wyboru, na podstawie założenia o wystąpieniu zapomnienia. W tym celu model „podwójnie mieszanym” rozszerzono o

efekt rozpadu wartości oczekiwanej („rozpad Q”). W modelu „podwójnie mieszanym z rozkładem Q” tempo rozkładu wartości Q determinowane jest parametrem S, gdzie im większa wartość, tym szybszy spadek. W przeciągu 10 minut wartość oczekiwanej nagrody spadłaby o 1,2% od wartości maksymalnej 1, na podstawie mediany S wszystkich osobników wynoszącej 0,0012. Natomiast dla parametru S o wartości 0 założenia tego rozwiązania sprowadzają się do modelu „podwójnie mieszanego”, co dotyczyło większości grupy alkoholowej (16/28) i sacharynowej (16/28) oraz kilku osobników grupy z mieszaniną substancji (4/13) i kontrolnej (1/14). Wartości optymalnych parametrów dotyczących rzeczywistego wzmocnienia (α_+ i β) wyniosły 0 dla części osobników grupy alkoholowej (10/28), sacharynowej (10/28), z mieszaniną substancji (5/13) oraz kontrolnej (7/14), co oznaczono „*” na wykresach (Rycina 5-8). Sugeruje to, że za dopasowanie modelu do danych mogły odpowiadać inne czynniki poza wzmocnieniem.

Alternatywnie, wpływ czasu na decyzje mógł polegać na wzroście niepewności wraz z długością interwału do kolejnej próby. W modelu „podwójnie mieszanym” z rozpadem β w każdej próbie wartość parametru β maleje od wartości domyślnej β_0 , zależnie od czasu jaki minął od poprzedniej próby i współczynnika rozpadu δ (Tabela 5—1). W przeciągu 10 minut wartość β , określona na podstawie mediany δ wszystkich osobników (0,00053), spadłaby o 0,529% od wartości $\sim 3,44$ stanowiącej medianę β_0 dla wszystkich osobników. Wartość parametru δ była większa od 0 dla większości osobników w grupie alkoholowej (11/28) i sacharynowej (19/28), natomiast dla mniej niż połowy osobników grupy z mieszaniną substancji (5/13). Wartości parametrów związanych z rzeczywistym przyznaniem nagrody (α_+ i β) wynoszące 0 dotyczyły pojedynczych osobników (w grupie alkoholowej: 7/28, sacharynowej: 4/28, z mieszaniną substancji: 2/13). Podsumowując, wprowadzenie, przygotowanych na potrzeby niniejszych badań, założeń dotyczących zależności wyboru od czasu do kolejnej próby doprowadziło do nieznacznie lepszego dopasowania modelu „podwójnie mieszanego” dla części osobników w doświadczeniu.

W kolejnym kroku sprawdzono czy występują różnice w wartościach parametrów modeli zależne od rodzaju nagrody. Znamienne różnice zależne od substancji określono dla modelu „podstawowego”, „wartościowego”, „uwagowego” i „urojonego” (Tabela 5—3).

Tabela 5—3 Statystyka dla wartości parametrów modeli

Model	Parametr	Różnica między substancjami (alkohol, sacharyna i mieszanina substancji) test Kruskala-Wallisa
		wartość p
wzmocnienie	ε	0,746
podstawowy	α_+	0,00000125
	β	0,148
wartościowy	α_-	0,00351
	α_+	0,444
	β	0,731
uwagowy	α_0	0,0089
	β	0,275

Model	Parametr	Różnica między substancjami (alkohol, sacharyna i mieszanina substancji)
		test Kruskala-Wallisa wartość p
urojony	μ	0,066
	α	0,00014
	β	0,183
mieszany	α_{-}	0,216
	α_{+}	0,96
	β	0,301
incydentalny	α_{ϕ}	0,738
	α_p	0,944
	β	0,908
podwójnie mieszany	$\alpha_{-\phi}$	0,409
	α_{-p}	0,274
	$\alpha_{+\phi}$	0,612
	α_{+p}	0,633
	β	0,174
podwójnie mieszany + rozpad Q	$\alpha_{-\phi}$	0,519
	α_{-p}	0,0357
	$\alpha_{+\phi}$	0,953
	α_{+p}	0,975
	β	0,494
podwójnie mieszany + rozpad β	S	0,367
	$\alpha_{-\phi}$	0,0919
	α_{-p}	0,355
	$\alpha_{+\phi}$	0,938
	α_{+p}	0,502
	β	0,0543
	δ	0,0886

Należy wziąć pod uwagę, że różnice uzasadnione są wartością optymalnych parametrów związanych z rzeczywistym dostępem do nagrody wynoszącą 0, dla przynajmniej połowy osobników grupy alkoholowej II: w modelu „podstawowym” (α i β) dla 13/14 osobników, „wartościowym” (α_{+} i β) dla 7/14 osobników i „urojonym” (α i β) dla 13/14 osobników. Wskazuje to na niezależność dopasowania modelu od efektu wzmocnienia w grupie alkoholowej II, dla tych modeli.

W modelu „uwagowym” wielkość kroku bieżącego działania zależna jest każdorazowo od różnicy między oczekiwanym a rzeczywistym skutkiem działania, z założeniem, że im większy błąd przewidywania tym większa informatywność, i tym większa wielkość kroku dla danej próby uzyskania nagrody. Wartość α_0 w tym modelu wskazuje na domyślną wielkość kroku natomiast współczynnik μ na siłę zależności α_0 i błędu przewidywania. W modelu „uwagowym” wartość α_0 wyniosła 0 nie tylko dla połowy grupy alkoholowej II (7/14), ale też większości osobników grupy alkoholowej I (12/14) i sacharynowej I (8/14), co poddaje pod wątpliwość możliwość wykorzystania tego modelu do wytłumaczenia zachowania tych zwierząt.

W modelu „wartościowym” różnica między nagrodami dotyczyła parametru α_- , czyli sytuacji braku nagrody, i także była spowodowana wartością parametru wynoszącą 0 dla większości grupy alkoholowej II (9/14 osobników). Natomiast w przypadku różnicy dla modelu „podwójnie mieszanego z rozpadem Q” wartość parametru α_{-p} dla grupy alkoholowej II, w odróżnieniu od pozostałych grup, była większa od 0, sugerując występowanie efektu rzeczywistego braku nagrody (Tabela 5–1). Opisane wyniki sugerowały gorsze przewidywanie zachowania zwierząt grupy alkoholowej II przez zastosowane modele.

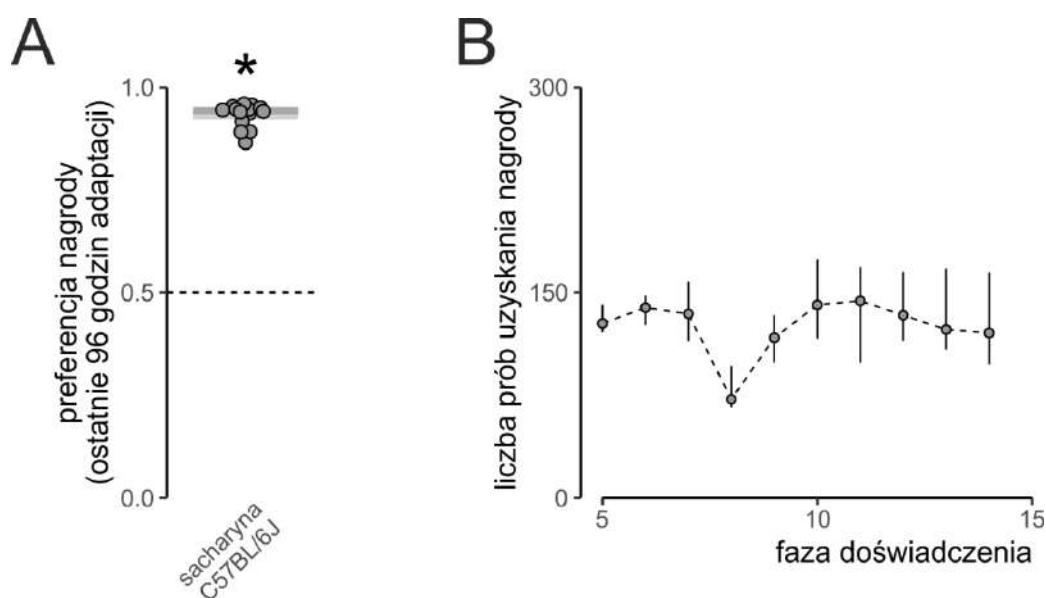
Podsumowując, w teście z probabilistycznym odwróceniem, u myszy szczepu C57BL/6J, zgodnie z oczekiwaniami, zaobserwowano uczenie się ze wzmocnieniem motywowane nagrodą. Zarówno częstość wyboru opcji o większym prawdopodobieństwie nagrody (Rycina 5-3) jak i częstość sięgania po nagrodę (Rycina 5-4, Rycina 5-5 i Rycina 5-6) różnicowały zwierzęta zależnie od rodzaju przyjmowanej nagrody, wskazując na słabsze właściwości wzmacniające samego alkoholu oraz rzadsze sięganie po niego, w porównaniu zarówno do roztworu sacharyny jak i mieszaniny sacharyny i alkoholu. Etanol jak i mieszanina etanolu i sacharyny działały wzmacniająco na zachowanie, chociaż nagroda ze słodzikiem (grupa sacharynowa i z mieszaniną substancji) charakteryzowała się większą preferencją u zwierząt. W odróżnieniu zarówno do grupy sacharynowej jak i grupy z mieszaniną alkoholu i sacharyny, w grupie alkoholowej zanotowano mniejszą preferencję opcji o wyższym prawdopodobieństwie uzyskania nagrody oraz, analogicznie, mniejszy efekt nagrody w przewidywaniu prawdopodobieństwa działania powtórzenia decyzji. Poza tym dla powyżej ¼ zwierząt grupy alkoholowej modelem o najlepszym dopasowaniu na podstawie ułamka poprawnych przewidywań było kontrolne rozwiązanie „incydentalne” (Rycina 5-9), co sugerowało zaburzone uczenie się. Natomiast analogicznych różnic nie zanotowano między grupą sacharynową i z mieszaniną substancji. Dlatego nie potwierdzono hipotezy o zaburzeniu procesu podejmowania decyzji pod wpływem alkoholu. Gorsze wykonanie zadania przez grupę alkoholową mogło wynikać z awersyjnego smaku alkoholu. Jedną z niewielu różnic między grupą sacharynową i z mieszaniną substancji był wynik na podstawie analizy wielkości kroku (α) w modelu „podwójnie mieszanym”, w którym wykazano silniejsze zróżnicowanie pomiędzy wartościami parametrów dla grupy sacharynowej. Natomiast stwierdzenie różnic w sposobie podejmowania decyzji jedynie na podstawie tego wyniku nie jest uzasadnione.

Nie określono dla zwierząt jednego optymalnego modelu naśladowującego zachowanie (Rycina 5-7 i Rycina 5-8), chociaż stosunkowo lepsze dopasowanie do danych uzyskano przy założeniu o zdolności zwierząt do przewidywania skutków niepodjętego działania wraz z możliwością zróżnicowania wrażliwości na pozytywną i negatywną lub neutralną informację zwrotną. Zanotowano zależność wyboru i czasu podjęcia próby uzyskania nagrody polegającą na proporcjonalnym wzroście prawdopodobieństwa powtórzenia poprzedniego działania wraz z długością interwału od ostatniej próby (Jabłońska i in., 2021). Natomiast pomimo występowania zależności wyboru od czasu mijającego od poprzedniego działania, (Rycina 5-4 i Rycina 5-5), przygotowane na potrzebę niniejszych badań, rozszerzenia modeli, polegające na włącze-

niu efektu zapominania oraz wzrostu niepewności wraz z czasem, mijającym od poprzedniej próby, spowodowały nieznaczne polepszenie dopasowania do danych.

5.2 Analiza uczenia ze wzmocnieniem zależna od wybranych receptorów w teście probabilistycznego odwrócenia w klatkach IntelliCage

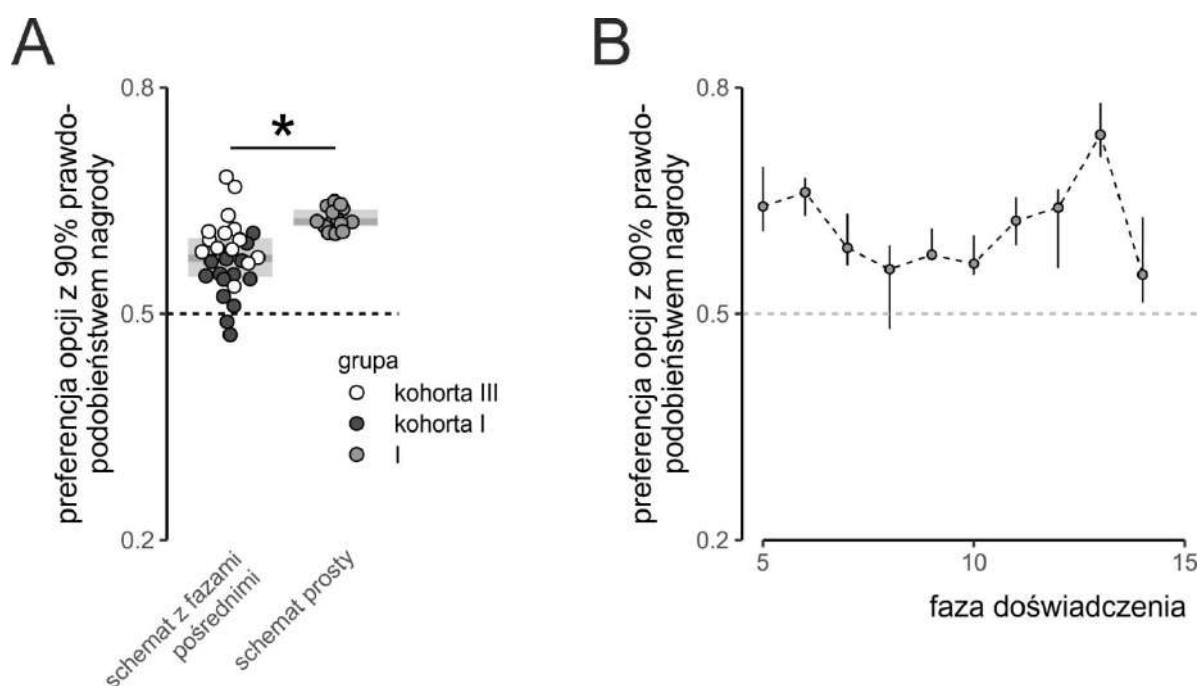
Schemat testu w poprzednim doświadczeniu charakteryzowało występowanie faz pośrednich, o jednakowym prawdopodobieństwie uzyskania nagrody w obu narożnikach aparatu ($30\% \times 30\%$ i $90\% \times 90\%$), co mogło utrudniać wykonanie zadania przez myszy. Dlatego, w celu optymalizacji testu, w kolejnym doświadczeniu zastosowałam schemat testu z probabilistycznym odwróceniem, w którym przypisanie prawdopodobieństwa uzyskania nagrody, w dwóch narożnikach klatki, określone jako faza, było zmieniane z 90% i 30% ($90\% \times 30\%$) na 30% i 90% ($30\% \times 90\%$) oraz odwrotnie przez 10 kolejnych, ~48-godzinnych, odwróceń (Tabela 4–2). Wstępne doświadczenie z takim schematem wykonano z udziałem myszy C57BL/6J i zastosowaniem nagrody w postaci $0,1\%$ (w./obj.) roztworu sacharyny. W etapie adaptacji prawdopodobieństwo uzyskania nagrody wynosiło 90% w obu narożnikach z dostępem do nagrody (Tabela 4–2). Zwierzęta preferowały sacharynę silnie od wody pod koniec etapu adaptacji (Rycina 5-10A) oraz stale wykonywały próby uzyskania dostępu do substancji w etapie odwróceń (Rycina 5-10B). Wysoka preferencja sacharyny oraz utrzymująca się na stałym poziomie liczba prób uzyskania nagrody potwierdziły nagradzające właściwości sacharyny.



Rycina 5-10 Preferencja substancji i liczba prób uzyskania nagrody w czasie doświadczenia. (A) Preferencja substancji jako proporcja liczby „liźnięć” butelek z substancją, względem wszystkich, w okresie ostatnich 96 godzin etapu adaptacji. Na wykresie długość prostokąta reprezentuje 50% środkowych obserwacji, ciemniejsza linia medianę, punkty wartości dla pojedynczych osobników. Przerywana linia wskazuje próg losowości. Znakiem „*” oznaczono wartość znamienne różną od $0,5$ (test Wilcoxona z wartością $0,5$, $p < 0,05$). Liczebność grupy to 14 osobników. (B) Liczba prób uzyskania nagrody dla kolejnych faz doświadczenia. Punkty wskazują medianę, zakresy 50% środkowych obserwacji. Pokazano obserwacje wyłącznie dla etapu

odwróceń. Liczebność grupy to 14 osobników. Mniejsza liczba prób w fazie 8 była spowodowana kilkugodzinnym brakiem rejestracji danych.

Występowanie faz pośrednich, w schemacie poprzedniego doświadczenia, mogło utrudniać test, co byłoby widoczne w mniejszej preferencji strony o wyższym prawdopodobieństwie nagrody. W celu porównania tej miary między doświadczeniami zestawiono dane z faz o zróżnicowanym prawdopodobieństwie uzyskania nagrody dla dwóch grup zwierząt poprzedniego doświadczenia, testowanych z rozwiązaniem sacharyny, z bieżącym doświadczeniem. Wartość preferencji strony o wyższym prawdopodobieństwie nagrody była niższa dla zwierząt testowanych w schemacie z fazami pośrednimi, w porównaniu do grupy zwierząt w bieżącym schemacie (bez faz pośrednich–schemat prosty) (Rycina 5-11A), co sugeruje, że występowanie faz pośrednich utrudniało test. W bieżącym doświadczeniu możliwe było rozpoznanie przez zwierzęta zmian faz, co skutkowałoby wartością preferencji opcji z większym prawdopodobieństwem nagrody zwiększającą się dla kolejnych faz. Nie zostało to zaobserwowane (Rycina 5-11B).

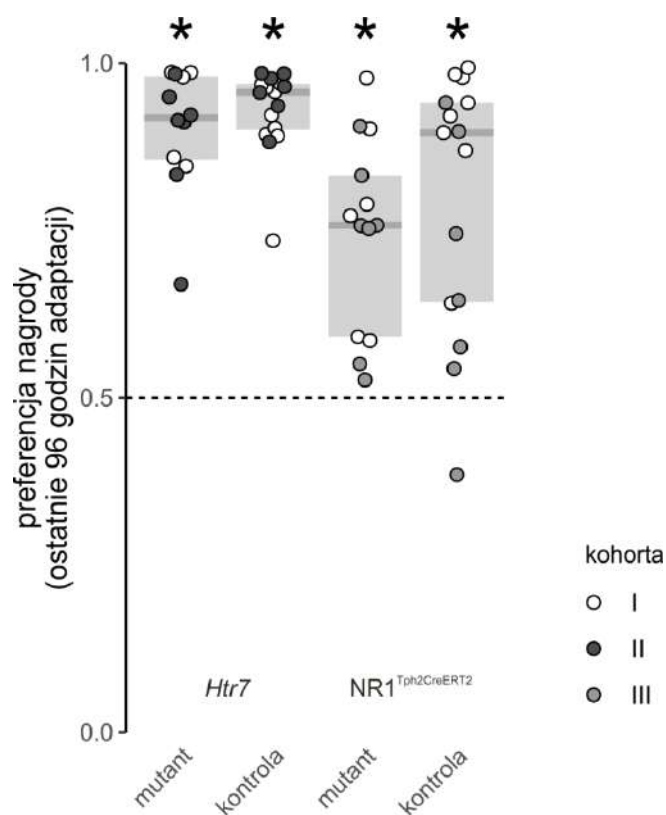


Rycina 5-11 Preferencja opcji o wyższym prawdopodobieństwie nagrody. Preferencja jako liczba prób uzyskania nagrody w narożniku o 90% prawdopodobieństwie uzyskania nagrody względem wszystkich prób, dla faz o zróżnicowanym prawdopodobieństwie uzyskania nagrody. (A) Na wykresie długość prostokąta reprezentuje 50% środkowych obserwacji, pionowa linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu kohortę. Znakiem „*” oznaczono statystycznie istotną różnicę (test Wilcozona, $p < 0,05$). Liczebność grup wyniosła po 14 osobników. (B) Preferencja strony o wyższym prawdopodobieństwie nagrody w zależności od fazy doświadczenia dla grupy zwierząt w doświadczeniu z prostym schematem. Na wykresie punkt wskazuje medianę, zakresy 50% środkowych obserwacji (granice pierwszego i trzeciego kwartylu). Liczebność grupy wyniosła 14 osobników. Wpływ na ten wynik mogło mieć skrócenie faz 8 i 9 związane z trudnościami technicznymi doświadczenia (Tabela 4–2).

Uproszczony schemat zastosowano do kolejnego doświadczenia, w którym celem było scharakteryzowanie strategii podejmowania decyzji oraz określenie zależności strategii od czasu decyzji dla samic myszy pobawionych receptorów 5-HT₇ (nazywane później *Htr7*) oraz sa-

mic myszy z selektywną indukowaną mutacją powodującą brak funkcjonalnych receptorów NMDA na komórkach serotoninowych (NR1^{Tph2CreERT2}). Do analiz włączono po dwie kohorty zwierząt obu szczepów (oznaczone jako I-III, chronologicznie). Zwierzęta były testowane w grupach mieszanych, czyli w skład grupy znajdującej się w jednej klatce IntelliCage wchodziły zarówno myszy z mutacją jak i myszy kontrolne tej samej linii. Dla wykluczenia wpływu modyfikacji genetycznej w grupach kontrolnych wskazanych linii oraz efektu mieszanej grupy, do porównań włączono wyniki dla opisanej wcześniej jednolitej grupy myszy C57BL/6J. Porównania przeprowadzono dla określenia wpływu genotypu (mutant i kontrola) oraz wykluczenia wskazanych czynników (pomiędzy grupą kontrolną danego szczepu i grupą myszy C57BL/6J). W tym doświadczeniu zastosowano, tak jak poprzednio, nagrodę w postaci 0,1% (w./obj.) roztworu sacharyny.

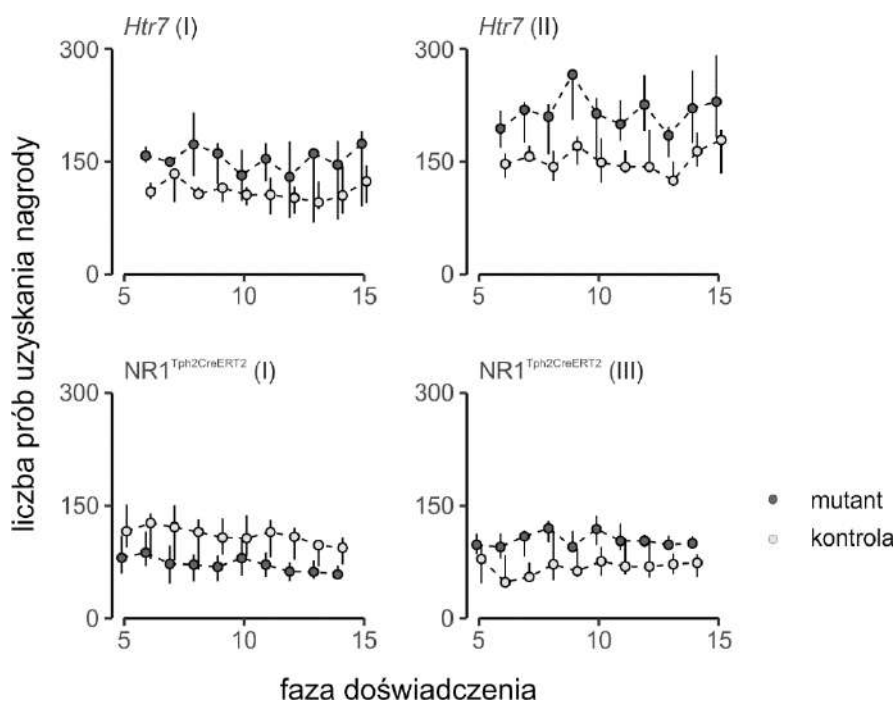
Nagradzające działanie substancji w każdej z badanych grup potwierdziła wartość preferencji przekraczająca 0,5 (Rycina 5-12). Genotyp nie wpłynął znacząco na preferencję nagrody dla grupy *Htr7* (test Wilcoxon $p = 0,664$) ani NR1^{Tph2CreERT2} ($p = 0,316$). Preferencja nagrody nie różniła się znacząco między grupą kontrolną *Htr7* i C57BL/6J ($p = 0,644$), w odróżnieniu od grupy kontrolnej NR1^{Tph2CreERT2} w porównaniu do myszy C57BL/6J ($p = 0,005$). Możliwe, że na wskazaną różnicę wpłynęło skrócenie etapu adaptacji do 4 faz w grupie zwierząt szczepu NR1^{Tph2CreERT2} (Tabela 4–2).



Rycina 5-12 Efekt mutacji na preferencję sacharyny. Preferencja nagrody jako wartość całkowitej liczby liżnięć butelek z nagrodą względem wszystkich liżnięć, w okresie ostatnich 96 godzin etapu adaptacji. Na wykresach pudełkowych długość prostokąta reprezentuje 50% 50% obserwacji, ciemniejsza linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu oznaczenie kohorty. Przerywana linia wskazuje próg wystąpienia preferencji. Znakem „*” oznaczono znaczącość

(test Wilcoxona z wartością 0,5). Liczebność grup to: „mutant *Htr7*”: 12, „kontrola *Htr7*”: 16, „mutant NR1^{Tph2CreERT2}”: 13 i „kontrola NR1^{Tph2CreERT2}”: 15 osobników.

Zwierzęta w trakcie testu stale wykonywały próby uzyskania nagrody (Rycina 5-13). Odbiegająca wartość opisująca aktywność zwierząt dotyczyła grupy *Htr7* kohorty II i była spowodowana przedłużeniem fazy 9 do 72 godzin, z powodu kilkunastogodzinnego braku rejestracji danych (Tabela 4–2). W grupie *Htr7* całkowita liczba prób uzyskania nagrody w etapie odwróceń zależała od genotypu (test Wilcoxona, $p = 0,0113$) oraz nie różniła się między grupą kontrolną i grupą myszy C57BL/6J ($p = 0,208$). Zwierzęta z mutacją wykonywały więcej prób uzyskania nagrody od kontrolnych. Z kolei dla zwierząt szczepu NR1^{Tph2CreERT2} wskazana wartość nie różnicowała genotypu ($p = 0,964$) natomiast była różna dla grupy kontrolnej w porównaniu do myszy C57BL/6J ($p = 0,007$). Zwierzęta tego szczepu wykonały mniej prób uzyskania nagrody w porównaniu do myszy C57BL/6J.

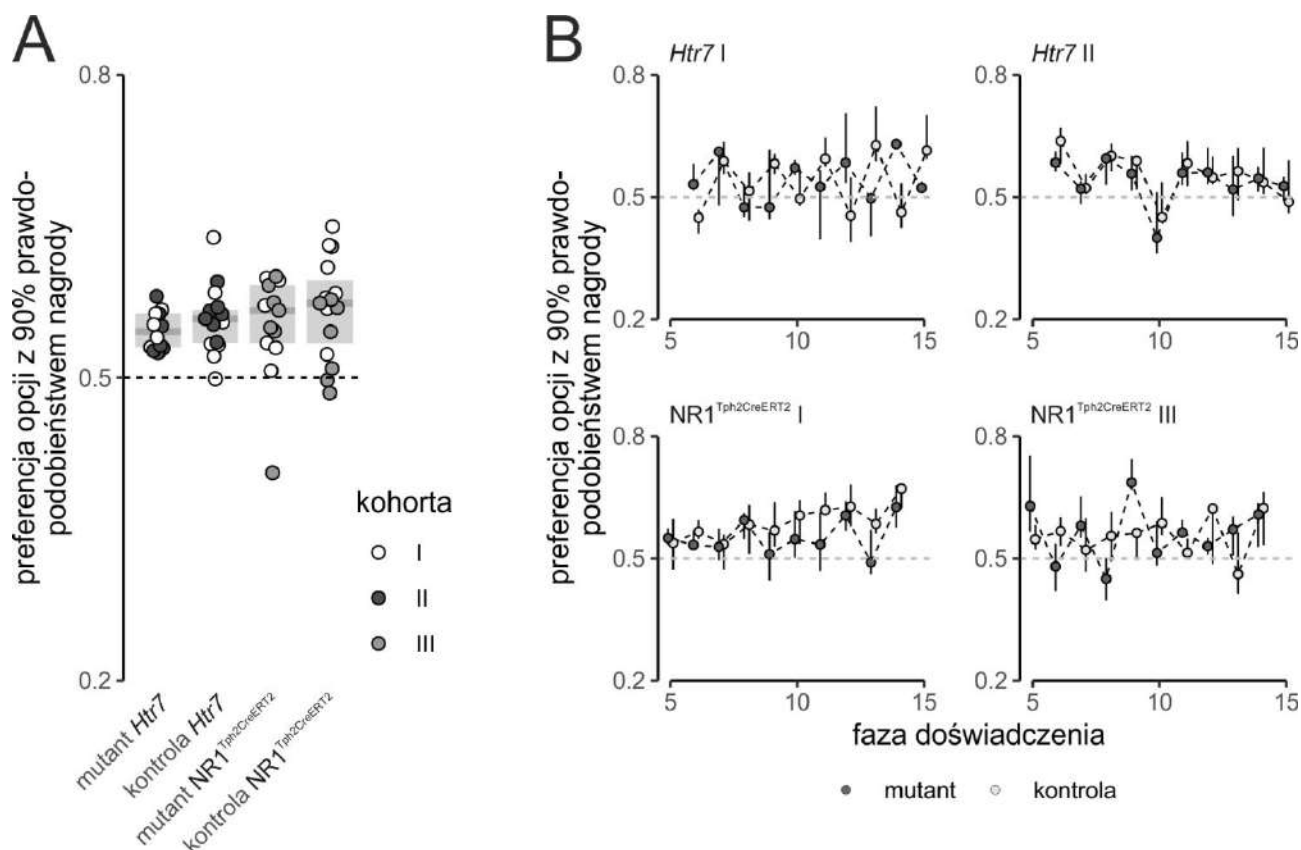


Rycina 5-13 Liczba prób uzyskania nagrody w etapie odwróceń. Liczba prób uzyskania nagrody dla kolejnych faz doświadczenia. Punkt wskazuje medianę, zakresy 50% środkowych obserwacji, odcień punktu genotyp. Dla przejrzystości pokazano obserwacje wyłącznie dla etapu odwróceń. Liczebność grup to: mutant *Htr7*: 12, kontrola *Htr7*: 16, mutant NR1^{Tph2CreERT2}: 13, kontrola NR1^{Tph2CreERT2}: 15 i C57BL/6J: 14 osobników.

Założeniem dotyczącym sposobu wykonania testu przez zwierzęta było występowanie zdolności do rozpoznania zróżnicowanego prawdopodobieństwa uzyskania nagrody pomiędzy dwoma dostępnymi opcjami i częstszy wybór opcji związanej z wyższym prawdopodobieństwem nagrody. W celu weryfikacji tej hipotezy określono czy proporcja wyborów między na-
 różnikami, w zależności od prawdopodobieństwa uzyskania w nich nagrody, przekracza granicę losowości na poziomie 50% liczby prób uzyskania nagrody. Preferencja opcji o wyższym prawdopodobieństwie uzyskania nagrody była znamienne różna od 0,5 dla każdej grupy (test

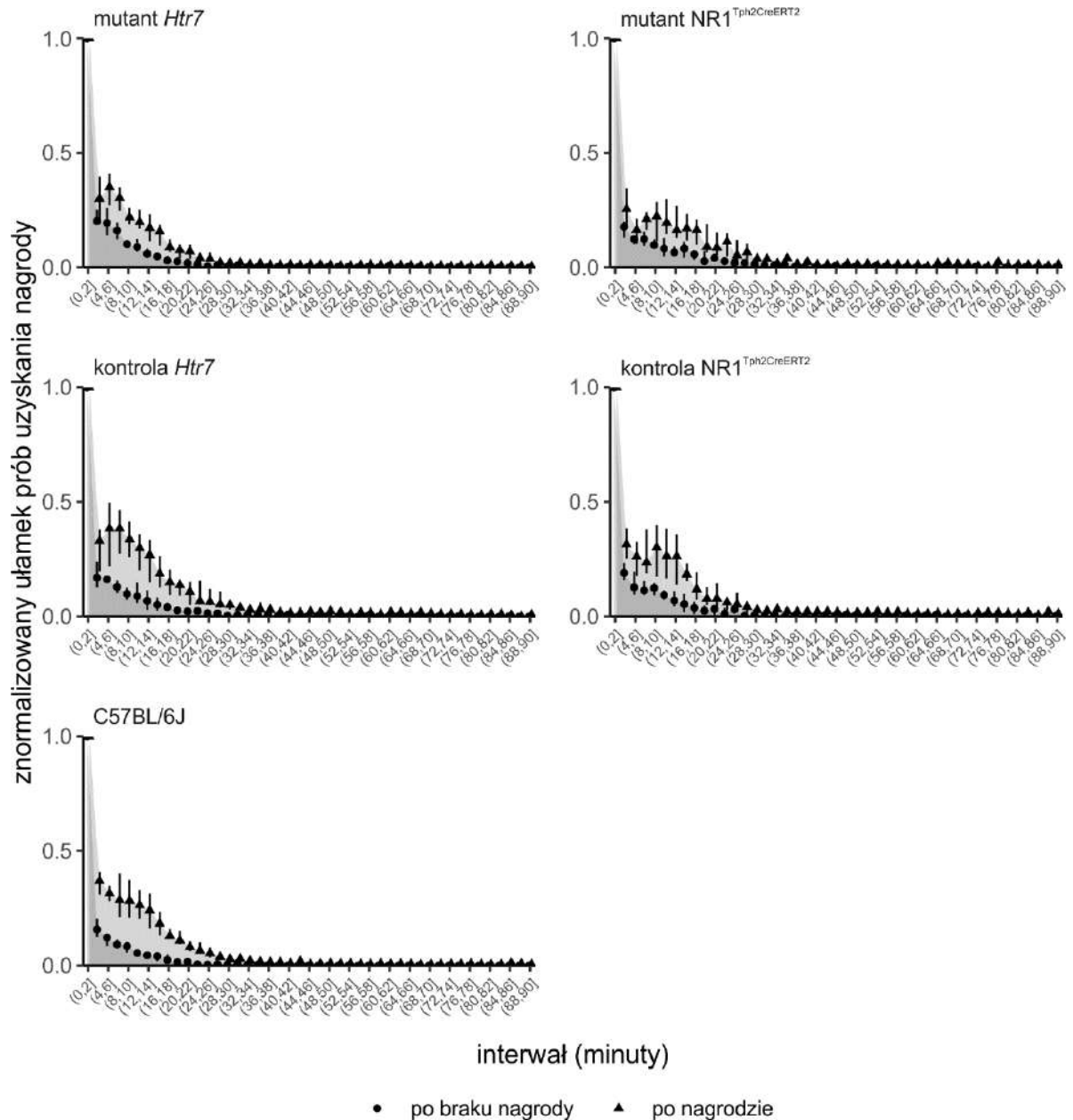
Wilcoxona z wartością 0,5, dla grupy mutantów *Htr7*: $p < 0,001$, kontrolnej *Htr7*: $p < 0,001$, mutantów $NR1^{Tph2CreERT2}$: $p = 0,0105$, kontrolnej $NR1^{Tph2CreERT2}$: $p < 0,001$ (Rycina 5-14). Ten wynik potwierdził działanie zwierząt nastawione na cel. Genotyp nie wpłynął na preferencję strony o wyższym prawdopodobieństwie nagrody dla myszy linii *Htr7* (test Wilcoxona, $p = 0,302$) ani $NR1^{Tph2CreERT2}$ ($p = 0,555$). Różnicę wykazano wyłącznie dla myszy C57BL/6J w porównaniu do grup kontrolnych *Htr7* ($p < 0,001$) i $NR1^{Tph2CreERT2}$ ($p = 0,002$). Preferencja strony o większym prawdopodobieństwie nagrody była w tych grupach mniejsza w porównaniu do myszy C57BL/6J (por. Rycina 5-11 i Rycina 5-14A).

Dodatkowa zależność opisanej preferencji od fazy doświadczenia wskazałaby na bardziej złożony wzorzec zachowania zwierząt. Spadek wartości co drugą fazę oznaczałby wystąpienie preferencji lokalizacji narożnika, natomiast stopniowy wzrost wartości dla kolejnych faz wskazałby adaptację do testu. Spadek co drugą fazę zauważalny był wyłącznie dla grupy *Htr7* kohorty I oraz grupy mutantów $NR1^{Tph2CreERT2}$ kohorty III (Rycina 5-14B). Jednak dla żadnej z grup nie wystąpiła skrajna indywidualna preferencja lokalizacji narożnika analogiczna do poprzedniego doświadczenia (Rycina 5-3). Wskazane wyniki potwierdziły zdolność zwierząt do wyboru opcji o pewniejszej nagrodzie. Alternatywne wytłumaczenie uzyskanych wyników stanowiłaby zmniejszona liczba prób uzyskania nagrody w fazach gdy niższe prawdopodobieństwo uzyskania nagrody było związane z faworyzowanym narożnikiem. Nie zaobserwowano takiej tendencji (Rycina 5-13).



Rycina 5-14 Preferencja opcji o wyższym prawdopodobieństwie nagrody. Preferencja jako liczba prób uzyskania nagrody w narożniku o 90% prawdopodobieństwie uzyskania nagrody względem liczby wszystkich prób, dla faz etapu odwróceń. (A) Na wykresie długość prostokąta reprezentuje 50% środkowych obserwacji, pionowa linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu kohortę. Przerywana linia wskazuje przyjętą granicę losowości. Liczebność grup to: mutant *Htr7*: 12, kontrola *Htr7*: 16, mutant $\text{NR1}^{\text{Tph2CreERT2}}$: 13 i kontrola $\text{NR1}^{\text{Tph2CreERT2}}$: 15 osobników. (B) Na wykresie punkt wskazuje medianę, zakresy 50% środkowych obserwacji, odcień punktu genotyp. Liczebność grup to: mutant *Htr7* kohorta I: 5, kohorta II: 7, kontrola *Htr7* kohorta I: 9, kohorta II: 7, mutant $\text{NR1}^{\text{Tph2CreERT2}}$ kohorta I: 6, kohorta III: 7 i kontrola $\text{NR1}^{\text{Tph2CreERT2}}$ kohorta I: 8, kohorta III: 7 osobników. Wartość odbiegającą od pozostałych zaobserwowano dla fazy 10 w grupie myszy *Htr7* II, co mogło być spowodowane kilkunastogodzinnym (12 godzin) brakiem rejestracji danych w poprzedzającej fazie (Tabela 4–2).

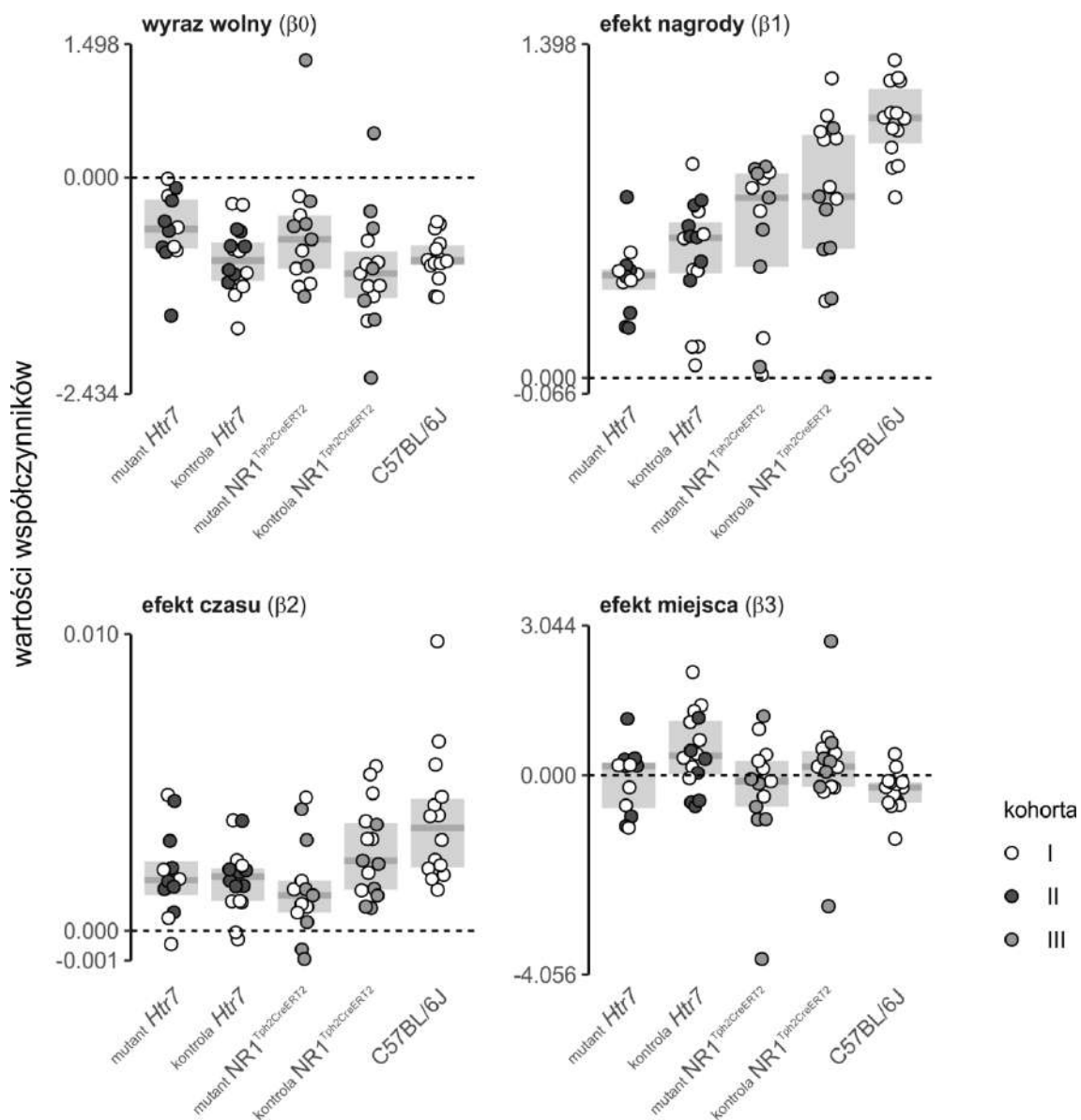
W kolejnym kroku analizowano czas upływający do kolejnej próby uzyskania nagrody. Częstość występowania prób uzyskania nagrody, w etapie odwróceń, w danym zakresie interwału, przedstawia Rycina 5-15. Na podstawie rozkładu częstości interwałów widoczne było, że zarówno po uzyskaniu nagrody jak i po działaniu nie zakończonym uzyskaniem nagrody, dominowało natychmiastowe (0–2 min) podjęcie kolejnego działania. Kształt rozkładu nie odbiegał od tendencji zaobserwowanej w poprzednim doświadczeniu (Rycina 5-4).



Rycina 5-15 Rozkład czasu mijającego do kolejnej próby uzyskania nagrody w 2-minutowych przedziałach. Ułamek liczby prób uzyskania nagrody w etapie odwróceń znormalizowany względem referencyjnej wartości maksymalnej liczby prób uzyskania nagrody. Przedział z największą liczbą prób uzyskania nagrody wyznaczono indywidualnie dla każdego osobnika. Liczba prób uzyskania nagrody dla tego zakresu służyła jako wartość referencyjna dla pozostałych zakresów. Punkt wskazuje medianę, zakresy 50% środkowych pomiarów. Interwał określa czas mijający do kolejnej próby uzyskania nagrody. Dla przejrzystości zakres interwału przedstawiono do wartości 90 minut i pole pod krzywą zaprezentowano w ciemniejszym odcieniu. Liczebność grup to: mutant *Htr7*: 12, kontrola *Htr7*: 16, mutant NR1^{Tph2CreERT2}: 13, kontrola NR1^{Tph2CreERT2}: 15 i C57BL/6J: 14 osobników.

Na podstawie wskazanych analiz widoczne było, że na wybory zwierząt wpłynęły: czas decyzji, skutek poprzedniego działania oraz indywidualna preferencja lokalizacji narożnika. Określona na podstawie regresji logistycznej (Rycina 5-16) szansa działania powtórzenia działania („zostań”) była większa w sytuacji dostępu do nagrody w poprzedniej próbie (β_1), wydłużenia interwału (β_2) i zależnie od lokalizacji narożnika w poprzedniej próbie (β_3). Ujemna wartość wyrazu wolnego w każdej grupie wskazała na domyślnie zwiększoną szansę działania

„zmień”. Prawdopodobieństwo działania „zostań”, określone na podstawie wyników regresji, wzrastało wraz z wartością interwału do kolejnej próby uzyskania nagrody zależnie od przyznania dostępu do nagrody oraz lokalizacji narożnika, dla każdej testowanej grupy zwierząt, natomiast efekt czasu mijającego do kolejnej próby widoczny był dla interwałów dłuższych od kilkunastu minut, podobnie jak w poprzednim doświadczeniu (Rycina 5-6). Z analizy wykluczono jednego osobnika z grupy alkoholowej I i jednego z grupy sacharynowej I, ponieważ nie skorzystały z obu dostępnych opcji, tj. narożników klatki IntelliCage (stosunki liczby prób uzyskania nagrody w narożnikach wyniosły 1 z 111 i 1 z 430 całkowitej liczby prób faz etapu odwróceń o zróżnicowanej szansie uzyskania nagrody, odpowiednio).

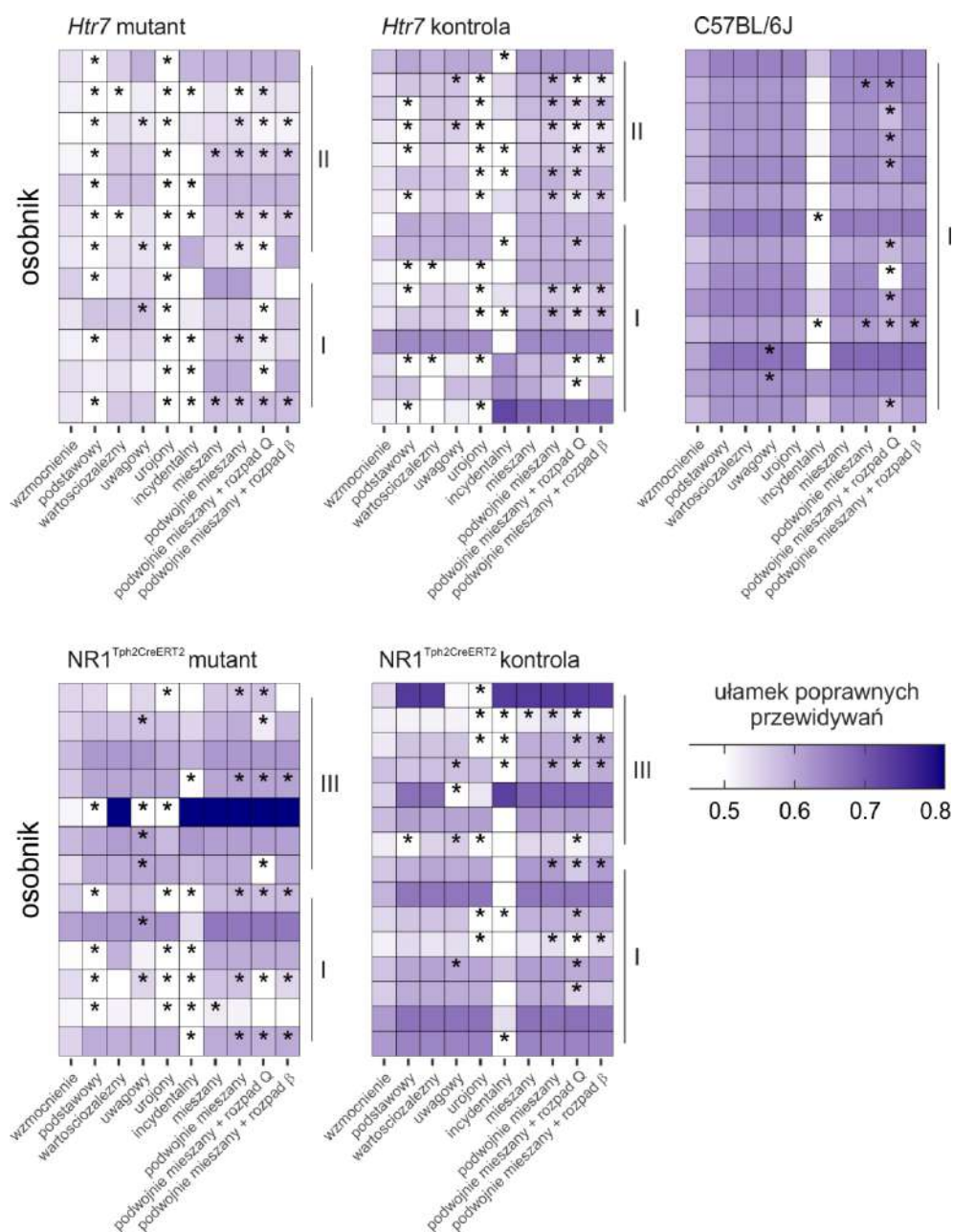


Rycina 5-16 Współczynniki regresji logistycznej. Na wykresach pudełkowych długość prostokąta reprezentuje 50% środkowych obserwacji, pionowa linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu oznaczenie kohorty. Przerywana linia wskazuje kierunek zmiany (wartość powyżej wskazuje zwiększoną szansę działania „zostań”). Liczebność grup to: mutant *Htr7*: 12, kontrola *Htr7*: 16, mutant *NR1*^{Tph2CreERT2}: 13, kontrola *NR1*^{Tph2CreERT2}: 15 i C57BL/6J: 14 osobników.

W kolejnym kroku sprawdzono czy genotyp wpłynął na siłę wskazanych efektów (Rycina 5-16). W linii *Htr7* genotyp nie wpłynął na wielkość efektu nagrody (test Wilcoxon $p = 0,121$), czasu ($p = 0,945$), ani preferencji lokalizacji narożnika ($p = 0,371$). Dla grupy $NR1^{TphCreERT2}$ znamiennej różnicę względem grupy kontrolnej wskazano dla wielkości efektu czasu do kolejnej próby ($p = 0,0325$), ale nie dla efektu nagrody ($p = 0,339$), ani lokalizacji narożnika ($p = 0,65$). Zgodnie ze wskazanym wynikiem prawdopodobieństwo działania „zostań” wzrastałoby wolniej w grupie z mutacją, w porównaniu do grupy kontrolnej. Natomiast w zachowaniu myszy C57BL/6J widoczny był większy efekt nagrody oraz czasu ale nie lokalizacji narożnika, w porównaniu zarówno do grupy kontrolnej *Htr7* ($\beta 1: p < 0,001$, $\beta 2: p = 0,003$, $\beta 3: p = 0,42$) jak i $NR1^{Tph2CreERT2}$ ($\beta 1: p < 0,001$, $\beta 2: p = 0,025$, $\beta 3: p = 0,872$). Sugerowało to większe znaczenie dostępu do nagrody oraz szybszy wzrost prawdopodobieństwa działania „zostań” w grupie myszy C57BL/6J w porównaniu do pozostałych grup.

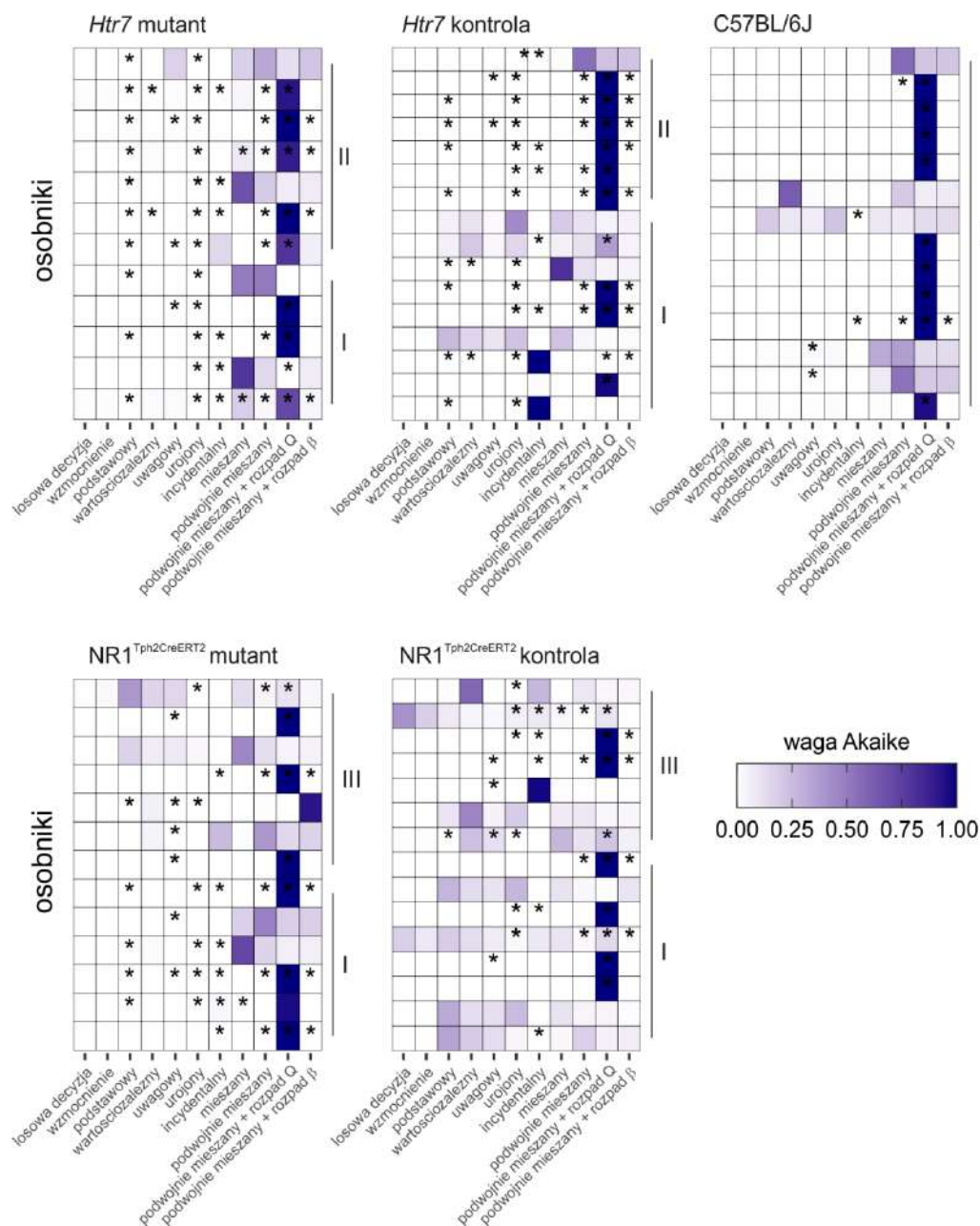
Dalszą analizę procesu podejmowania decyzji przeprowadzono z wykorzystaniem modeli obliczeniowych, analogicznie do poprzedniego doświadczenia. Do określenia dopasowania modelu do danych wykorzystano miary ułamka poprawnie przewidzianych działań oraz wartość AIC.

W pierwszym kroku wskazano modele, które nie przewidywały wyborów zwierząt lepiej niż losowe zgadywanie, więc o wartości ułamka poprawnych przewidywań równej i poniżej 0,5. Taki wynik dotyczył 58,6% (41/70) osobników w doświadczeniu dla modelu „incydentalnego”, 50% dla modelu „urojonego” i 34,3% dla modelu „podstawowego” oraz dla mniej niż 16% w pozostałych przypadkach (Rycina 5-17). Wskazuje to, że założenia tych modeli odbiegają od procesu decyzyjnego u badanych zwierząt. Wyrażone w taki sposób gorsze dopasowanie kontrolnego modelu „incydentalnego” sugeruje dodatkowo, że zwierzęta rozróżniają szansę uzyskania nagrody dla dostępnych opcji. Wartość ułamka poprawnych przewidywań była mniejsza bądź równa 0,5 w części przypadków zastosowania modeli w grupach z mutacją i kontrolnych (od 21/150 dla grupy kontrolnej $NR1^{Tph2CreERT2}$ do 37/120 dla grupy mutantów *Htr7*). Sugerowało to inny sposób działania tych zwierząt, w porównaniu do myszy C57BL/6J, dla których wskazany wynik był rzadki (8 z 140 przypadków, z czego 7 ze wskazanych 8 dla modelu „incydentalnego”).



Rycina 5-17 Ułamek poprawnych przewidywań modeli zastosowanych w badaniach. Odcień pola odpowiada wartości ułamka poprawnych przewidywań uzyskanych dla danego modelu. Rząd określa wartości pojedynczego osobnika, kolumna wartości danego modelu. Dla przejrzystości wartości niższe i równe 0,5 oznaczono jednakowo. Znak „*” określa przypadki dla których nie uzyskano dodatniej wartości optymalnych parametrów związanych z rzeczywistym wzmocnieniem. Wyniki przygotowano na podstawie prób z okresu odwróceń.

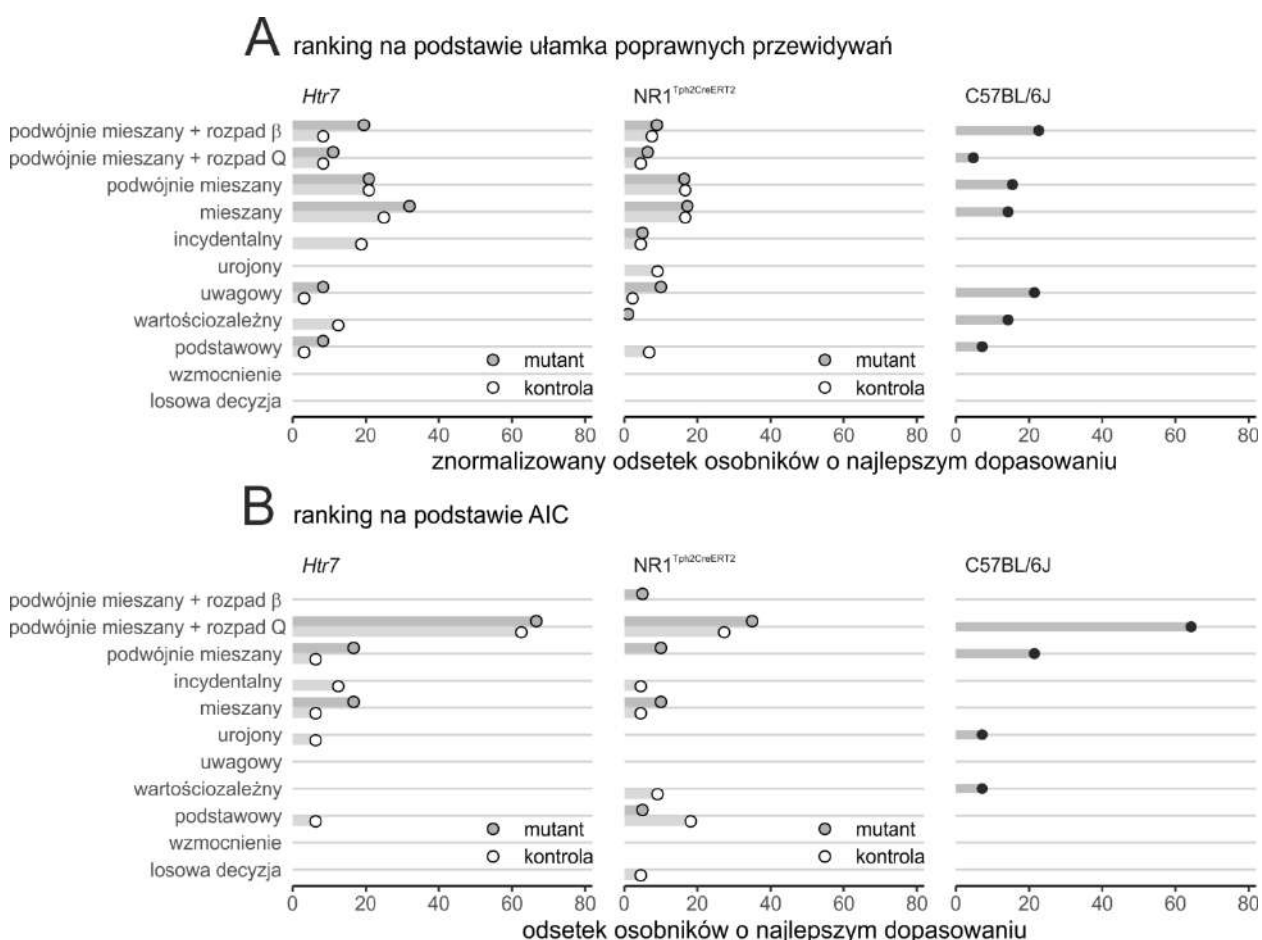
W kolejnym kroku do porównania dopasowania modeli do danych wykorzystano wagę Akaike w (Rycina 5-18). Wartość w wskazuje prawdopodobieństwo najlepszego dopasowania modelu względem pozostałych rozwiązań, gdzie prawdopodobieństwo całego zbioru sumuje się do wartości 1. Większe prawdopodobieństwo modelu względem pozostałych było widoczne dla większości osobników (40/70) dla modelu „podwójnie mieszanego z rozpadem Q”.



Rycina 5-18 Prawdopodobieństwo warunkowe modelu na podstawie wagi Akaike. Odcień pola odpowiada wartości w uzyskanej dla modelu. Rząd określa wartości pojedynczego osobnika (wartości sumują się do 1), kolumna wartości danego modelu. Znak „*” wskazuje przypadek wartości optymalnych parametrów związanych z rzeczywistym wzmoocnieniem wynoszących 0. Wyniki przygotowano na podstawie prób z okresu odwróceń.

W celu wskazania modelu o najlepszym dopasowaniu utworzono ranking modeli na podstawie liczby osobników, dla których dany model uzyskał najlepsze dopasowanie, czyli najwyższą wartość ułamka poprawnych przewidywań (Rycina 5-19A), oraz najniższą wartość AIC (Rycina 5-19B). Ułamek poprawnych przewidywań był równy dla dwóch lub więcej modeli dla 27/70 zwierząt w doświadczeniu. Dlatego, analogicznie do poprzedniego doświadczenia, zastosowano korektę w postaci wag. Pod względem ułamka poprawnych przewidywań do dwóch najlepszych rozwiązań należały w grupie mutantów *Htr7* oraz $\text{NR1}^{\text{Tph2CreERT2}}$ modele: „mieszany” i „podwójnie mieszany”, podobnie jak w grupach kontrolnych wskazanych linii. Natomiast w

grupie myszy C57BL/6J dwoma najlepszymi modelami były „podwójnie mieszany z rozpadem β ” i „uwagowy”. Pod względem wartości AIC najlepiej dopasowanym modelem był dla każdej z grup „podwójnie mieszany z rozpadem Q”. Na drugim miejscu znalazły się modele: „mieszany”, „podwójnie mieszany”, „podstawowy” i „incydentalny”, zależnie od grupy. Analiza na podstawie ułamka poprawnych przewidywań oraz wartości AIC wskazała dominujące dopasowanie modeli zakładających możliwość przewidywania rezultatu niewybranej opcji (urojona aktualizacja).



Rycina 5-19 Ranking modeli na podstawie odsetka grupy z największym dopasowaniem. Punkt wskazuje wartość odsetka grupy o największym dopasowaniu danego modelu. Pokazano wartości dla wszystkich zastosowanych modeli. Dla wyników na podstawie ułamka poprawnych przewidywań zastosowano normalizację w postaci przypisania wagi zależnej od liczby modeli o jednakowej wartości ułamka poprawnych przewidywań dla osobnika. Oznacza to, że jeśli dwa modele uzyskały jednakową wartość to waga dla każdego z nich wynosi 1/2. Liczebność grup to: mutant *Htr7*: 12, kontrola *Htr7*: 16, mutant *NR1^{Tph2CreERT2}*: 13 i kontrola *NR1^{Tph2CreERT2}*: 15 osobników.

W celu dalszej analizy zachowania zwierząt przeanalizowano optymalne wartości parametrów (Tabela 5–4). Optymalne wartości parametrów związanych z rzeczywistym przyznaniem nagrody wyniosły 0 dla części osobników każdej z grup (Rycina 5-18). Wskazuje to trudności w znalezieniu optymalnych wartości parametrów i sugeruje zachowanie zwierząt trudne do wytłumaczenia za pomocą założenia o wystąpieniu uczenia ze wzmocnieniem. Wtedy za zdolność modelu do przewidywania zachowania zwierząt odpowiadałyby inne czynniki, poza

wzmacniającym działaniem nagrody. Wartości optymalne parametrów związane z rzeczywistym wzmocnieniem nieprzekraczające wartości 0 uzyskano głównie dla modelu „podstawowego”, „urojonego”, „podwójnie mieszanego” i „podwójnie mieszanego z rozpadem Q” (Tabela 5–4).

Tabela 5–4 Wartości optymalne parametrów modeli

model	parameter	mutant <i>Htr7</i>	kontrola <i>Htr7</i>	mutant NR1 ^{Tph2CreERT2}	kontrola NR1 ^{Tph2CreERT2}	C57BL/6J
wzmocnienie	ε	0,939 (0,914–0,949)*	0,931 (0,914–0,966)	0,904 (0,886–0,928)	0,885 (0,858–0,926)	0,814 (0,792–0,844)
podstawowy**	α	0 (0–0)	0,00771 (0–0,0154)	0,0178 (0–0,0357)	0,134 (0,0351–0,173)	0,133 (0,0549–0,151)
	β	0 (0–0)	0,222 (0–0,445)	0,722 (0,324–1,44)	0,772 (0,259–1,03)	1,04 (0,865–1,13)
wartościowo zależny	α_-	0,222 (0,123–0,369)	0,149 (0,0543–0,282)	0,0316 (0–0,0512)	0,125 (0–0,189)	0,191 (0,0949–0,272)
	α_+	0,00528 (0–0,00792)	0,0335 (0–0,0644)	0,109 (0–0,153)	0,126 (0,0262–0,239)	0,137 (0,0569–0,195)
	β	1,19 (0–1,78)	1,31 (0–2,15)	1,13 (0–1,76)	1,37 (0–1,91)	1,05 (0,813–1,11)
uwagowy	α_0	0,0168 (0–0,0324)	0,105 (0,0343–0,165)	0,00589 (0–0,0118)	0,0909 (0,0274–0,182)	0,103 (0,0423–0,175)
	β	0,618 (0,366–0,896)	0,548 (0,0507–0,705)	0,722 (0–1,14)	0,772 (0,336–1,08)	1,12 (0,687–1,25)
	μ	0,117 (0–0,201)	0,107 (0–0,214)	0,0382 (0–0,0764)	0,101 (0–0,203)	0,0016 (0–0,0032)
urojony**	α	0 (0–0)	0 (0–0)	0,0222 (0–0,0444)	0,0635 (0,041–0,127)	0,0638 (0,0406–0,0684)
	β	0 (0–0)	0 (0–0)	0,558 (0,174–1,12)	0,612 (0,101–1,22)	1,01 (0,876–1,09)
mieszany	α_-	0,624 (0,248–1)	0,0626 (0–0,125)	0 (0–0)	0,0589 (0–0,0959)	0,0704 (0,0411–0,104)
	α_+	0,0337 (0,00343–0,0638)	0,0254 (0,00273–0,0409)	0,0264 (0–0,04)	0,0676 (0,0578–0,0887)	0,0699 (0,0547–0,0935)
	β	0,349 (0–0,465)	0,676 (0–0,98)	1,04 (0,619–1,65)	0,634 (0,0895–0,741)	1,03 (0,904–1,17)
incydentalny***	α_ϕ	0,124 (0–0,246)	0,00645 (0–0,0129)	0,32 (0–0,626)	0,0861 (0–0,169)	0,00978 (0–0,0182)
	α_p	0,00138 (0–0,00276)	0,00287 (0–0,00573)	0,00529 (0–0,0106)	0,0205 (0–0,0411)	0,0458 (0,0303–0,052)
	β	11 (0–18,6)	4,23 (0–4,72)	7,47 (0–7,69)	20,9 (0–37,9)	3,98 (0–4,2)
podwójnie mieszanym	$\alpha_{-\phi}$	0,0397 (0–0,0753)	0,0767 (0–0,151)	0,0189 (0–0,0378)	0,0845 (0–0,169)	0,073 (0–0,124)
	α_{-p}	0,3 (0–0,599)	0,146 (0–0,292)	0,000801 (0–0,0016)	0,152 (0–0,303)	0,0632 (0–0,111)
	$\alpha_{+\phi}$	0,144 (0–0,288)	0,0795 (0–0,159)	0,0231 (0–0,0462)	0,0971 (0–0,192)	0,0344 (0–0,0635)
	α_{+p}	0 (0–0)	0,00157 (0–0,00314)	0,00609 (0–0,0122)	0,0119 (0–0,0237)	0,0132 (0–0,0218)
	β	0,967 (0–1,63)	1,83 (0–3,25)	1,78 (0,241–2,6)	1,35 (0–1,89)	1,58 (0–1,9)
podwójnie mieszanym z rozpadem Q**	$\alpha_{-\phi}$	0,0667 (0–0,133)	0,141 (0–0,275)	0,0195 (0–0,0389)	0,167 (0–0,258)	0,344 (0–0,538)
	α_{-p}	0,0677 (0–0,135)	0 (0–0)	0 (0–0)	0,239 (0–0,474)	0,0764 (0–0,147)

model	para- ra- met- r	mutant <i>Htr7</i>	kontrola <i>Htr7</i>	mutant NR1 ^{Tph2CreERT2}	kontrola NR1 ^{Tph2CreERT2}	C57BL/6J
	$\alpha_{+\phi}$	1 (1–1)	0,985 (0,97–1)	0,88 (0,761–1)	0,6 (0,2–1)	0,657 (0,314–1)
	α_{+p}	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)
	β	1,16 (0,875– 1,34)	1,87 (0,719– 2,59)	1,93 (0,523– 2,27)	1,51 (0–1,64)	2,17 (0,897– 2,57)
	s	0,149 (0–0,262)	0,142 (0,0782– 0,284)	0,0455 (0–0,091)	0,0423 (0– 0,0846)	0,0828 (0,0307– 0,164)
	$\alpha_{-\phi}$	0,00662 (0– 0,0132)	0,0943 (0–0,189)	0,0189 (0– 0,0316)	0,0782 (0–0,156)	0,0481 (0– 0,0853)
	α_{-p}	0,461 (0–0,921)	0,661 (0,323–1)	0 (0–0)	0,156 (0–0,312)	0,063 (0–0,108)
podwójnie mie- szany z rozpadem β^{***}	$\alpha_{+\phi}$	0,309 (0–0,617)	0,721 (0,552– 1,44)	0,0178 (0– 0,0355)	0,155 (0–0,299)	0,0678 (0–0,128)
	α_{+p}	0,00857 (0– 0,0171)	0,000646 (0– 0,00129)	0,0118 (0– 0,0236)	0,00813 (0– 0,0163)	0,0519 (0– 0,0956)
	β_0	0,758 (0–1,1)	1,27 (0–1,84)	1,63 (0–2,52)	1,35 (0–1,89)	1,6 (0–2,03)
	δ	0,00354 (0– 0,00546)	0,0258 (0– 0,0502)	0,000929 (0– 0,00186)	0,00103 (0– 0,00207)	0,000147 (0– 0,000294)

(*) Wartości przedstawiono jako mediana i 95% przedział ufności. Liczebność grup to: mutant *Htr7*: 12, kontrola *Htr7*: 16, mutant NR1^{Tph2CreERT2}: 13, kontrola NR1^{Tph2CreERT2}: 15 i C57BL/6J: 14 osobników. (**) W modelu „podstawowym” wartość α i β wyniosła 0 dla większości grupy mutantów (10/12) i połowy grupy kontrolnej *Htr7* (8/16) oraz wyłącznie kilku osobników grupy NR1^{Tph2CreERT2} (5/13 mutantów i 1/15 kontroli). Także w modelu „urojonym” dla linii zwierząt *Htr7* określono wartości parametrów równe 0 dla wszystkich osobników grupy mutantów i większości osobników kontrolnych (11/16). W modelu „podwójnie mieszanym z rozpadem Q” wartość α_{+p} związana z rzeczywistym uzyskaniem dostępu do nagrody była równa 0 dla większości osobników grupy mutantów *Htr7* (9/12), kontroli *Htr7* (11/16), mutantów NR1^{Tph2CreERT2} (7/13), kontroli NR1^{Tph2CreERT2} (9/15) oraz myszy C57BL/6J (9/14). W modelu „podwójnie mieszanym z rozpadem β ” wartość δ wyniosła 0 dla kilku osobników każdej z grup (2/12 mutantów *Htr7*, 4/16 kontroli *Htr7*, 6/13 mutantów NR1^{Tph2CreERT2}, 6/15 kontroli NR1^{Tph2CreERT2} oraz 7/14 myszy C57BL/6J). (***) W modelu „incydentalnym” optymalna wartość parametru β równała się górnemu ograniczeniu optymalizacji (50) dla kilku osobników w każdej grupie (4/12 mutantów *Htr7*, 3/16 kontroli *Htr7*, 2/13 mutantów NR1^{Tph2CreERT2}, 6/15 kontroli NR1^{Tph2CreERT2} i 3/14 myszy C57BL/6J).

Dla modelu „podwójnie mieszanego” zestawiono ze sobą wielkości kroku poszczególnych sytuacji dla określenia czy zdarzenie uzyskania nagrody ma większe znaczenie od rezultatu działania w postaci braku nagrody (α_{+p} względem α_{-p}) oraz czy wielkość kroku zdarzenia rzeczywistego będzie większa od założonego wyniku dla opcji alternatywnej (α_{+p} względem $\alpha_{+\phi}$ oraz α_{-p} względem $\alpha_{-\phi}$) (Tabela 5–5). Efekt rezultatu działania wyrażony różnicą α_{+p} i α_{-p} był znamienny wyłącznie w grupie kontrolnej *Htr7* i kontrolnej NR1^{Tph2CreERT2}, dla których wartość α_{+p} była mniejsza od α_{-p} , wskazując na większe znaczenie uzyskania dostępu do nagrody niż jej braku. Efekt przewidywanego i rzeczywistego uzyskania dostępu do nagrody, określony jako różnica między α_{+p} i $\alpha_{+\phi}$, w omawianym modelu, był znamienne różny także wyłącznie w grupie kontrolnej *Htr7* i kontrolnej NR1^{Tph2CreERT2}, dla których wartość α_{+p} była mniejsza od $\alpha_{+\phi}$ dla większości osobników. Oznaczało to, że wielkość kroku dla działania z przewidywaną nagrodą była większa niż rzeczywiste uzyskanie dostępu do nagrody. Natomiast nie wykazano znamiennych efektów wyrażonych różnicami w wartościach parametrów α dla grupy zwierząt C57BL/6J. Wskazane tendencje różniły się od wyniku uzyskanego dla prostszego modelu „wartościowego”, w którym wielkość kroku dla uzyskania dostępu do nagrody (α_{+}) nie była

znamiennie różna od wielkości kroku przy braku nagrody (α_-) dla żadnej z grup (sparowany test Wilcozona, mutant *Htr7*: $p = 0,185$; kontrola *Htr7*: $p = 0,0687$; mutant $NR1^{Tph2CreERT2}$: $p = 0,327$; kontrola $NR1^{Tph2CreERT2}$: $p = 1$; kontrola C57BL/6J: $p = 0,426$), podobnie do modelu „mieszanego” (mutant *Htr7*: $p = 0,064$; kontrola *Htr7*: $p = 0,117$; mutant $NR1^{Tph2CreERT2}$: $p = 0,376$; kontrola $NR1^{Tph2CreERT2}$: $p = 0,252$; kontrola C57BL/6J: $p = 0,583$).

Tabela 5–5 Różnice w parametrach α w modelu "podwójnie mieszanym"

	parametr bieżący (a)	parametr referencyjny (b)	Grupa	Sparowany test Wilcozona p	Osobników o a>b	Osobników o b>a
Efekt rezultatu działania	α_{+p}	α_{-p}	mutant <i>Htr7</i>	0,118	5/12	6/12
			kontrola <i>Htr7</i>	0,0182	5/16	10/16
			mutant $NR1^{Tph2CreERT2}$	0,889	7/13	6/13
			kontrola $NR1^{Tph2CreERT2}$	0,0287	5/15	10/15
			C57BL/6J	0,217	5/14	9/14
Efekt przewidywania w sytuacji dostępu do nagrody	α_{+p}	$\alpha_{+\phi}$	mutant <i>Htr7</i>	0,0771	5/12	7/12
			kontrola <i>Htr7</i>	0,0361	6/16	10/16
			mutant $NR1^{Tph2CreERT2}$	0,839	8/13	5/13
			kontrola $NR1^{Tph2CreERT2}$	0,0413	5/15	10/15
			C57BL/6J	1	8/14	6/14
Efekt przewidywania w sytuacji braku nagrody	α_{-p}	$\alpha_{-\phi}$	mutant <i>Htr7</i>	0,266	6/12	5/12
			kontrola <i>Htr7</i>	0,0516	9/16	5/16
			mutant $NR1^{Tph2CreERT2}$	0,455	7/13	6/13
			kontrola $NR1^{Tph2CreERT2}$	0,164	10/15	5/15
			C57BL/6J	0,67	7/14	7/14

Dla określenia efektu czasu, do modelu „podwójnie mieszanego” dołączono rozszerzenie dotyczące spadku wartości oczekiwanej nagrody w czasie, czyli „rozpad Q”. W takim modelu tempo spadku wartości oczekiwanej nagrody determinowane było przez wartość S. Rozpad wartości oczekiwanej nagrody, od wartości równej 1, wyrażony medianą dla wszystkich osobników w doświadczeniu (0,0802), wyniósłby 55,16% w ciągu 10 minut. W modelu „podwójnie mieszanym” wartość α_{+p} związana z rzeczywistym uzyskaniem dostępu do nagrody była równa 0 dla większości osobników wszystkich grup. Sugerowało to, że za dopasowanie modelu do danych odpowiadają inne czynniki, w tym efekt czasu. Dlatego do modelu „podwójnie mieszanego” włączono efekt czasu działania w postaci efektu wzrostu niepewności wraz z czasem upływającym od poprzedniej próby uzyskania nagrody. W modelu „podwójnie mieszanym z rozpadem β ” wartość parametru β spadałaby o 1,902% swojej wartości domyślnej (β_0) po 10 minutach, zakładając wartość δ i β_0 jako medianę dla wszystkich osobników w doświadczeniu (0,00192, 1,43, odpowiednio).

W modelu „uwagowym” wartość wielkości kroku α zależy każdorazowo od błędu przewidywania, a siłę tej zależności determinuje parametr μ . W odróżnieniu od poprzedniego doświadczenia wartość parametru μ większą od zera określono dla przynajmniej połowy osobników każdej z grup (10/12 mutantów *Htr7*, 11/16 kontroli *Htr7*, 9/13 mutantów $NR1^{TphCreERT2}$,

10/15 kontroli NR1^{Tph2CreERT2} i 7/14 myszy C57BL/6J). W pozostałych przypadkach założenia modelu pokryły się z modelem „podstawowym”. Mediana μ dla wszystkich osobników wskazała na 5,83% zależność wartości α bieżącego wyboru od błędu przewidywania.

W kolejnym kroku określono różnice w wartościach parametrów zależne od genotypu (Tabela 5–6). Znamienne różnice w wartościach parametrów dotyczyły linii *Htr7* dla modelu „podstawowego”, „urojonego” i „podwójnie mieszanego z rozpadem Q” oraz linii NR1^{Tph2CreERT2} dla modelu „podstawowego” i „podwójnie mieszanego z rozpadem Q”. Wystąpiły także różnice dla grup kontrolnych obu linii w porównaniu do myszy C57BL/6J. Na uzyskane różnice wpłynęły wartości parametrów równe 0 dla części osobników w grupach (Tabela 5–4).

Tabela 5–6 Znamienność różnic wartości parametrów

		p test Wilcoxona (genotyp <i>Htr7</i>)*	p test Wilcoxona (genotyp NR1 ^{Tph2CreERT2})*	p test Wilcoxona (C57BL/6J i kontro- la <i>Htr7</i>)*	p test Wilcoxona (C57BL/6J i kontro- la NR1 ^{Tph2CreERT2})*
wzmocnienie	ε	0,265	0,44	0,000181	0,005
podstawowy	α	0,0423	0,0171	0,012	0,813
	β	0,0617	0,367	0,000849	0,217
wartościозależny	α_{-}	0,5	0,49	0,924	0,924
	α_{+}	0,416	0,618	0,16	0,451
	β	0,871	0,294	1	1
uwagowy	α_0	0,0731	0,572	1	1
	β	0,982	0,856	0,018	0,031
	μ	0,657	0,761	0,194	0,206
urojony	α	0,0403	0,362	0,002	0,525
	β	0,0403	0,701	0,000948	0,067
mieszany	α_{-}	0,679	0,178	1	1
	α_{+}	0,693	0,24	0,052	0,88
	β	0,053	0,856	0,8	0,651
incydentalny	α_{ϕ}	0,468	0,378	1	1
	α_{ρ}	0,518	0,274	0,094	0,488
	β	0,468	0,456	0,638	0,638
podwójnie miesza- ny	$\alpha_{-\phi}$	0,726	0,743	1	1
	$\alpha_{-\rho}$	0,981	0,259	1	1
	$\alpha_{+\phi}$	0,981	0,307	1	0,975
	$\alpha_{+\rho}$	0,882	0,815	0,037	0,709
	β	0,0999	0,892	1	1
podwójnie miesza- ny z rozpadem Q	$\alpha_{-\phi}$	0,85	0,288	0,22	0,59
	$\alpha_{-\rho}$	0,654	0,000448	0,272	0,292
	$\alpha_{+\phi}$	0,278	0,696	1	1
	$\alpha_{+\rho}$	1	0,491	1	1
	β	0,0197	0,555	0,416	0,369
	S	0,468	0,815	0,566	0,566
podwójnie miesza- ny z rozpadem β	$\alpha_{-\phi}$	0,235	0,926	1	1
	$\alpha_{-\rho}$	0,612	0,264	0,321	1
	$\alpha_{+\phi}$	0,611	0,0509	0,666	0,588
	$\alpha_{+\rho}$	0,128	0,709	0,007	0,246
	β_0	0,189	0,928	1	1
	δ	0,944	0,867	0,054	0,356

*Liczebność grup to: mutant *Htr7*: 12, kontrola *Htr7*: 16, mutant NR1^{Tph2CreERT2}: 13, kontrola NR1^{Tph2CreERT2}: 15 i C57BL/6J: 14 osobników.

Podsumowując, zachowanie zwierząt można opisać w kategoriach uczenia ze wzmocnieniem z powodu wzmacniającego działania nagrody (Rycina 5-10 i Rycina 5-12) oraz częstości

działania na korzyść pewniejszej opcji powyżej granicy losowego wyboru (Rycina 5-11 i Rycina 5-14). W tym doświadczeniu test został uproszczony względem poprzedniego doświadczenia. Doprowadziło to do częstszego działania związanego z pewniejszym uzyskaniem nagrody (Rycina 5-11). Uproszczenie schematu nie wpłynęło na zależność od czasu podjęcia działania, która była zgodna z tendencją, opisaną w poprzednim doświadczeniu, stanowiącą o zwiększeniu prawdopodobieństwa powtórzenia działania wraz z wydłużaniem interwału (Rycina 5-16). Nie wskazano jednoznacznie jednego modelu najlepiej naśladującego zachowanie zwierząt w doświadczeniu, natomiast widoczne było lepsze dopasowanie modeli mieszanych, zakładających zdolność do określenia wartości niepodjętego, hipotetycznego działania (Rycina 5-19).

W bieżącym doświadczeniu analizowano rolę receptorów 5-HT₇ oraz NMDAR-zależnej transmisji serotoniny przy użyciu myszy transgenicznych. Myszy pozbawione receptorów 5-HT₇ nie wykazały zaburzenia procesu uczenia się oraz podejmowania decyzji. Dla grupy eksperymentalnej zanotowano więcej prób uzyskania nagrody w porównaniu do grupy kontrolnej. Indukowany w dorosłości zanik receptorów NMDA wpłynął na wielkość efektu czasu na decyzję, wskazując mniej nasiloną skłonność grupy eksperymentalnej do powtórzenia wyboru wraz z upływającym czasem w porównaniu do grupy kontrolnej.

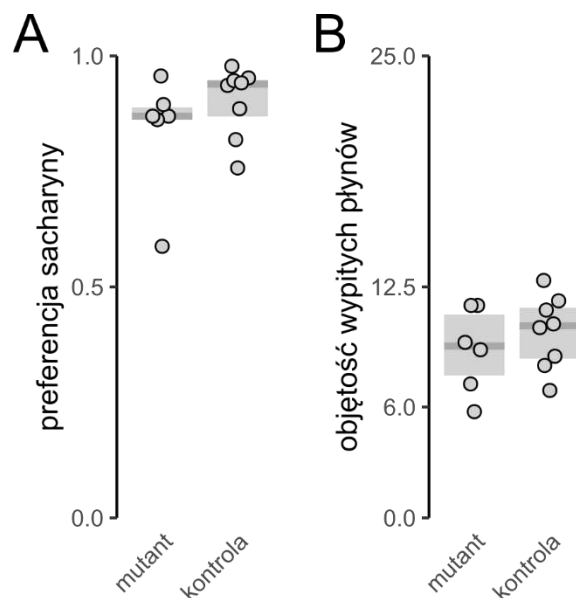
Testowane grupy charakteryzowały się zróżnicowanym składem osobników, ponieważ w jednej kohorcie testowano zarówno osobniki z mutacją jak i kontrolne. W porównaniu do wskazanych grup jednolita kohorta myszy C57BL/6J charakteryzowała się lepszą zdolnością do rozpoznania zróżnicowanego prawdopodobieństwa uzyskania nagrody w obu narożnikach (Rycina 5-11 i Rycina 5-14), większym efektem nagrody (Rycina 5-16) oraz silniejszą tendencją do powtórzenia działania zależnie od czasu mijającego od poprzedniej próby (Rycina 5-16).

5.3 Analiza uczenia ze wzmocnieniem u zwierząt szczepu NR1^{Tph2CreERT2} w teście probabilistycznego odwrócenia w klatkach do warunkowania instrumentalnego

Poprzednie doświadczenia dostarczyły opisu zachowania samic myszy w nowym teście uczenia ze wzmocnieniem. Aby określić różnice w zachowaniu w porównaniu do tradycyjnego sposobu testowania test probabilistycznego odwrócenia został przeprowadzony w klatkach do warunkowania instrumentalnego. Doświadczenie zostało wykonane wraz z dr Zofią Hardą (Zakład Neurofarmakologii Molekularnej, Instytut Farmakologii PAN). Uczenie ze wzmocnieniem opisano u myszy linii NR1^{Th2CreERT2}, analogicznie do poprzedniego doświadczenia, jednak w bieżącym doświadczeniu, ze względu na dostępność, u samców.

W celu określenia potencjalnego wpływu mutacji na wrażliwość na nagrodę u samców, przed rozpoczęciem doświadczenia, w osobnej grupie zwierząt, przeprowadzono test preferencji sacharyny. W tym teście myszy przebywające pojedynczo w klatce mogły, przez 24 godziny, bez ograniczeń, pić z butelek z 0,1% (w./obj.) z roztworem sacharyny oraz wodą. Preferencja

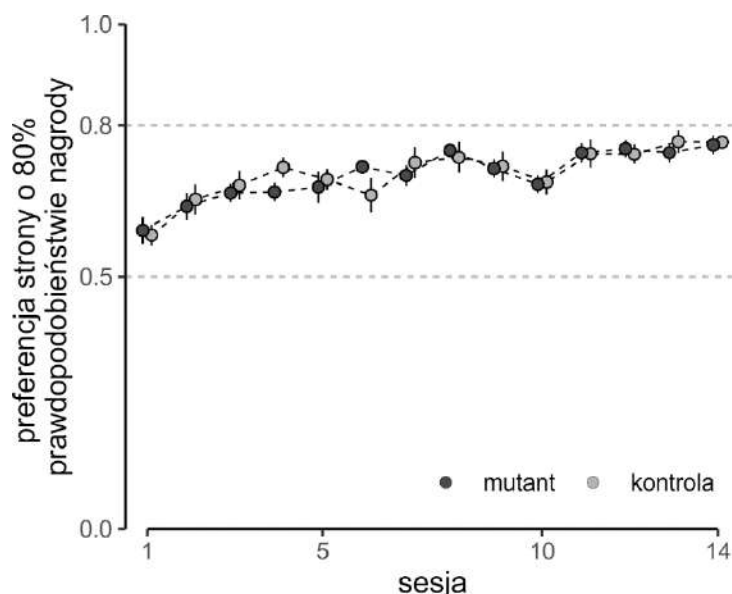
sacharyny, określona na podstawie spożytej objętości względem całości wypitych płynów, była wyższa od 0,5 w grupie myszy z mutacją i kontrolnych (test Wilcoxona z wartością 0,5: $p = 0,0178$ i $p = 0,00391$, odpowiednio) (Rycina 5-20). Nie wykazano różnicy zależnej od genotypu w objętości wypitych płynów (test Wilcoxona $p = 0,561$), ani preferencji sacharyny ($p = 0,332$).



Rycina 5-20 Preferencja sacharyny. (A) Objętość wypitego roztworu sacharyny względem całkowitej objętości wypitych płynów (roztworu sacharyny i wody). (B) Objętość wypitych płynów w ml. Określono z dokładnością do 0,5ml. Na wykresach długość prostokąta reprezentuje 50% środkowych obserwacji, pionowa linia medianę, punkty wartości dla pojedynczych osobników, odcień punktu oznaczenie kohorty. Liczebność: 6 mutantów i 8 osobników kontrolnych.

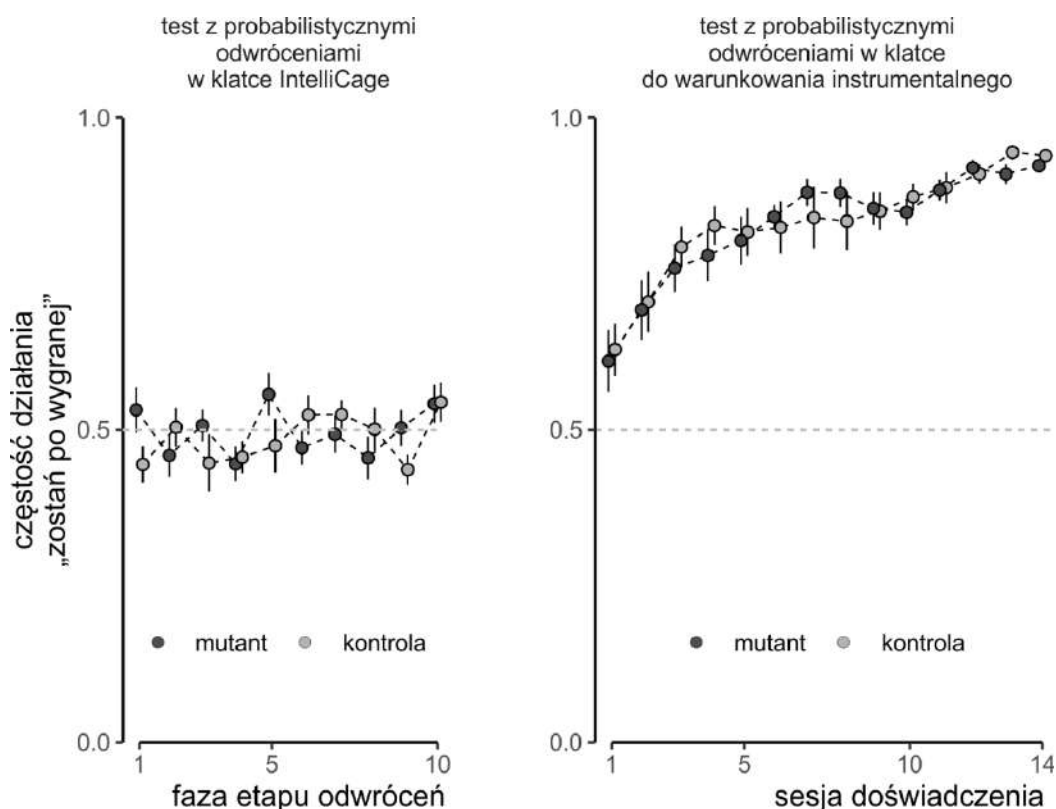
W teście z probabilistycznym odwróceniem przybliżoną efektywność wykonania testu określono za pomocą częstości wyboru opcji związanej z pewniejszym dostępem do nagrody. W tym teście, aby uzyskać nagrodę w postaci chrupki, mysz mogła wybrać jeden z dwóch dostępnych portów aparatu, którym przypisano prawdopodobieństwo uzyskania nagrody na poziomie 80% i 20%. Przypisanie prawdopodobieństwa do portów klatki było odwracane w środku sesji o stałej liczbie 120 możliwych wyborów. Każda testowana mysz wykonała 14 sesji, co oznacza maksymalną możliwą w doświadczeniu liczbę 1680 prób. Do analiz nie włączono 4 osobników (1 mutant i 3 kontrolne), dla których całkowita liczba prób nie przekroczyła 99% maksymalnej liczby wyborów (1663). Sprawdzone czy wartość preferencji strony o wyższym prawdopodobieństwie nagrody, określona jako ułamek wyborów portu o wyższej szansie nagrody względem obu portów, różniła się zależnie od genotypu i numeru sesji. ANOVA z powtórzonymi wykazała zmianę preferencji lepszej strony w zależności od sesji ale nie od genotypu (ANOVA efekt sesji $F(13/208) = 9,451$ i $p < 0,001$; efekt genotypu $F(1/16) = 0,035$ i $p = 0,854$, interakcja sesji i genotypu $F(13/208) = 0,723$ i $p = 0,74$). Wartość preferencji zwiększyła się z $58,8 \pm 1,82\%$ (średnia $\pm$ błąd standardowy średniej) w pierwszej sesji do $76,3 \pm 1,26\%$ w ostatniej sesji (Rycina 5-21, Rycina 5-22). Takiej tendencji nie zanotowano dla preferencji opcji o wyższym

prawdopodobieństwie uzyskania nagrody dla kolejnych faz etapu odwróceń w teście dedykowanym do aparatu IntelliCage (Rycina 5-14).



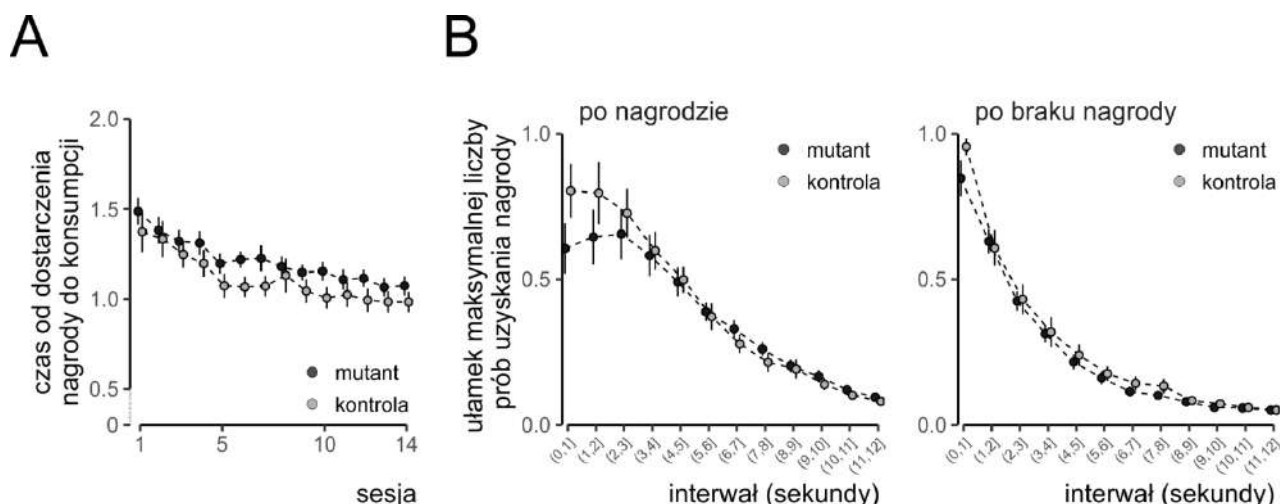
Rycina 5-21 Preferencja opcji o wyższym prawdopodobieństwie uzyskania nagrody. Ułamek wyborów opcji o wyższym prawdopodobieństwie nagrody dla następujących po sobie sesji doświadczenia. Punkt oznacza wartość średnią, zakresy błąd standardowy średniej, odcień genotyp. Linie przerywane stanowią odniesienie do wartości 0,5 i 0,8 i służą uwidocznieniu wskazanej zależności preferencji od sesji doświadczenia. Liczebność to 14 mutantów i 8 zwierząt kontrolnych.

W celu weryfikacji czy strategia podejmowania decyzji zwierząt była oparta o efekt wzmocnienia określono częstość działania powtórzenia tego samego wyboru po uzyskaniu nagrody („zostań po wygranej”) dla kolejnych sesji testu. Zmiana ułamka wyborów „zostań po wygranej” w kolejnych sesjach była znamienna i niezależna od genotypu (ANOVA efekt sesji $F(13/260) = 19,683$ i $p < 0,001$, efekt genotypu $F(1/20) = 0,058$, $p = 0,812$, interakcja sesji i genotypu $F(13/260) = 0,442$ i $p = 0,953$). Częstość tego działania wzrastała dla kolejnych sesji doświadczenia, od $0,617 \pm 0,0343$ w pierwszej do $0,929 \pm 0,0076$ w ostatniej sesji (Rycina 5-22). Taka tendencja nie była widoczna dla testu z wykorzystaniem aparatu IntelliCage, dla kolejnych odwróceń, czyli faz doświadczenia (ANOVA efekt fazy $F(9/234) = 1,727$ i $p = 0,084$; efekt genotypu $F(1/26) = 0,185$; $p = 0,671$, interakcja sesji i genotypu $F(9/234) = 2,019$ i $p = 0,038$) (Rycina 5-22). Sugeruje to szybsze, w domenie czasu, zwiększenie częstości wyborów strony o większym prawdopodobieństwie uzyskania nagrody u testowanych w klatce do warunkowania instrumentalnego.



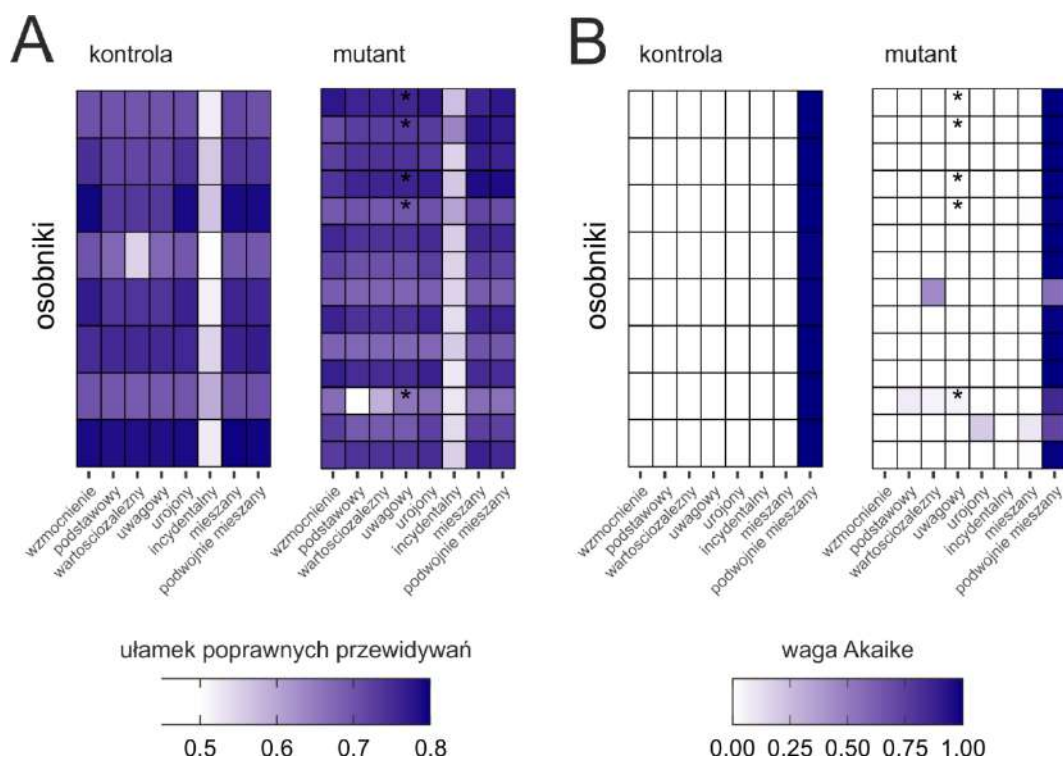
Rycina 5-22 Działanie w sytuacji zróżnicowanego prawdopodobieństwa nagrody. Na wykresie punkt oznacza wartość średnią, zakresy błęd standardowy średniej. (A) Częstość działania „zostań po wygranej” dla kolejnych faz etapu odwróceń w doświadczeniu z wykorzystaniem klatki IntelliCage. Liczebność to 13 mutantów i 15 zwierząt kontrolnych. (B) Częstość działania „zostań po wygranej” dla kolejnych sesji doświadczenia w doświadczeniu z wykorzystaniem klatki do warunkowania instrumentalnego. Liczebność to 14 mutantów i 8 zwierząt kontrolnych.

W kolejnym kroku określono czy genotyp wpłynął na czas reakcji. Czas od reakcji podajnika do wetknięcia pyszczka do otworu wyniósł nie więcej niż 4,98 s dla każdego zwierzęcia i jego wartość, uśredniona dla osobnika i sesji, malała wraz z kolejnymi sesjami doświadczenia (Rycina 5-23A). Analiza wariancji dla powtórzonych pomiarów nie wykazała różnicy zależnej od genotypu oraz sesji w czasie sięgnięcia po nagrodę (ANOVA efekt sesji $F(13/260) = 27,821$ i $p < 0,001$, efekt genotypu $F(1/20) = 1,51$ i $p = 0,233$, interakcja sesji i genotypu $F(13/260) = 0,516$ i $p = 0,914$). Z kolei czas podjęcia decyzji, określony jako czas mijający od rozpoczęcia próby do wetknięcia pyszczka do portu, zależał od rezultatu poprzedniego działania. Podjęcie decyzji po uzyskaniu nagrody, szacowane na podstawie rzeczywistej częstości występowania interwałów decyzji było wysoce prawdopodobne w zakresie do 5 sekund od rozpoczęcia próby (Rycina 5-23B). Natomiast podjęcie decyzji po braku nagrody było wysoce prawdopodobne dla interwału krótszego od 3 sekund (Rycina 5-23B). Genotyp nie wpłynął na częstość prób dla żadnego zakresu interwału w żadnym z rozważanych przypadków (ANOVA z powtórzonymi pomiarami, efekt genotypu $F(1/20) = 0,446$ i $p = 0,512$, efekt zakresu $F(11/231) = 106,084$ i $p < 0,001$, efekt nagrody $F(1/263) = 87,791$ i $p < 0,001$).



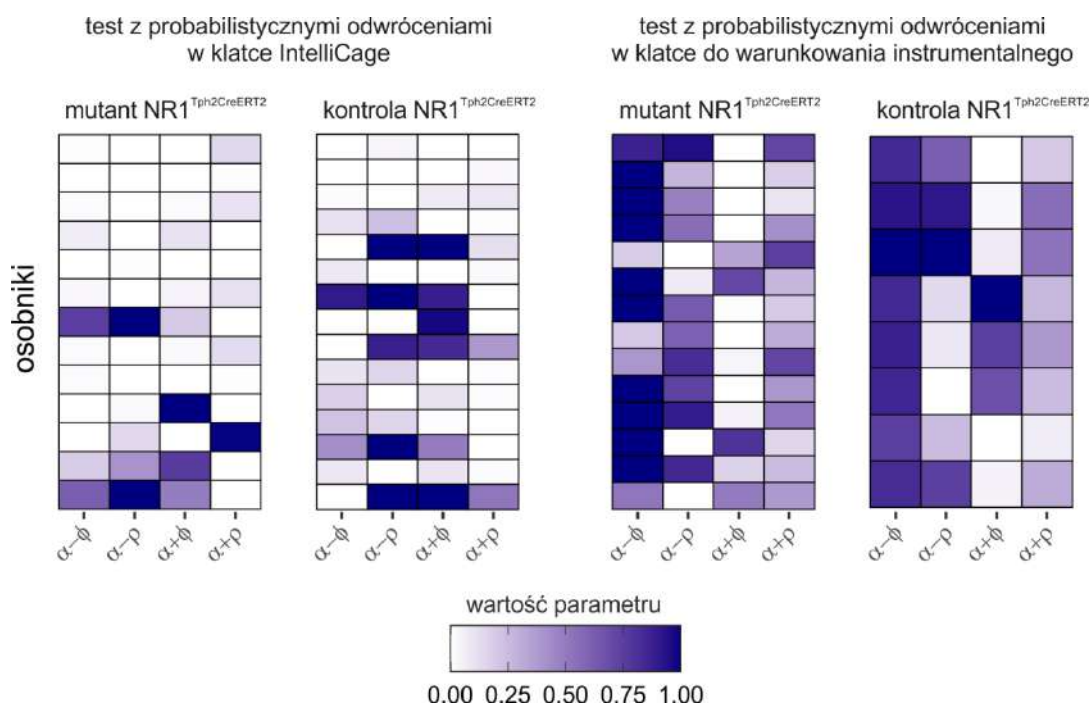
Rycina 5-23 Opóźnienie między odpowiedziami instrumentalnymi. (A) Średni czas od dostarczenia nagrody do rozpoczęcia spożycia dla kolejnych sesji doświadczenia, w sekundach. Liczebność to 14 mutantów i 8 zwierząt kontrolnych. (B) Znormalizowana częstość wyborów osobnika o danym zakresie interwału decyzji (od początku próby do wetknięcia pyszczka do portu), w zależności od nagrody w poprzedniej próbie, w sekundach. Znormalizowano względem referencyjnej wartości maksymalnej liczby prób uzyskania nagrody. Przedział z największą liczbą prób uzyskania nagrody wyznaczono indywidualnie dla każdego osobnika. Liczba prób uzyskania nagrody dla tego zakresu służyła jako wartość referencyjna dla pozostałych zakresów. Punkt oznacza wartość średnią, zakresy błąd standardowy średniej. Liczebność to 14 mutantów i 8 zwierząt kontrolnych.

W celu określenia najbardziej prawdopodobnej strategii decyzji porównano dopasowania modeli obliczeniowych do danych. Na podstawie odsetka poprawnych przewidywań, modelami o największej liczbie osobników z najlepszym dopasowaniem, zarówno dla grupy mutantów i kontrolnej, były: „wzmocnienie” (w 40,6% i 33,3% przypadków, odpowiednio), „mieszany” (37,5% i 16,7%) oraz „podwójnie mieszany” (9,38% i 16,7%) (Rycina 5-24A) z zastrzeżeniem, że dla jednego osobnika dwa modele dostarczyły równej wartości ułamka poprawnych przewidywań, więc każdemu z nich przypisano wagę $\frac{1}{2}$, jak opisano dla wcześniejszych doświadczeń. Kontrolny model „incydentalny” charakteryzował się mniejszym ułamkiem poprawnych przewidywań zarówno dla grupy mutantów ($0,558 \pm 0,0082$, średnia $\pm$ błąd standardowy średniej) jak i kontrolnej ($0,541 \pm 0,0108$) (Rycina 5-24A). Na podstawie wagi Akaike najlepiej dopasowanym modelem spośród zaproponowanych rozwiązań, dla większości grupy, zarówno mutantów i kontrolnej był model „podwójnie mieszany” (Rycina 5-24B).



Rycina 5-24 Dopasowanie modelu do danych. (A) Ułamek poprawnych przewidywań modeli zastosowanych w badaniach. Odcień pola odpowiada wartości ułamka poprawnych przewidywań. Dla przejrzystości wartości niższe i równe 0,5 oznaczono jednakowo. Rząd określa wartości pojedynczego osobnika, kolumna wartości danego modelu. Znak „*” określa przypadki dla których nie uzyskano dodatniej wartości optymalnych parametrów związanych z rzeczywistym wzmocnieniem. (B) Prawdopodobieństwo warunkowe modelu na podstawie wagi Akaike (w). Odcień pola odpowiada wartości w uzyskanych dla danego modelu. Rząd określa wartości pojedynczego osobnika, kolumna wartości danego modelu. Znak „*” określa przypadki dla których nie uzyskano dodatniej wartości optymalnych parametrów związanych z rzeczywistym wzmocnieniem.

W modelu „podwójnie mieszanym” szacowanie oczekiwanej nagrody odbywa się w sytuacji uzyskania nagrody w wybranym narożniku (α_{+p}) i alternatywnym do wybranego ($\alpha_{-\phi}$) oraz po braku nagrody w narożniku wybranym (α_{-p}) i alternatywnym ($\alpha_{+\phi}$) (Rycina 5-25). Efekt rzeczywistej nagrody określony różnicą między α_{+p} i α_{-p} nie był znamienny w grupie kontrolnej ani mutantów (sparowany test Wilcozona, $p = 0,195$ i $p = 0,217$, odpowiednio), co wskazało na jednakowe znaczenie rzeczywistego uzyskania i braku nagrody. Podobnie, efekt przewidywanego działania z pozytywnym skutkiem względem rzeczywistego przyznania nagrody określony jako różnica między α_{+p} i $\alpha_{+\phi}$ nie był znamienny dla grupy kontrolnej ani mutantów ($p = 0,945$ i $p = 0,0906$), co oznaczało, że waga rzeczywistego uzyskania nagrody była równoważna z przewidywanym uzyskaniem nagrody. Efekt skutku przewidywanego działania z brakiem nagrody ($\alpha_{-\phi}$) względem rzeczywistego braku nagrody (α_{-p}) był znamienny zarówno dla grupy kontrolnej ($p = 0,0225$) jak i mutantów ($p = 0,0203$). Wartość α_{-p} była mniejsza od $\alpha_{-\phi}$ dla większości osobników grupy kontrolnej (7/8) i z mutacją (11/14). Wielkość kroku dla opcji z przewidywanym brakiem nagrody była większa od rzeczywistego braku nagrody. Widoczne było także zróżnicowanie wartości α wewnątrz grupy oraz wielokrotnie większe wartości parametrów α , dla tego modelu, w bieżącym doświadczeniu, w porównaniu do doświadczenia z wykorzystaniem systemu IntelliCage (Rycina 5-25).



Rycina 5-25 Optymalne wartości parametrów α modelu „podwójnie mieszanego”. Jasność pola odpowiada wartości α . Rząd określa wartości pojedynczego osobnika, kolumna wartości danego parametru α . Wykresy prezentują wartości w zależności od doświadczenia i genotypu zwierząt.

Warto zestawzić uzyskany wynik z uzyskanym dla analogicznych, ale prostszych, rozwiązań: modelu „wartościowego” i „mieszanego” (Tabela 5–7), w których wielkość kroku w sytuacji braku nagrody i przy uzyskaniu nagrody była zróżnicowana. W modelu „wartościowym” wielkość kroku w sytuacji braku nagrody (α_-) i przy uzyskaniu nagrody (α_+) była zróżnicowana w grupie kontrolnej (sparowany test Wilcoxona $p = 0,00781$) ale nie w grupie mutantów ($p = 0,153$). Wartość α_- była większa od α_+ dla wszystkich osobników grupy kontrolnej wskazując większe znaczenie braku nagrody w porównaniu do jej uzyskania. Także w modelu „mieszanym” wielkości α_+ i α_- różniły się istotnie w grupie mutantów ($p < 0,001$) i kontrolnej ($p = 0,00781$). Wartość α_+ była większa od α_- dla grupy kontrolnej (dla 8/8 osobników) oraz grupy mutantów (13/14). Interpretacja zachowania na podstawie modelu „wartościowego” wskazuje, że czułość na negatywną informację zwrotną była większa od czułości na pozytywną informację zwrotną, w grupie kontrolnej. Natomiast, w modelu „mieszanym”, przeciwnie, to wartość α próby zakończonej nagrodą byłaby większa w porównaniu do sytuacji braku nagrody, zarówno w grupie mutantów i kontrolnej.

Tabela 5–7 Wartości optymalne parametrów modeli

model	parametr	mutant NR1 ^{Tph2CreERT2}	kontrola NR1 ^{Tph2CreERT2}	test Wilcoxona p (mutant x kontrola)
wzmocnienie	ε	0,548 (0,493–0,6)*	0,508 (0,413–0,594)	0,267
podstawowy	α	0,636 (0,466–0,726)	0,74 (0,61–0,981)	0,441
	β	1,84 (1,69–2,18)	1,79 (1,52–1,94)	0,815
uwagowy	α_0	0,555 (0,419–1)	0,733 (0,595–0,966)	0,0689
	β	1,99 (1,19–2,41)	1,95 (1,37–2,26)	0,714

model	parametr	mutant NR1 ^{Tph2CreERT2}	kontrola NR1 ^{Tph2CreERT2}	test Wilcoxona p (mutant x kontrola)
	μ	0,168 (0–0,337)	0,0413 (0–0,0826)	0,129
wartościowo zależny	α_{-}	0,662 (0,485–0,742)	0,802 (0,644–0,959)	0,33
	α_{+}	0,56 (0,419–0,633)	0,641 (0,585–1,03)	0,815
	β	1,96 (1,83–2,32)	1,99 (1,73–2,25)	0,570
urojony	α	0,399 (0,325–0,504)	0,432 (0,289–0,523)	0,365
	β	1,88 (1,78–2,3)	1,8 (1,54–2,09)	0,714
mieszany	α_{-}	0,275 (0,154–0,366)	0,384 (0,247–0,471)	0,145
	α_{+}	0,598 (0,483–0,69)	0,637 (0,483–0,754)	0,525
	β	1,77 (1,69–2,11)	1,71 (1,45–1,94)	0,714
podwójnie mieszany	$\alpha_{-\phi}$	1 (1–1)	0,851 (0,774–0,872)	0,251
	$\alpha_{-\rho}$	0,579 (0,377–1)	0,434 (0–0,776)	0,891
	$\alpha_{+\phi}$	0,0228 (0–0,0455)	0,0615 (0–0,123)	0,382
	$\alpha_{+\rho}$	0,344 (0,168–0,476)	0,285 (0,0402–0,371)	0,714
	β	2,05 (1,7–2,26)	2,27 (1,17–2,65)	0,664
incydentalny	α_{ϕ}	0,0326 (0,0281–0,0485)	0,027 (0,0186–0,0334)	0,664
	α_{ρ}	0,0701 (0,0476–0,0828)	0,0863 (0,0782–0,108)	0,57
	β	10,6 (0–15,8)	6,74 (2,23–8,27)	0,482

*Wartości przedstawiono jako mediana i 95% przedział ufności. Liczebność to 14 mutantów i 8 myszy kontrolnych.

Optymalne wartości parametrów nie różnicowały genotypu zwierząt w żadnym z zastosowanych modeli (Tabela 5–7). W modelu „wzmocnienie”, dla którego wskazano lepsze dopasowanie do danych na podstawie ułamka poprawnych przewidywań, uzyskana wartość mediany parametru poniżej 1 wskazała większe prawdopodobieństwo działania optymalnego „zostań po wygranej” i „zmień po przegranej” zgodnie z wcześniej zaprezentowanymi wynikami (Rycina 5-22).

Podsumowując, w teście z probabilistycznym odwróceniem wykonanym w klatkach do warunkowania instrumentalnego widoczna była adaptacja zwierząt do wykonania zadania (Rycina 5-22) oraz szybsze zwiększanie preferencji opcji o pewniejszym uzyskaniu nagrody (Rycina 5-21) w porównaniu do doświadczenia z wykorzystaniem klatki IntelliCage (Rycina 5-14). Zwierzęta wykonywały próby uzyskania nagrody w bieżącym doświadczeniu relatywnie wielokrotnie częściej (Rycina 5-23) niż w doświadczeniu z zastosowaniem klatki IntelliCage (Rycina 5-15) (liczba prób/jednostka czasu). Natomiast modelowanie obliczeniowe zachowania, obserwowanego w tym doświadczeniu, charakteryzowało się dopasowaniem modeli do danych przekraczającym próg losowego wyboru dla wszystkich modeli, poza kontrolnym modelem „incydentalnym” (Rycina 5-24A), upewniając o zachowaniu zwierząt zgodnym z założeniami uczenia ze wzmocnieniem oraz uczeniu się w tym teście (por. Rycina 5-21 i Rycina 5-22). W bieżącym doświadczeniu modelem o najlepszym dopasowaniu do danych było rozwiązanie „podwójnie mieszane” (Rycina 5-24B). Wartości optymalnych parametrów tego modelu wskazywały na szczególne znaczenie dewaluacji wartości działania alternatywnego w sytuacji uzyskania dostępu do nagrody (Rycina 5-25B). W żadnej z miar opisujących zachowanie zwierząt oraz reprezentujących wyniki modelowania nie wskazano różnicy zależnej od genotypu.

6. Dyskusja

W badaniach zaprezentowanych w niniejszej rozprawie opracowano nowy test behawioralny uczenia się ze wzmocnieniem, w którym decyzje podejmowane są w warunkach zbliżonych do naturalnych. Celem było także opisanie różnic w efekcie działania nagrody w postaci etanolu i sacharyny oraz wskazanie roli wybranych składowych związanych z transmisją serotonergiczną w uczeniu ze wzmocnieniem. W celu opisu strategii podejmowania decyzji zastosowano modele obliczeniowe procesu podejmowania decyzji, dla których określono zdolność do przewidywania zachowania zwierząt.

Ważna obserwacja z niniejszych badań stanowi, że zachowanie zwierząt podlega założeniom modeli uczenia ze wzmocnieniem, co było widoczne m.in. w lepszym dopasowaniu do danych modeli opartych o algorytm *Q-learning*, w porównaniu do rozwiązań kontrolnych. Dlatego modele obliczeniowe uczenia ze wzmocnieniem mogą być stosowane do tłumaczenia obserwowanego zjawiska. Porównanie modeli obliczeniowych o zróżnicowanych założeniach względem sposobu aktualizacji wartości oczekiwanej nagrody, służyło określeniu, który z proponowanych modeli stanowi najlepsze wyjaśnienie obserwowanego zjawiska. W doświadczeniach z zastosowaniem klatki IntelliCage modele oparte o algorytm *Q-learning* uzyskały porównywalne dopasowanie, natomiast w doświadczeniu z zastosowaniem klatki do warunkowania instrumentalnego zdecydowanie najdokładniejszego wyniku dostarczyło rozwiązanie „podwójnie mieszane”, przygotowane na potrzeby tych badań. Zaproponowane na potrzeby badań z zastosowaniem klatki IntelliCage modyfikacje tego modelu, odwołujące się do efektu zapomnienia i wzrostu niepewności z czasem nie dostarczyły oczekiwanego polepszenia dopasowania modelu.

Zarówno nagroda w postaci sacharyny, etanolu jak i mieszaniny obu działała wzmacniająco na zachowanie myszy, chociaż nagroda ze słodzikiem charakteryzowała się większą preferencją u zwierząt. W odróżnieniu zarówno do grupy sacharynowej jak i grupy z mieszaniną alkoholu i sacharyny, w grupie alkoholowej zanotowano mniejszą preferencję opcji o wyższym prawdopodobieństwie uzyskania nagrody oraz analogicznie, mniejszy efekt nagrody w przewidywaniu prawdopodobieństwa powtórzenia decyzji. Niższe wartości wskazanych współczynników były prawdopodobnie spowodowane awersyjnym smakiem alkoholu, ponieważ analogicznych różnic nie zanotowano między grupą sacharynową i z mieszaniną substancji. Zbliżone wyniki uzyskane dla grupy sacharynowej i z mieszaniną substancji oznaczają, że hipoteza o zaburzeniu procesu podejmowania decyzji pod wpływem alkoholu spożywanym o dowolnym czasie, chronicznie i w niskim stężeniu nie została potwierdzona. Wyniki modelowania nie dostarczyły jednoznacznej koncepcji wpływu rodzaju nagrody na strategię podejmowania decyzji.

Myszy pozbawione receptorów 5-HT₇ nie wykazały zaburzenia procesu podejmowania decyzji. Dla grupy eksperymentalnej zanotowano więcej prób uzyskania nagrody w porównaniu

do grupy kontrolnej, jednak nie przełożyło się to na inne miary behawioralne. Podobnie, indukowany w dorosłości zanik receptorów NMDA na neuronach serotoninowych nie zaburzył zachowania testowanego zarówno z wykorzystaniem klatki IntelliCage jak i klatki do warunkowania instrumentalnego chociaż wpłynął na wielkość efektu czasu na decyzję opisany w teście z wykorzystaniem klatki IntelliCage.

Co ciekawe, zaobserwowano także zależność czasu działania i strategii podejmowania decyzji stanowiącą, że im więcej czasu minie od poprzedniej decyzji tym większe będzie prawdopodobieństwo powtórzenia poprzedniego wyboru (Jabłońska i in., 2021).

6.1 Uczenie z odwróceniem testowane behawioralnie

6.1.1 Potwierdzenie wystąpienia uczenia ze wzmocnieniem

W pierwszym kroku konieczne było potwierdzenie wystąpienia uczenia ze wzmocnieniem. Stwierdzono to na podstawie właściwości nagrody do silnego wzmocniania zachowania oraz zaangażowania zwierząt w zadanie podczas całego doświadczenia. W każdym z zaprezentowanych doświadczeń preferencja substancji przekroczyła 50% w etapie adaptacji (metodyka z wykorzystaniem klatki IntelliCage) lub w badaniach wstępnych (metodyka z wykorzystaniem klatki do warunkowania instrumentalnego). Zwierzęta były zaangażowane w wykonanie zadania wykonując próby uzyskania nagrody każdej doby czy sesji. W przypadku braku wzmocniającego efektu substancji liczba prób uzyskania nagrody prawdopodobnie malałaby w trakcie trwania doświadczenia. Tymczasem ich liczba utrzymywała się na stałym poziomie.

Ze względu na charakterystykę testu z probabilistycznym odwróceniem, określenie bezwzględnej miary efektywności wykonania zadania czy uczenia się nie było możliwe w tych badaniach. Charakterystyczną cechą testu z odwróceniem jest następujący po odwróceniu czas potrzebny na przeuczenie, w którym można oczekiwać wyborów na korzyść gorszej alternatywy oraz ponownej eksploracji środowiska. W tym teście, w zachowaniu notuje się trudną do ilościowego opisu rozbieżność między jednoczesnym wykorzystaniem dostępnej wiedzy a eksploracją otoczenia. Dodatkowo, w probabilistycznej wersji testu z odwróceniem podczas wyboru strony o wyższym prawdopodobieństwie nagrody (0,9) istnieje niższe co do wartości (1-0,9) prawdopodobieństwo braku nagrody czyli rezultatu wprowadzającego w błąd (ang. *misleading output*). Podczas testu mysz wykona także wybory eksploracyjne oraz związane z błędem wynikającym z konieczności przeuczenia się. Dlatego wykorzystana w niniejszych badaniach miara preferencji strony o wyższym prawdopodobieństwie nagrody nie odpowiada bezpośrednio efektywności uczenia się i stanowi wyłącznie przybliżoną miarę efektywności uczenia się. Natomiast można określić, że jeśli wskazana preferencja przekracza 50%, czyli granicę losowości, oznacza to zachowanie nakierowane na cel oraz jeśli wskazana wartość wzrasta wraz z czasem trwania doświadczenia to następuje adaptacja do testu i uczenie się. Uzyskanie miary efektywności byłoby możliwe przy zmianie schematu doświadczenia. Występowanie odwrócenia przy-

pisania szansy uzyskania nagrody mogłoby następować zgodnie z zadaniem kryterium, np. po 8 wyborach strony o pewniejszej nagrodzie, co pozwoliłoby na określenie liczby wskazanych sekwencji powtórzeń w sekwencji wszystkich wyborów, i tym samym efektywności wykonania zadania (np. (Ineichen i in., 2012; Bryce i Floresco, 2021)). Natomiast mogłoby to ograniczyć możliwości wnioskowania o procesie podejmowania decyzji na podstawie wyników modelowania obliczeniowego z wykorzystaniem rozwiązań uczenia ze wzmocnieniem.

Wykorzystanie testu pozwalającego na obserwację procesu podejmowania decyzji w warunkach zbliżonych do normalnych w klatce IntelliCage związane jest z szeregiem okoliczności, które należy wziąć pod uwagę podczas analizy zachowania. Wykorzystanie systemu IntelliCage pozwala na testowanie zachowania grupy myszy składającej się z od kilku do kilkunastu osobników w sposób ciągły i automatyczny przez kilka tygodni. Takie cechy stoją w kontraście do standardowych testów, w których zachowanie osobnika obserwowane jest indywidualnie w krótkim czasie i w warunkach innych niż klatka domowa (Kiryk i in., 2020), np. w testach z wykorzystaniem klatki do warunkowania instrumentalnego (Cieślak i in., 2018; Noworyta-Sokolowska i in., 2019). Badania z wykorzystaniem systemu IntelliCage prowadzą do opisu wpływu różnych czynników na wybrane aspekty zdolności poznawczych i wzorców aktywności zwierząt (Endo i in., 2011; Kiryk i in., 2020). Opracowano wiele różnych schematów doświadczalnych, w których notowane zachowanie stanowi niewymuszoną aktywność prowadzącą do osiągnięcia celu, a dokładniej dostępu do substancji znajdującej się w narożnikach klatki (np. (Radwanska i Kaczmarek, 2012; Smutek i in., 2014; Beroun i in., 2018; Koskela i in., 2018; Kiryk i in., 2020)). W badaniach opisanych w niniejszej rozprawie, w celu uzyskania napoju o właściwościach nagradzających, zwierzęta mogły przez miesiąc swobodnie wykonywać wizyty w narożnikach klatki IntelliCage. Zwierzęta nie były poddane ograniczeniu dostępu do pokarmu ani nie narzucono im ograniczeń czasowych w podejmowaniu decyzji. Dlatego obserwowane zachowanie zwierząt można określić jako spontaniczne i zbliżone do warunków naturalnych. Jako próbę określenia czynników wpływających na zachowanie, a dokładniej na proces podejmowania decyzji, zastosowano regresję logistyczną, przewidującą prawdopodobieństwo działania powtórzenia wyboru zależnie od uzyskania nagrody, czasu mijającego od poprzedniej próby oraz indywidualnej preferencji miejsca. Ten model wskazał na zgodny z oczekiwaniami efekt uzyskania dostępu do nagrody w zwiększeniu szansy powtórzenia wyboru przez mysz.

Innymi czynnikami wpływającymi na obserwowane zachowanie mogły być cykl okołodobowy aktywności, sen i interakcje społeczne, występujące w czasie pomiędzy kolejnymi wizytami w narożnikach aparatu IntelliCage. Wydaje się, że z dostępnych aktualnie rozwiązań, do określenia wpływu wskazanych czynników możliwy do zastosowania byłby system rejestracji zachowania zwierząt za pomocą kamery wraz z rozwiązaniami uczenia maszynowego do automatycznego rozpoznawania zachowania zwierząt na podstawie pozycji i lokalizacji ciała np. (Kabra i in., 2013; Mathis i in., 2018), natomiast takie rozwiązania nie osiągają jeszcze dokładności pozwalającej na jednoczesne monitorowanie kilkunastu osobników przebywających blisko siebie. Biorąc pod uwagę, że w doświadczeniach z wykorzystaniem klatki IntelliCage kontakt z

eksperymentatorem był ograniczony, a zwierzęta znajdowały się w izolowanym i spokojnym pomieszczeniu, można założyć, że myszy utrzymywały swój naturalny cykl okołodobowy (Kiryk i in., 2020).

W badaniach z wykorzystaniem klatki IntelliCage możliwe jest wystąpienie uczenia grupowego. W badaniu Anny Kiryk i współpracowników wykazano zależność efektywności uczenia się od składu grupy, wskazując unormowanie wartości wskaźnika uczenia się, dla myszy modelowych choroby Alzheimera, testowanych w grupie mieszanej z myszami typu dzikiego, w klatce IntelliCage (Kiryk i in., 2011). Kłopotliwe pozostaje nadal określenie w jaki sposób takie uczenie mogłoby zachodzić. W innych badaniach wskazano, że w aktywności gryzoni następuje podążanie za sobą (Winiarski i in., 2021). Natomiast zarówno kryteria występowania podążania m.in. czas następujący po wizycie poprzedniego osobnika jak i wzorzec sugerowanego podążania nie zostały jednoznacznie określone. Stanowi to ciekawy obszar dalszych badań biorąc pod uwagę modele teoretyczne transmisji informacji w homo- czy heterogennych sieciach społecznych (np. (Karamched i in., 2020)). Należy tutaj jednak wziąć pod uwagę sposób funkcjonowania organizmu modelowego. U myszy transmisja informacji o pożywieniu odbywa się głównie za pomocą węchu (Gire i in., 2016; Loureiro i in., 2019). Ograniczenia w modelowaniu zjawiska uczenia się grupowego polegałyby więc na dynamicznym przemieszczaniu się zwierząt w grupie, nietrwałości zapachu oraz niejednakowym dostępie do informacji dla osobników zaangażowanych w danym momencie w inne czynności. Dlatego do zbadania tego zjawiska konieczne jest opracowanie innego sposobu testowania behawioralnego, w którym zredukowane zostałoby nakładanie się wskazanych czynników. W badaniach wstępnych dla niniejszej pracy nie określono wzorca podążania zwierząt za sobą. Natomiast biorąc pod uwagę możliwe wystąpienie takiego uczenia się, efekt heterogeniczności grupy zweryfikowano za pomocą porównania wyników grup kontrolnych linii *Htr7* oraz *Tph2*^{CreERT2} z wynikami grupy 14 samic myszy C57BL/6J, w teście probabilistycznego odwrócenia bez faz pośrednich. Zachowanie homogennej grupy myszy C57BL/6J sugerowało szybsze uczenie się w porównaniu do grup kontrolnych zwierząt, dlatego występowanie uczenia się grupowego w opisywanych badaniach jest możliwe.

Dodatkową kwestią dotyczącą testowania zwierząt w grupie jest dzielony dostęp do zasobów. Oznacza to potrzebę określenia czy wybór opcji zdeterminowany był wyłącznie procesem uczenia się a nie dostępnością narożników klatki w danym momencie. Biorąc pod uwagę to, że czas dostępu do butelki był ograniczony, myszy wykonywały krótkie wizyty, a całkowity czas zajęcia narożników w okresie doby nie przekraczał kilku godzin, odrzucono pomysł włączenia do niniejszych badań modelu symulującego wybór na podstawie dostępności narożników.

Zwierzęta testowane w doświadczeniu z wykorzystaniem klatki do warunkowania instrumentalnego wykazały adaptację do testu i efektywne uczenie się. Było to widoczne zarówno poprzez preferencję strony o wyższym prawdopodobieństwie nagrody jak i częstość powtórzenia wyboru po uzyskaniu nagrody, gdzie wartości obu miar zwiększały się dla kolejnych sesji doświadczenia. Dopasowanie modeli obliczeniowych do tych danych nie pozostawiało wątpliwości co do przebiegu procesu zgodnego z założeniami uczenia się ze wzmocnieniem a uzyska-

ne parametry modeli były wielokrotnie większe w porównaniu do doświadczeń z wykorzystaniem klatki IntelliCage (Rycina 5-25). Porównanie wyników zaprezentowanych do doświadczenia z wykorzystaniem klatki do warunkowania instrumentalnego i dla doświadczeń w IntelliCage, sugerowałyby mniej efektywne uczenie się myszy w doświadczeniach w IntelliCage (Rycina 5-22). Przekonującym wyjaśnieniem wydaje się zaangażowanie zwierząt w inne aktywności poza testem w klatce IntelliCage.

Przy porównaniu wyników uzyskanych z testowania z wykorzystaniem klatki IntelliCage oraz klatki do warunkowania instrumentalnego należy wziąć pod uwagę występowanie dodatkowych czynników różniących doświadczenia. Czynnikiem różniącymi z pewnością był stopień deprywacji pokarmowej, zwiększający motywację zwierząt do wykonania testu (Makowiecki i in., 2012). Ze względu na deprywację pokarmową, koszt utraconej szansy, czyli sytuacji braku nagrody w przypadku doświadczenia w klatkach IntelliCage, jest niższy niż w analogicznej sytuacji w doświadczeniu w klatce do warunkowania instrumentalnego oraz bardziej awersyjny ze względu na włączenie oświetlenia klatki. Innymi czynnikami były być stres związany z przenoszeniem zwierząt przez eksperymentatora (Gouveia i Hurst, 2019) i izolacja społeczna. Izolacja dotyczyła testu trwającego do godziny dziennie, a więc względnie krótkiego czasu całej doby. Natomiast w celu obniżenia stresu kontaktu z eksperymentatorem, myszy były przygotowywane poprzez poprzedzającą adaptację procedurę przyzwyczajania do eksperymentatora (ang. *handling*). Czynnikiem utrudniającym bezpośrednie porównanie między sposobami testowania jest potencjalny efekt płci, ponieważ do doświadczeń z wykorzystaniem systemu IntelliCage włączono samice, natomiast w badaniach z zastosowaniem klatek do warunkowania instrumentalnego, z powodu dostępności zwierząt, samce. Różnice płciowe notuje się m.in. zależnie od poziomu hormonów płciowych, jako nasilone uczenie społeczne u samic myszy, wykazywany typ agresji, czy zróżnicowane wyniki w testach zdolności poznawczych (Choleris i in., 2018), które jednak można częściowo wytłumaczyć zróżnicowaną wrażliwością na stres, co opisano w zadaniu z probabilistycznym odwróceniem u stresowanych samic i samców szczurów (Bryce i Floresco, 2021). Ogólnie, mężczyźni są bardziej skłonni do podejmowania ryzyka (Byrnes i in., 1999). Natomiast pod względem zdolności poznawczych płeć nie stanowi czynnika różnicującego, poza umiarkowanie większymi zdolnościami przestrzennymi u mężczyzn m.in. do rotacji przestrzennej obiektów trójwymiarowych (Hyde, 2016). Mimo to, dokładne porównanie efektu sposobu testowania w niniejszych badaniach wymagałoby dalszych badań.

6.1.2 Porównanie działania etanolu i sacharyny

Na podstawie przedstawionych analiz nie określono zaburzenia w procesie podejmowania decyzji czy zdolności do uczenia się warunkowanego dostępem do roztworu etanolu w niskim stężeniu. Myszy preferują roztwór alkoholu zdecydowanie mniej niż roztwór sacharyny (Yoneyama i in., 2008). Dlatego oczekiwanym wynikiem było to, że zwierzęta będą sięgać po alkohol z mniejszą częstotliwością po sacharynę. Istotnie, w badaniach potwierdzono oczekiwa-

nią niższą preferencję samego alkoholu w porównaniu do słodkich substancji (Rycina 5-1) i co z tym związane, także mniejszy efekt nagrody w procesie podejmowania decyzji (Rycina 5-6). Zwierzęta pod wpływem alkoholu rzadziej korzystały z opcji o pewniejszej nagrodzie w porównaniu do słodkich substancji (Rycina 5-3), co mogło być związane zarówno z obniżoną zdolnością do rozpoznania zróżnicowanego prawdopodobieństwa uzyskania nagrody w teście z probabilistycznym odwróceniem jak i mniejszą motywacją do uzyskania tego typu nagrody. Wyniki modelowania obliczeniowego dostarczyły ciekawych obserwacji, ponieważ w przypadku grupy alkoholowej kohorty II dopasowania modeli obliczeniowych nie odbiegały od grupy kontrolnej, w której nagroda została zastąpiona wodą, co można interpretować jako zaburzone uczenie się. Na podstawie analizy optymalnych wartości parametrów różnice między grupą alkoholową a pozostałymi substancjami zostały określone dla najprostszych modeli ale nie dla rozwiązań mieszanych (Tabela 5–3). Nie określono zaburzenie procesu uczenia się w grupie alkoholowej.

Jednak stwierdzenie, że myszy podejmują decyzję w jednakowy sposób niezależny od rodzaju nagrody byłoby zbyt dużym uogólnieniem. Dodanie sacharyny do alkoholu istotnie zwiększa preferencję tej mieszaniny względem roztworu samego alkoholu (Yoneyama i in., 2008), bez zmiany kaloryczności. Do badań opisanych w niniejszej rozprawie włączono grupę zwierząt, dla których dostępna była mieszanina alkoholu i sacharyny. W przypadku tej grupy nie zanotowano zachowania zmienionego względem grup sacharynowych, poza wartościami parametrów modelu „podwójnie mieszanego”, wskazującymi na silniejszy efekt wartości rezultatu działania i efekt przewidywania w grupie sięgającej po roztwór samej sacharyny. Dodatkowo, na podstawie analizy z wykorzystaniem regresji logistycznej, wskazano efekt czasu znamienne różny wyłącznie pomiędzy grupą alkoholową i z mieszaniną substancji. Sformułowanie wiarygodnej interpretacji wskazanych obserwacji wymaga dalszych badań.

Dla sacharyny podawanej gryzoniom w wysokim stężeniu (5%) zaznaczono możliwy efekt kancerogeny w układzie moczowo-płciowym i innych narządach wewnętrznych (Reuber, 1978), który nie jest oczekiwany dla ludzi (Guy, 2005). Natomiast w badaniach opisywanych w niniejszej rozprawie zastosowano pięćdziesięciokrotnie niższe stężenie sacharyny, dlatego wykluczono możliwość wystąpienia takiego efektu. Sacharyna w stężeniach w zakresie 0,01 do 1% jest typowo wykorzystywana w badaniach z udziałem gryzoni.

Pomimo awersji do etanolu, w badaniach uzależnienia, zwierzęta w odpowiednich warunkach, sięgają po niego aż do odurzenia (Rhodes i in., 2007). W schematach doświadczalnych procesu uzależnienia wprowadza się naprzemienne fazy dostępu i braku dostępu do substancji mające na celu odwzorowanie czasu odstawienia (Leeman i in., 2010). Jednak czy odwzorowuje to rzeczywisty sposób sięgania po substancję uzależniającą u ludzi pozostaje trudne do zweryfikowania (Ahmed, 2018). Dlatego zamiast naprzemiennych okresów dostępności alkoholu, w niniejszych badaniach z wykorzystaniem klatki IntelliCage, schemat doświadczenia charakteryzował się ciągłym dostępem do substancji oraz niskim stężeniem roztworu etanolu (4% w./obj.). Z tego powodu założeniem testu nie było monitorowanie zmian towarzyszących uzależnieniu, tylko opisanie efektu spontanicznego spożywaniu alkoholu w niskim stężeniu. Ze

względem m.in. na awersyjny smak alkoholu w doświadczeniach niezbędny jest wydłużony czas adaptacji do używania tej substancji. Nieoczekiwanie, w badaniach przedstawionych w niniejszej rozprawie zanotowano zależność wyników grupy alkoholowej od badanej kohorty, które różniły się wyłącznie schematem etapu adaptacji. W kohorcie z 14 z naprzemiennymi fazami adaptacji aktywność zwierząt w doświadczeniu była niższa oraz zanotowano wykształcenie się silnej indywidualnej preferencji narożnika. Dla drugiej kohorty zwierząt, z 9 fazami adaptacji i kilkukrotnie wydłużoną ostatnią fazą tego etapu, preferencja opcji o wyższym prawdopodobieństwie nagrody była nie wyższa niż w grupie kontrolnej, sugerując zaburzone uczenie się. Podobieństwo tej grupy do grupy kontrolnej widoczne było także w dopasowaniu modeli uczenia ze wzmocnieniem, które było niewiele większe od modelu losowej decyzji, oraz charakteryzowało się zróżnicowaniem wewnątrz grupy. Zaburzone uczenie się w tej kohorcie sugerowały także wartości wolnych parametrów zbliżone do grupy kontrolnej.

Wykorzystany test nie stanowił jednak próby odwzorowania procesu uzależnienia. Zarówno u gryzoni jak u ludzi wystawienie na działanie etanolu powoduje uzależnienie dla niskiego odsetka populacji (Grant i in., 2015; Ahmed, 2018; Augier i in., 2018), co nadal nie zostało wyjaśnione. Aktualne badania wskazują, że istotny czynnik stanowią warunki środowiskowe (Ahmed, 2018). Dostęp do alternatywnej nagrody czy innych aktywności behawioralnych, u gryzoni, może prowadzić do zmniejszenia lub nawet rezygnacji z samopodawania substancji uzależniającej, co wskazywałoby na sięganie po substancję z powodu braku alternatywy (Ahmed, 2018). Przykładowo, w badaniach Erica Augier'a i współpracowników z 2018r, w teście samopodawania wykonanym przez szczury w klatce do warunkowania instrumentalnego, w sytuacji jednoczesnego wyboru między sacharyną i alkoholem, zachowanie wskazujące na uzależnienie zanotowano dla mniejszości grupy, podczas gdy dla większości zwierząt to etanol uzyskał niższą preferencję (Augier i in., 2018). Także w badaniach Magdaleny Smutek i współpracowników, opisujących spontaniczne zachowanie myszy w teście samopodawania etanolu, w klatce IntelliCage, w której zwierzęta testowane były w grupie w sposób ciągły przez kilka tygodni, nie wskazano oczekiwanego podziału na grupy o niskiej i wysokiej wrażliwości oraz nie potwierdzono dla tych zwierząt objawów uzależnienia (Smutek i in., 2014). W analogicznych badaniach, podział myszy na grupy o większej i mniejszej skłonności do alkoholu, wskazał, że czynnikami przewidującymi kompulsywne sięganie po alkohol mogą być cechy indywidualne związane z lękowością, objawiające się zmniejszonym poszukiwaniem nowości (ang. *novelty seeking*), zwiększoną kompulsywnością i impulsywnością oraz obniżoną odpornością na awersyjne konsekwencje działania (ang. *resistance to punishment*) (Radwanska i Kaczmarek, 2012). Występowanie zróżnicowania we wzorcu sięgania po alkohol w obrębie grupy można wytłumaczyć także hipotezą dotyczącą hierarchii społecznej, według której osobniki o niższej pozycji, odpowiadającej wykluczeniu społecznemu, czy też bardziej podległe, będą charakteryzowały się zmienionym spożyciem substancji uzależniającej (Heilig i in., 2016). Mogłoby to wytłumaczyć zaobserwowane zróżnicowanie wewnątrzgrupowe pod względem efektu nagrody obserwowane-

nym jako czynnik określający szansę powtórzenia poprzedniego wyboru, w kohorcie I grupy alkoholowej.

6.1.3 Transmisja serotoninowa ma znaczenie dla zależności decyzji i czasu działania ale nie dla strategii podejmowania decyzji.

Genetycznie modyfikowane myszy pozbawione receptora 5-HT₇, w zaprezentowanym teście, wykonywały więcej prób uzyskania nagrody. Wykonanie większej liczby prób może oznaczać szybsze uczenie się poprzez większą liczbę szans na poznanie działania środowiska. Natomiast w tym przypadku nie wiązało się to jednak z preferencją opcji z pewniejszą nagrodą, wielkością efektu nagrody, wielkością efektu związanego z preferencją narożnika ani wielkością efektu czasu, analizowanych w regresji logistycznej. W wynikach modelowania widoczne było częste wystąpienie wartości parametrów równych 0, co sugerowało zaburzony proces uczenia się lub założenia modelu nieodpowiadające przebiegowi zjawiska. Niską wartość parametru α można interpretować alternatywnie jako większe poleganie na wcześniejszych działaniach. Na tej podstawie można spekulować o wystąpieniu większej perseweracji u tych zwierząt. Natomiast w celu jednoznacznej interpretacji potrzebne są dalsze badania w odniesieniu do zjawiska dyskontowania oraz czasu decyzji jak i czasu sięgnięcia po nagrodę. Należy rozróżnić wskazane aspekty czasu działania, ponieważ czas reakcji na bodziec ale nie czas sięgnięcia po nagrodę może być krótszy, co wykazano dla szczurów niemal całkowicie pozbawionych serotoniny w mózgowiu poprzez dokomorowe podanie 5,7-dihydroksy-tryptaminy (Bari i in., 2010).

W zachowaniu zwierząt linii NR1^{Tph2CreERT2}, testowanym z wykorzystaniem IntelliCage, wskazano zmniejszony efekt czasu na prawdopodobieństwo powtórzenia poprzedniego wyboru, określony za pomocą regresji logistycznej, w porównaniu do własnej grupy kontrolnej. Biorąc pod uwagę występowanie w tej grupie efektu nagrody, istotnej preferencji strony o większej pewności nagrody oraz braku znamiennej różnicy dla efektu miejsca określającego preferencję narożnika, w porównaniu do grupy kontrolnej, interpretacja wyniku w oparciu o zwiększoną tendencję do powtórzenia tego samego wyboru jest mało prawdopodobna. Oznacza to, zgodnie z doniesieniami (np. (Fonseca i in., 2015; Lottem i in., 2018)), że znaczenie serotoniny w procesach poznawczych związane jest z czasem wyboru. Uzyskany wynik jest jednak niewystarczający do udowodnienia hipotezy o znaczeniu serotoniny w utrzymywaniu pozytywnego subiektywnego przekonania (Miyazaki i in., 2018) czy większej wytrwałości. Także dla wyników testu przeprowadzonego w klatce do warunkowania instrumentalnego, pomimo widocznego uczenia się i adaptacji do zadania, nie określono różnic zależnych od genotypu. Jednak w tym przypadku nie określono także efektu związanego z czasem podjęcia działania i sięgnięcia po nagrodę. Wyniki modelowania obliczeniowego nie wykazały znamiennych różnic zależnych od genotypu zarówno dla doświadczenia z wykorzystaniem klatki IntelliCage jak i klatki do warunkowania instrumentalnego.

Wątpliwości może dostarczyć preferencja sacharyny niższa od oczekiwanej w doświadczeniu z wykorzystaniem klatki IntelliCage dla myszy linii NR1^{Tph2CreERT2}. Taki efekt był najprawdopodobniej związany z krótszym okresem adaptacji dla dwóch badanych kohort (po 4 fazy) w porównaniu do pozostałych grup (po 5 faz) w doświadczeniu. Miara preferencji obliczona została z danych ostatnich 96 godzin adaptacji, a więc w tym przypadku włączała czas, w którym preferencja mogła być jeszcze nieustabilizowana. Nie powinno to wpłynąć na wykonanie zadania przez zwierzęta. Alternatywnie, na preferencją obliczoną z liczby liżnięć mogły wpłynąć trudności techniczne doświadczenia związane z kalibracją detektorów.

W niniejszych badaniach dorosłym myszom podawany był tamoksifen indukujący mutację powodującą stopniowy zanik wskazanych receptorów na komórkach serotoninowych. Należy rozważyć występowanie u tych zwierząt potencjalnych zmian związanych z podawaniem tego związku. Nie zaobserwowano, żeby podanie tamoksifenu zaburzyło zdolności do wykonania testu przez zwierzęta. Nie zanotowano u tych zwierząt zmienionej aktywności lokomotorycznej, określonej poprzez liczbę prób uzyskania nagrody, zmienionej preferencji nagrody ani spadkowej tendencji w liczbie prób uzyskania nagrody wraz z czasem trwania doświadczenia.

Podobnie jak w przypadku zwierząt linii *Htr7* zanotowano różnice między grupą kontrolną linii NR1^{Tph2CreERT2} a grupą myszy C57BL/6J, co wskazuje na potencjalne zmiany procesu podejmowania decyzji związane z uczeniem się grupowym. Różnice występujące między grupą kontrolną a grupą myszy C57BL/6J można uzasadnić gorszym uczeniem się zwierząt kontrolnych w grupie heterogenicznej ze zwierzętami z mutacją (Kiryk i in., 2011). Stanowi to pole do dalszych badań.

Coraz intensywniej badanym aspektem roli układu serotoninowego w uczeniu się, zwłaszcza badanym z wykorzystaniem testów z probabilistycznym odwróceniem, jest kodowanie przez neurony serotoninowe oczekiwanej niepewności wyniku wyboru (Behrens i in., 2007). W najnowszych badaniach Coopera Grossmana i współpracowników wskazano, że efekt behawioralny optogenetycznego wyłączenia neuronów serotoninowych jąder szwu jest zgodny z modelem teoretycznym zakładającym zależność tempa uczenia, czy też alternatywnie wielkości kroku (α), w modelu uczenia się ze wzmocnieniem od niepewności rezultatu (ang. *expected uncertainty*) (Grossman i in., 2022). Taki wynik może tłumaczyć wolniejsze uczenie się w kontekście opisywanego wydłużenia czekania czy zwiększoną persewerację, obserwowane w badaniach roli układu serotoninowego w podejmowaniu decyzji (Grossman i in., 2022). Wskazany model odwołuje się do koncepcji, w której bardziej nieoczekiwane zdarzenie będzie dostarczało więcej informacji, a więc będzie bardziej użyteczne w procesie uczenia się (Pearce i Hall, 1980), co stanowi kluczowy element modelu „uwagowego” (Bai i in., 2014) zastosowanego w niniejszych badaniach. Ten model odróżnia się przez założenie niestacjonarnego procesu uczenia się, w którym wartość parametru wielkości kroku (α) zależy każdorazowo od wielkości błędu przewidywania. O ile sposób sformułowania matematycznego tego założenia we wspomnianych badaniach różnił się od zastosowanego w niniejszej rozprawie to uzyskane wyniki wskazały

zgodnie, że wprowadzenie takiego rozszerzenia nie dostarcza wyraźnego polepszenia dopasowania do zachowania myszy w warunkach kontrolnych. Dodatkowo, nie zaobserwowano ani zmniejszonego dopasowania modelu „uwagowego” do zachowania myszy transgenicznych ani zmiany w parametrze μ , wskazującym na siłę zależności wielkości kroku i wartości błędu przewidywania. Nie oznacza to jednak, że uczenie się zaobserwowane w tym teście nie byłoby dobrze tłumaczone przez założenie o niestacjonarności procesu uczenia się (Belkaid i in., 2020), chociaż, biorąc pod uwagę schemat doświadczalny przeprowadzonego testu, sposób wykonania zadania wskazujący na adaptację do środowiska stacjonarnego byłby uzasadniony.

6.2 Modelowanie uczenia ze wzmocnieniem

6.2.1 Określenie modelu najlepiej przewidującego obserwowany proces uczenia

Testowane modele oparte o algorytm *Q-learning* lepiej tłumaczyły zachowanie niż losowe podejmowanie decyzji czy model opierający się o zasadę prostej odpowiedzi na efekt działania ‘zostań po wygranej, zmień po przegranej’. Na podstawie dopasowania do danych nie określono jednoznacznie najlepszego modelu z siedmiu proponowanych, w tym trzech rozszerzeń opracowanych na potrzeby niniejszych badań (model „podwójnie mieszany”, „rozpad Q” oraz „rozpad β ”). Najbardziej użytecznym modelem, z grupy testowanych rozwiązań, był model „podwójnie mieszany”.

Niższa wartość ułamka poprawnych przewidywań modelu w doświadczeniach z wykorzystaniem aparatu IntelliCage, w porównaniu do wyników dla doświadczenia z wykorzystaniem klatki do warunkowania instrumentalnego, może wskazywać, że zjawisko przebiega w inny sposób niż w założeniach modelu. Mogły to spowodować dodatkowe czynniki, stanowiące część zmienności w danych, albo wykonanie zadania w sposób zbliżony do losowego podejmowania decyzji. W warunkach laboratoryjnych można wymusić losowe wykonywanie zadania przez myszy (Belkaid i in., 2020) natomiast w niniejszych badaniach jest to mało prawdopodobne ze względu na opisane, powtarzające się między doświadczeniami, wzorce zachowania wskazujące na występowanie uczenia się. Jak wynika z przedstawionych analiz, działanie zwierząt było ukierunkowane na cel i opierało się na mechanizmie wzmocnienia, leżącego także u podłoża założenia modeli. Natomiast sposób wykonania zadania probabilistycznego odwrócenia w zaprezentowanych warunkach eksperymentalnych nie został wcześniej opisany. Wiązało się to z koniecznością dostosowania i testowania aktualnych rozwiązań.

Należy zauważyć, że do porównania włączono modele redukowalne do założeń prostszych rozwiązań za pomocą wartości parametrów. Zgodnie z tym przy równych wartościach parametru α_+ i α_- model „wartościowo zależny” nie odbiega założeniami od modelu „podstawowego”. Analogicznie, w przypadku modelu „podwójnie mieszanego” jego założenia mogą zostać zredukowane do rozwiązania „podstawowego”, „wartościowo zależnego”, „urojonego”, i „mieszanego”. Dlatego wartość dopasowania dla tego modelu w porównaniu do wskazanych rozwiązań

powinna wskazać o ile lepsze dopasowanie do danych uzyskuje się poprzez możliwość rozszerzenia modelu o założenie aktualizacji dla niewybranej opcji i zróżnicowanej wrażliwości na pozytywne zdarzenie uzyskania nagrody i nieprzyjemne odczucie jej braku. Na podstawie analiz widoczne było zwiększone dopasowanie rozwiązań „mieszanych” umożliwiających przyjęcie zróżnicowanej wielkości kroku (α) w zależności od rezultatu wyboru dla obu możliwych opcji do wyboru.

Trudność w analizie z wykorzystaniem modelowania obliczeniowego stanowi wnioskowanie na podstawie dopasowania modelu do danych. Skrajnie wysokie dopasowanie wskazuje na nadmierne dopasowanie modelu do bieżącego zbioru danych i tym samym niską zdolność do tłumaczenia zjawiska w innych warunkach, z kolei zbyt niskie dopasowanie oznacza niskie zdolności predykcyjne modelu i tym samym nie tłumaczy odpowiednio obserwowanego zjawiska. Także zwiększona liczba parametrów modelu powoduje większe dopasowanie modelu do danych poprzez włączenie zmienności spowodowanej losowością (Daw, 2011), dlatego, w tym zastosowaniu, lepszą strategią modelowania jest wykorzystanie jak najprostszych modeli możliwych do bezpośredniej interpretacji. Do określenia dopasowania modelu do danych wykorzystano dwie miary, które dostarczyły odrębnych wyników. Miara ułamka poprawnych przewidywań posłużyła do odrzucenia modeli nieodpowiadających obserwowanemu zachowaniu, nie tłumaczących zachowania lepiej niż przyjęcie losowego przebiegu zjawiska, a więc nieużytecznych do przewidywania zachowania zwierząt.

Zastosowanie wartości AIC pozwoliło na wyraźniejsze zróżnicowanie dopasowania testowanych modeli. Ograniczenie zastosowania tej miary stanowiła zależność wartości od liczby prób uzyskania nagrody wykonanych przez myszy. Różnica wartości AIC dla bardzo dobrego modelu (hipotetyczne $p = 0,8$ w każdej próbie) od losowego przewidywania ($p = 0,5$) będzie znacznie większa dla osobnika wykonującego większą liczbę prób. Ważną zaletą logarytmicznego przekształcenia funkcji dopasowania *nll* zależnej od prawdopodobieństwa wyboru jest to, że różnica wartości *nll* czy AIC dla jednakowej różnicy prawdopodobieństwa wyboru maleje wraz z jego wartością. Pozwala to na lepsze szacowanie dopasowania modelu w sytuacji znacznej niepewności. Natomiast w celu zniwelowania efektu zróżnicowania wewnątrz grupy, związanego ze zróżnicowaną liczbą prób wykonanych przez zwierzęta, w analizie, zamiast bezwzględnych wartości AIC, określono dla każdego osobnika różnicę w wartości AIC modelu względem rozwiązania referencyjnego (ΔAIC) oraz prawdopodobieństwo modelu względem pozostałych rozwiązań (waga Akaike) (Burnham i Anderson, 2004).

Postulowanym sposobem określenia dopasowania modeli byłoby także porównanie wyników z wynikami uzyskanymi z symulacji przeprowadzonymi poprzez generowanie danych na podstawie zakresu wartości parametrów (Conway i White, 2015; Palminteri i in., 2017; Wilson i Collins, 2019). Jednym z podstawowych sposobów odnoszących się do tego problemu stanowi walidacja krzyżowa (sprawdzian krzyżowy, ang. *cross validation*) (Conway i White, 2015) na podstawie zestawu danych dzielonego na pomniejsze części np. walidacja krzyżowa *k*-krotna (ang. *k-fold cross validation*) czy walidacja krzyżowa z pominięciem jednego wyjścia (ang.

leave-one-out cross validation) (Lin i in., 2007). Wskazane metody nie zostały zastosowane ze względu na możliwe wystąpienie dodatkowych wzorców zachowania, zależnych od etapu doświadczenia, sekwencji wyborów czy czasu mijającego między kolejnymi próbami, które zostałyby pominięte podczas interwałowego czy losowego podziału danych stosowanego w tych metodach. Na przykładzie danych z doświadczenia z wykorzystaniem klatki IntelliCage, przeszkodę stanowiła także różna liczba prób wykonanych przez zwierzęta. Im więcej prób wykona mysz w trakcie fazy odwrócenia tym więcej uzyska informacji o środowisku i tym większa będzie adaptacja do wykonania zadania. Jako że nie ma sposobu na określenie w jakim zakresie uzyskane dane odpowiadałyby danym rzeczywistym nie zdecydowano się na włączenie wskazanych metod do badań.

6.2.2 Wnioskowanie na podstawie wartości parametrów

Wykorzystany w niniejszych badaniach sposób określania wartości parametrów jest standardowy (Daw, 2011; Wilson i Collins, 2019). Mimo to, w doświadczeniach w klatce IntelliCage, dla części zwierząt, wartości optymalnych parametrów związanych z rzeczywistym uzyskaniem nagrody (α , α_0 , α_+ , α_+ i β) wyniosły 0, co można zinterpretować jako brak uczenia ze wzmocnieniem, przynajmniej na podstawie pozytywnej informacji zwrotnej. Natomiast za pomocą miar behawioralnych wykazano istnienie takiego efektu. Ze względu na złożoność obliczeniową modeli z większą liczbą parametrów określenie wartości optymalnych wiązało się z koniecznością ograniczenia liczby punktów startowych optymalizacji, co z kolei mogło wpłynąć na dokładność wyników. Z tego powodu, w niniejszych badaniach, wartości optymalne parametrów kluczowych dla rzeczywistego wzmocnienia wynoszące 0 mogły stanowić przybliżenie skrajnie małej wartości. Niskie wartości wskazanych parametrów mogły wiązać się charakterystyką testu, w którym zwierzęta mogą być zaangażowane w inne aktywności oraz są, potencjalnie, mniej zmotywowane do sięgania po nagrodę niż w doświadczeniach z wykorzystaniem klatki do warunkowania instrumentalnego lub za dopasowanie modelu do danych odpowiadałyby inne czynniki poza wzmocnieniem, jak w przypadku modelu „podwójnie mieszanego z rozpadem Q”. Stanowi to pole dalszych badań.

Najbardziej użytecznym modelem, z grupy testowanych rozwiązań jest model „podwójnie mieszany”. W celu dokładnego przeanalizowania wartości parametrów tego modelu zastosowano porównania parami dla określenia efektu wartości (α_+ i α_-), przewidywanego rezultatu dla uzyskania nagrody (α_+ i α_+) oraz przewidywanego rezultatu braku nagrody (α_- i α_-). Różnice wartości pomiędzy wartościami parametrów odnoszące się do wskazanych efektów mogą posłużyć do ilościowego określenia sposobu rozwiązania zadania, a więc stanowić użyteczne narzędzie badawcze. Ograniczeniem, które należy rozważyć, w kontekście zastosowania w innych badaniach, byłaby zależność dopasowania i wartości optymalnych parametrów, od częstości każdego z możliwych działań, czyli (1) wyboru rzeczywistego i nagrody uzyskanej w bieżącej próbie (związanego z parametrem α_+), (2) wyboru alternatywnego i nagrody uzy-

skanej w bieżącej próbie ($\alpha_{-\phi}$), (3) wyboru rzeczywistego i braku nagrody w bieżącej próbie (α_{-p}) oraz (4) wyboru alternatywnego i braku nagrody w bieżącej próbie ($\alpha_{-\phi}$) (Tabela 2—1).

Na tej podstawie wskazano na większą wielkość kroku dla zdarzenia rzeczywistego uzyskania nagrody w porównaniu do zdarzenia przewidywanego oraz rzeczywistego braku nagrody w grupie sacharynowej myszy C57BL/6J testowanych w klatce IntelliCage, ale nie w grupach alkoholowych (Tabela 5—2). Takiego wzorca nie zaobserwowano natomiast dla grupy myszy C57BL/6J testowanych w uproszczonej wersji testu (Tabela 5—5). Mogło to być spowodowane różnicą w sposobie testowania, jednak bardziej prawdopodobnym wytłumaczeniem byłoby znaczne zróżnicowanie wewnątrzgrupowe w wartościach parametrów, widoczne na przykładzie grupy kontrolnej NR1^{Tph2CreERT2} testowanej w klatce IntelliCage (Rycina 5-25). Zróżnicowanie wewnątrz grupy może wynikać z cechy wrodzonej osobników do większej wagi przykładanej do zdarzenia o korzystnym bądź niekorzystnym skutku (Noworyta-Sokolowska i in., 2019). Natomiast dokładnie odwrotnie — większą wartość kroku dla braku nagrody w porównaniu do jej rzeczywistego uzyskania oraz dla oczekiwanego uzyskania nagrody w porównaniu do jej rzeczywistego uzyskania określono w grupach kontrolnych *Htr7* oraz NR1^{Tph2CreERT2} (Tabela 5—5). Mogłoby to oznaczać także większą wagę położoną na sytuację braku nagrody w porównaniu do jej uzyskania. Natomiast interpretacja wystąpienia takiego efektu wyłącznie dla grup kontrolnych nie jest jasna, podobnie jak przyczyna większej wartości wielkości kroku dla oczekiwanego braku nagrody nad rzeczywistym brakiem nagrody, u zwierząt NR1^{Tph2CreERT2}, testowanych z wykorzystaniem klatki do warunkowania instrumentalnego (Rycina 5-25).

Ze względu na schemat testu z odwróceniem, w niniejszych badaniach, wybór wyłącznie jednej opcji przez cały czas trwania doświadczenia jest strategią porównywalną do losowego podejmowania decyzji, a więc mniej korzystną niż wybór opcji z pewniejszą nagrodą. Mimo to, w badaniach zaobserwowano skłonność wybranych osobników do wyboru wyłącznie jednej opcji, co było widoczne w wartości preferencji opcji o większym prawdopodobieństwie uzyskania nagrody wzrastającej i opadającej okresowo, co drugą fazę doświadczenia. Stanowi to ograniczenie możliwości wnioskowania o zachowaniu na podstawie nie tylko wskazanego modelu ale ogólniej modelowania obliczeniowego, dla którego nie ma na razie dobrych rozwiązań. Jako, że niniejsze badania stanowiły pierwszą znaną próbę integracji testu z probabilistycznym odwróceniem, z wykorzystaniem klatki IntelliCage, i modelowania obliczeniowego, opartego o uczenie ze wzmocnieniem, starałam się przeanalizować uzyskane dane w sposób jak najbardziej szczegółowy, unikając normalizowania i standaryzowania w odniesieniu do badanych grup.

6.3 Zależność między czasem i strategią podejmowania decyzji

W niniejszych badaniach, na podstawie analizie rozkładu interwału między kolejnymi próbami, zaobserwowano bardzo częste próby uzyskania nagrody w czasie kilku minut od po-

przedniej próby. Wskazuje to występowanie wzorca aktywności zwierząt polegającego na powtarzającym się cyklu, okresu zaangażowania w test, a następnie dłuższego czasu odpoczynku czy innej aktywności.

W okresie mniejszego zaangażowania w test wartość oczekiwanej nagrody mogła ulegać zmianie dlatego rozwiązanie modelowania rozszerzono o efekt zapominania („rozpad Q”) oraz wzrostu niepewności w czasie („rozpad β ”). Zależność czasu mijającego między wyborami jako efektu zapominania oraz efektu wzrostu losowości w czasie nie odpowiadała bezpośrednio tendencji zaobserwowanej w badaniach polegającej na wzrastającym prawdopodobieństwie powtórzenia działania wraz z wydłużaniem czasu od poprzedniej próby. Prawdopodobnie dlatego wskazane rozszerzenia modeli obliczeniowych spowodowały nieznaczne zwiększenie dopasowania do danych. Trudniejszym do wyjaśnienia wynikiem było nadrzędne dopasowanie do danych modelu „podwójnie mieszanego” z rozpadem Q, odnoszącego się do efektu zapominania w czasie, dla którego wartości parametrów związanych z rzeczywistym wyborem nie były dodatnie, sugerując wpływ innych czynników lub trudności w określeniu optymalnych wartości parametrów.

Alternatywne wytłumaczenie braku polepszenia dopasowania dla zaproponowanych w niniejszych badaniach rozszerzeń („rozpad Q” i „rozpad β ”) polega na, zasugerowanej na podstawie wyników modelu regresji logistycznej, tendencji myszy do powtórzenia poprzedniego wyboru wraz z wydłużaniem czasu do kolejnej próby, dla długich interwałów. Według niej, w czasie dłuższego odstąpienia od wykonywania testu wzrastałaby skłonność zwierząt do powtórzenia wyboru, co mogłoby sugerować występowanie konsolidacji śladu pamięciowego ostatniego działania lub awersji do mniej pewnej opcji po dłuższym czasie od wyboru. Jednak takie hipotezy powinny zostać zweryfikowane biorąc pod uwagę także miarę impulsywności czy dyskontowania czasowego. Możliwe jest także włączenie innych sposobów analizy np. w odniesieniu do częstości działania (Lindström i in., 2021), czasu mijającego do większej pewności rezultatu w modelu dryfu i dyfuzji (Pedersen i in., 2017) czy na podstawie analizy powtarzających się zdarzeń (ang. *recurrent event analysis*) (Amorim i Cai, 2015).

Znaczenie interwału decyzji dla typu podjętego działania okazało się złożone. Na podstawie częstości działania „wygrana zostać” oraz analizy z wykorzystaniem regresji logistycznej można stwierdzić, że wraz z wydłużeniem interwału zwiększało się prawdopodobieństwo działania typu „zostań” i ta zależność była prawdziwa niezależnie od rezultatu próby (Jabłońska i in., 2021). Zwierzęta podejmowały kolejną próbę uzyskania nagrody szybciej gdy poprzednia zakończyła się nieprzyznaniem dostępu do napoju. Jednak w tym przypadku rozkłady mają dwie wartości szczytowe częstości występowania interwału do kolejnej próby – dla wartości krótszej niż minuta – oraz w zakresie 6–20 minut. Przy wnioskowaniu o strategii zachowania zwierząt trzeba wziąć pod uwagę częste występowanie prób następujących szybko po sobie i tym samym dopasowanie wartości regresji logistycznej czy opisanych dalej parametrów modeli obliczeniowych także do prób o potencjalnym charakterze eksploracyjnym. Interwały dłuższe

od 60 minut, dla których prawdopodobieństwo działania „zostań”, szacowane z wykorzystaniem modelu regresji logistycznej było wyższe od poziomu losowego, były rzadkie.

Alternatywnie, wzorzec aktywności zwierząt w IntelliCage polegający na wykonywaniu wielu prób uzyskania nagrody, a następnie czasu odstąpienia od zadania, można przedstawić jako aktywność blokowa z naprzemiennymi okresami aktywności i odpoczynku od zadania. Przykładowo, w zadaniu probabilistycznego rozróżnienia wizualnego u młodych dorosłych, analogicznym do wspomnianego doświadczenia Wolframa Schultza i wsp. u naczelnych, rozdzielanie zadania na bloki z kilkudniowymi przerwami skutkowało efektywniejszym zapamiętywaniem wartości opcji po kilku tygodniach (Wimmer i in., 2018). Dodatkowo, w tych badaniach autorzy wskazali, że efektywność uczenia się w tym teście była zależna od indywidualnych zasobów pamięci roboczej (Wimmer i in., 2018). Możliwe, że w badaniach opisywanych w niniejszej rozprawie myszy optymalizują swoją aktywność w zależności od procesów pamięciowych.

Na podstawie przedstawionych w niniejszej rozprawie analiz czasu, sugeruję występowanie trzech różnych strategii decyzji zależnych od czasu decyzji. W mojej koncepcji krótki interwał przewiduje aktywność eksploracyjną, więc wysoką szansę zmiany w stosunku do poprzedniego wyboru, skrajnie długi wskazuje na reakcję powtórzenia tego samego wyboru niezależnie od wyniku poprzedniego działania. Z kolei wartości pomiędzy nimi pozwalałyby na przewidywanie działania zgodne z zasadą wzmocnienia, biorąc pod uwagę probabilistyczną oraz niestacjonarną charakterystykę testu.

7. Wnioski

W opracowanym modelu behawioralnym testu probabilistycznego wyboru, przybliżającym warunki naturalne, widoczne było uczenie ze wzmocnieniem. Zachowanie zwierząt opisano z wykorzystaniem modeli obliczeniowych. Wśród zastosowanych rozwiązań lepsze dopasowanie do danych oraz możliwości interpretacji oferowały modele mieszane, w których zakładane jest myślenie o konsekwencjach niepodjętego działania oraz zróżnicowana wrażliwość na pozytywną i negatywną/neutralną informację zwrotną.

Pomimo właściwości wzmacniających etanolu, dla grupy alkoholowej zanotowano mniejszą preferencję opcji o wyższym prawdopodobieństwie uzyskania nagrody oraz mniejszy efekt nagrody w przewidywaniu prawdopodobieństwa działania powtórzenia decyzji. Było to prawdopodobnie spowodowane awersyjnym smakiem alkoholu, ponieważ analogicznych różnic nie zanotowano między grupą sacharynową i z mieszaniną sacharyny i etanolu. Oznacza to także, że hipoteza o zaburzeniu procesu podejmowania decyzji pod wpływem alkoholu spożywanym o dowolnym czasie, chronicznie i w niskim stężeniu nie została potwierdzona. Wyniki modelowania nie dostarczyły jednoznacznej koncepcji wpływu rodzaju nagrody na strategię podejmowania decyzji, chociaż na podstawie miar behawioralnych i dopasowania modeli można spekulować o przyjęciu mniej efektywnej strategii wykonania zdania w grupie alkoholowej. W celu jednoznacznego określenia takiego efektu konieczne byłyby dalsze badania.

Myszy pozbawione receptorów 5-HT₇ nie wykazały zaburzenia procesu uczenia się oraz podejmowania decyzji. Dla grupy eksperymentalnej zanotowano więcej prób uzyskania nagrody w porównaniu do grupy kontrolnej, jednak nie przełożyło się to na inne miary behawioralne.

Podobnie, indukowany w dorosłości zanik receptorów NMDA nie zaburzył zachowania testowanego zarówno z wykorzystaniem klatki IntelliCage jak i klatki do warunkowania instrumentalnego chociaż wpłynął na wielkość efektu czasu na decyzję opisany w teście z wykorzystaniem klatki IntelliCage. Wskazany wynik potwierdza, że znaczenie serotoniny w procesach poznawczych związane jest z czasem wyboru oraz nie przeczy hipotezie o znaczeniu serotoniny w utrzymywaniu pozytywnego subiektywnego przekonania i tym samym większej wytrwałości.

W doświadczeniach z wykorzystaniem systemu IntelliCage zaobserwowano zależność między czasem decyzji i wyborem polegającą na zwiększaniu się prawdopodobieństwa powtórzenia działania wraz z wydłużeniem interwału od poprzedniego wyboru. Oznacza to, że im więcej czasu mija od ostatniej próby tym mysz będzie bardziej skłonna sięgnąć po opcję, którą wybrała poprzednio. Wysunęłam hipotezę, że po dłuższym czasie mysz sięgnie po ostatnią opcję, o której pamięta, a więc, że takie zachowanie będzie związane z procesem zapominania. Jednak wprowadzone do modeli obliczeniowych modyfikacje, odwołujące się do efektu zapominania („rozkład Q”) i wzrostu niepewności („rozkład β ”) z czasem dostarczyły jedynie nieznacznego

polepszenia dopasowania modeli do danych. Możliwe, że zależność strategii decyzji i czasu jest jeszcze bardziej złożona, np. tak jak przedyskutowano, dla krótkich interwałów dominuje aktywność eksploracyjna, dla długich zjawisko zapominania, a dla pośrednich między nimi, z indywidualnymi granicami, można efektywnie opisać z wykorzystaniem narzędzi uczenia ze wzmocnieniem. Alternatywnie, do analizy efektu czasu można zastosować modelowanie innego typu, np. celujące w odpowiedź na pytanie o zależność między siłą wzmocnienia i częstotliwością sięgania po nagrodę.

Metody opisane w rozprawie pozwalają na ilościowy opis przebiegu uczenia ze wzmocnieniem w teście probabilistycznego odwrócenia. Zaproponowane w niniejszej rozprawie rozszerzenia modeli można wykorzystać zarówno w badaniach laboratoryjnych z wykorzystaniem zwierząt modelowych jak i w analizie danych pochodzących od wolontariuszy czy osób chorych, co wskazuje na znaczną interdyscyplinarność i translacyjność badań. Badania włączające myślenie modele rozwoju chorób, dla których notuje się objawy w procesie podejmowania decyzji, np. model rozwojowy choroby Alzheimera, choroby Parkinsona, depresji czy schizofrenii oraz substancje o potencjale leczniczym, mają potencjał do ilościowego opisu przebiegu zaburzenia, określenia potencjalnego celu terapeutycznego oraz oceny wpływu zastosowanych oddziaływań farmakologicznych.

8. Piśmiennictwo

- Abraham K. P., Salinas A. G., Lovinger D. M., 2017, *Alcohol and the Brain: Neuronal Molecular Targets, Synapses, and Circuits*, „Neuron” 96, nr 6: 1223-1238,
- Ahmed S. H., 2018, *Trying to make sense of rodents' drug choice behavior*, „Progress in Neuro-Psychopharmacology and Biological Psychiatry” 87: 3-10,
- Akaike H., 1974, *A new look at the statistical model identification*, „IEEE Transactions on Automatic Control” 19, nr 6: 716-723,
- Alhadeff A. L., Goldstein N., Park O., Klima M. L., Vargas A., Betley J. N., 2019, *Natural and Drug Rewards Engage Distinct Pathways that Converge on Coordinated Hypothalamic and Reward Circuits*, „Neuron” 103, nr 5: 891-908.e6,
- Amorim L. D. A. F., Cai J., 2015, *Modelling recurrent events: a tutorial for analysis in epidemiology*, „International Journal of Epidemiology” 44, nr 1: 324-333,
- Augier E., Barbier E., Dulman R. S., Licheri V., Augier G., Domi E., ... Heilig M., 2018, *A molecular mechanism for choosing alcohol over an alternative reward*, „Science” 360, nr 6395: 1321-1326,
- Awasthi J. R., Tamada K., Overton E. T. N., Takumi T., 2021, *Comprehensive topographical map of the serotonergic fibers in the male mouse brain*, „Journal of Comparative Neurology” 529, nr 7: 1391-1429,
- Bai Y., Katahira K., Ohira H., 2014, *Dual learning processes underlying human decision-making in reversal learning tasks: functional significance and evidence from the model fit to human behavior*, „Frontiers in Psychology” 5,
- Bari A., Theobald D. E., Caprioli D., Mar A. C., Aidoo-Micah A., Dalley J. W., Robbins T. W., 2010, *Serotonin Modulates Sensitivity to Reward and Negative Feedback in a Probabilistic Reversal Learning Task in Rats*, „Neuropsychopharmacology” 35, nr 6: 1290-1301,
- Becquet D., Héry M., Deprez P., Faudon M., Fache M. P., Giraud P., Héry F., 1993, *N-Methyl-d-Aspartic Acid/Glycine Interactions on the Control of 5-Hydroxytryptamine Release in Raphe Primary Cultures*, „Journal of Neurochemistry” 61, nr 5: 1692-1697,
- Behrens T. E. J., Woolrich M. W., Walton M. E., Rushworth M. F. S., 2007, *Learning the value of information in an uncertain world*, „Nature Neuroscience” 10, nr 9: 1214-1221,
- Belkaid M., Bousseyrol E., Durand-de Cuttoli R., Dongelmans M., Duranté E. K., Ahmed Yahia T., ... Faure P., 2020, *Mice adaptively generate choice variability in a deterministic task*, „Communications Biology” 3, nr 1: 34,
- Berke J. D., 2018, *What does dopamine mean?*, „Nature Neuroscience” 21, nr 6: 787-793,
- Bernardin F., Maheut-Bosser A., Paille F., 2014, *Cognitive Impairments in Alcohol-Dependent Subjects*, „Frontiers in Psychiatry” 5, nr JUL: 1-6,
- Beroun A., Nalberczak-Skóra M., Harda Z., Piechota M., Ziółkowska M., Cały A., ... Radwanska K., 2018, *Generation of silent synapses in dentate gyrus correlates with development of alcohol addiction*, „Neuropsychopharmacology” 43, nr 10: 1989-1999,
- Berridge K. C., 1996, *Food reward: Brain substrates of wanting and liking*, „Neuroscience & Biobehavioral Reviews” 20, nr 1: 1-25,
- Berridge K. C., 2000, *Reward learning: Reinforcement, incentives, and expectations*, *Psychology of Learning and Motivation - Advances in Research and Theory*, t. 40, 223-278,
- Bijoch L., Borczyk M., Czajkowski R., 2020, *Bases of Jerzy Konorski's theory of synaptic*

- plasticity, „European Journal of Neuroscience” 51, nr 9: 1857-1866,
- Birrell J. M., Brown V. J., 2000, *Medial Frontal Cortex Mediates Perceptual Attentional Set Shifting in the Rat*, „The Journal of Neuroscience” 20, nr 11: 4320-4324,
- Bonaventure P., Dugovic C., Kramer M., De Boer P., Singh J., Wilson S., ... Lovenberg T. W., 2012, *Translational Evaluation of JNJ-18038683, a 5-Hydroxytryptamine Type 7 Receptor Antagonist, on Rapid Eye Movement Sleep and in Major Depressive Disorder*, „Journal of Pharmacology and Experimental Therapeutics” 342, nr 2: 429-440,
- Bryce C. A., Floresco S. B., 2021, *Central CRF and acute stress differentially modulate probabilistic reversal learning in male and female rats*, „Behavioural Brain Research” 397, nr August 2020: 112929,
- Burnham K. P., Anderson D. R., 2004, *Multimodel Inference*, „Sociological Methods & Research” 33, nr 2: 261-304,
- Byrnes J. P., Miller D. C., Schafer W. D., 1999, *Gender differences in risk taking: A meta-analysis.*, „Psychological Bulletin” 125, nr 3: 367-383,
- Carr K. D., 2007, *Chronic food restriction: Enhancing effects on drug reward and striatal cell signaling*, „Physiology & Behavior” 91, nr 5: 459-472,
- Choleris E., Galea L. A. M., Sohrabji F., Frick K. M., 2018, *Sex differences in the brain: Implications for behavioral and biomedical research*, „Neuroscience & Biobehavioral Reviews” 85, nr January 2017: 126-145,
- Cieślak P. E., Ahn W.-Y., Bogacz R., Rodriguez Parkitna J., 2018, *Selective Effects of the Loss of NMDA or mGluR5 Receptors in the Reward System on Adaptive Decision-Making*, „eneuro” 5, nr 4: ENEURO.0331-18.2018,
- Clarke H. F., Roberts A. C., 2011, *Reversal learning in fronto-striatal circuits: a functional, autonomic, and neurochemical analysis*, w: M. R. Delgado, E. A. Phelps, T. W. Robbins (red.), *Decision Making, Affect, and Learning*, 205-234,
- Collins A. G. E., Frank M. J., 2012, *How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis*, „European Journal of Neuroscience” 35, nr 7: 1024-1035,
- Conway D., White J. M., 2015, *Uczenie maszynowe dla programistów*, ,
- Cools R., 2015, *Neuropsychopharmacology of Cognitive Flexibility*, *Brain Mapping*, t. 3, 349-353,
- Daw N. D., 2011, *Trial-by-trial data analysis using computational models*, *Decision Making, Affect, and Learning*, 3-38,
- Daw N. D., Kakade S., Dayan P., 2002, *Opponent interactions between serotonin and dopamine*, „Neural Networks” 15, nr 4-6: 603-616,
- den Ouden H. E. M., Daw N. D., Fernandez G., Elshout J. A., Rijpkema M., Hoogman M., ... Cools R., 2013, *Dissociable Effects of Dopamine and Serotonin on Reversal Learning*, „Neuron” 80, nr 4: 1090-1100,
- Di Chiara G., Imperato A., 1988, *Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats.*, „Proceedings of the National Academy of Sciences” 85, nr 14: 5274-5278,
- Dick D. M., Smith G., Olausson P., Mitchell S. H., Leeman R. F., O'Malley S. S., Sher K., 2010, *Understanding the construct of impulsivity and its relationship to alcohol use disorders*, „Addiction Biology” 15, nr 2: 217-226,
- Endo T., Maekawa F., Vöikar V., Haijima A., Uemura Y., Zhang Y., ... Kakeyama M., 2011, *Automated test of behavioral flexibility in mice using a behavioral sequencing task in IntelliCage*, „Behavioural Brain Research” 221, nr 1: 172-181,

- Feil R., Brocard J., Mascrez B., LeMeur M., Metzger D., Chambon P., 1996, *Ligand-activated site-specific recombination in mice.*, „Proceedings of the National Academy of Sciences” 93, nr 20: 10887-10890,
- Fiorillo C. D., Tobler P. N., Schultz W., 2003, *Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons*, „Science” 299, nr 5614: 1898-1902,
- Fonseca M. S., Murakami M., Mainen Z. F., 2015, *Activation of Dorsal Raphe Serotonergic Neurons Promotes Waiting but Is Not Reinforcing*, „Current Biology” 25, nr 3: 306-315,
- Frank M. J., Seeberger L. C., O'Reilly R. C., 2004, *By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism*, „Science” 306, nr 5703: 1940-1943,
- Gartside S. E., Cole A. J., Williams A. P., McQuade R., Judge S. J., 2007, *AMPA and NMDA receptor regulation of firing activity in 5-HT neurons of the dorsal and median raphe nuclei*, „European Journal of Neuroscience” 25, nr 10: 3001-3008,
- Gellynck E., Heynink K., Andressen K. W., Haegeman G., Levy F. O., Vanhoenacker P., Van Craenenbroeck K., 2013, *The serotonin 5-HT₇ receptors: two decades of research*, „Experimental Brain Research” 230, nr 4: 555-568,
- Gire D. H., Kapoor V., Arrighi-Allisan A., Seminara A., Murthy V. N., 2016, *Mice Develop Efficient Strategies for Foraging and Navigation Using Complex Natural Stimuli*, „Current Biology” 26, nr 10: 1261-1273,
- Gläscher J., Hampton A. N., O'Doherty J. P., 2009, *Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making*, „Cerebral Cortex” 19, nr 2: 483-495,
- Glimcher P. W., 2011, *Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis*, „Proceedings of the National Academy of Sciences” 108, nr supplement_3: 15647-15654,
- Gouveia K., Hurst J. L., 2019, *Improving the practicality of using non-aversive handling methods to reduce background stress and anxiety in laboratory mice*, „Scientific Reports” 9, nr 1: 20305,
- Grant B. F., Goldstein R. B., Saha T. D., Chou S. P., Jung J., Zhang H., ... Hasin D. S., 2015, *Epidemiology of DSM-5 Alcohol Use Disorder*, „JAMA Psychiatry” 72, nr 8: 757,
- Grogan J. P., Tsivos D., Smith L., Knight B. E., Bogacz R., Whone A., Coulthard E. J., 2017, *Effects of dopamine on reinforcement learning and consolidation in Parkinson's disease*, „eLife” 6,
- Grossman C. D., Bari B. A., Cohen J. Y., 2022, *Serotonin neurons modulate learning rate through uncertainty*, „Current Biology” 32, nr 3: 586-599.e7,
- Guy R. C., 2005, *Saccharin*, *Encyclopedia of Toxicology*, 750-752,
- Hampton A. N., Adolphs R., Tyszka J. M., O'Doherty J. P., 2007, *Contributions of the Amygdala to Reward Expectancy and Choice Signals in Human Prefrontal Cortex*, „Neuron” 55, nr 4: 545-555,
- Harada T., 2020, *Learning From Success or Failure? – Positivity Biases Revisited*, „Frontiers in Psychology” 11, nr July: 1-11,
- Hedlund P. B., Danielson P. E., Thomas E. A., Slanina K., Carson M. J., Sutcliffe J. G., 2003, *No hypothermic response to serotonin in 5-HT₇ receptor knockout mice*, „Proceedings of the National Academy of Sciences” 100, nr 3: 1375-1380,
- Hedlund P. B., Huitron-Resendiz S., Henriksen S. J., Sutcliffe J. G., 2005, *5-HT₇ Receptor Inhibition and Inactivation Induce Antidepressantlike Behavior and Sleep Pattern*, „Biological Psychiatry” 58, nr 10: 831-837,

- Hedlund P. B., Sutcliffe J. G., 2007, *The 5-HT₇ receptor influences stereotypic behavior in a model of obsessive-compulsive disorder*, „Neuroscience Letters” 414, nr 3: 247-251,
- Heilig M., Epstein D. H., Nader M. A., Shaham Y., 2016, *Time to connect: bringing social context into addiction neuroscience*, „Nature Reviews Neuroscience” 17, nr 9: 592-599,
- Heilig M., Spanagel R., 2021, *Neurobiology of Alcohol*, The American Psychiatric Association Publishing Textbook of Substance Use Disorder Treatment,
- Hollerman J. R., Schultz W., 1998, *Dopamine neurons report an error in the temporal prediction of reward during learning*, „Nature Neuroscience” 1, nr 4: 304-309,
- Hoyer D., 2019, *Serotonin receptors nomenclature*, *The Serotonin System*, 63-93,
- Hyde J. S., 2016, *Sex and cognition: gender and cognitive functions*, „Current Opinion in Neurobiology” 38: 53-56,
- Ineichen C., Sigrist H., Spinelli S., Lesch K.-P., Sautter E., Seifritz E., Pryce C. R., 2012, *Establishing a probabilistic reversal learning test in mice: Evidence for the processes mediating reward-stay and punishment-shift behaviour and for their modulation by serotonin*, „Neuropharmacology” 63, nr 6: 1012-1021,
- Izquierdo A., Brigman J. L., Radke A. K., Rudebeck P. H., Holmes A., 2017, *The neural basis of reversal learning: An updated perspective*, „Neuroscience” 345: 12-26,
- Jabłońska J., Szumiec Ł., Zieliński P., Rodriguez Parkitna J., 2021, *Time elapsed between choices in a probabilistic task correlates with repeating the same decision*, „European Journal of Neuroscience” 53, nr 8: 2639-2654,
- Jastrzebska K., Walczak M., Cieslak P. E., Szumiec Ł., Turbasa M., Engblom D., ... Parkitna J. R., 2016, *Loss of NMDA receptors in dopamine neurons leads to the development of affective disorder-like symptoms in mice*, „Scientific Reports” 6, nr July: 1-11,
- Kabra M., Robie A. A., Rivera-Alba M., Branson S., Branson K., 2013, *JAABA: interactive machine learning for automatic annotation of animal behavior*, „Nature Methods” 10, nr 1: 64-67,
- Karamched B., Stickler M., Ott W., Lindner B., Kilpatrick Z. P., Josić K., 2020, *Heterogeneity Improves Speed and Accuracy in Social Networks*, „Physical Review Letters” 125, nr 21: 218302,
- Kassambara A., 2020, *rstatix: Pipe-Friendly Framework for Basic Statistical Tests.*, .
- Kehagia A. A., Murray G. K., Robbins T. W., 2010, *Learning and cognitive flexibility: Frontostriatal function and monoaminergic modulation*, „Current Opinion in Neurobiology” 20, nr 2: 199-204,
- Kelley A. E., Berridge K. C., 2002, *The neuroscience of natural rewards: relevance to addictive drugs.*, „The Journal of neuroscience : the official journal of the Society for Neuroscience” 22, nr 9: 3306-3311.
- Kiryk A., Janusz A., Zglinicki B., Turkes E., Knapska E., Konopka W., ... Kaczmarek L., 2020, *IntelliCage as a tool for measuring mouse behavior – 20 years perspective*, „Behavioural Brain Research” 388, nr March: 112620,
- Kiryk A., Mochol G., K. Filipkowski R., Wawrzyniak M., Lioudyno V., Knapska E., ... Kaczmarek L., 2011, *Cognitive Abilities of Alzheimers Disease Transgenic Mice are Modulated by Social Context and Circadian Rhythm*, „Current Alzheimer Research” 8, nr 8: 883-892,
- Konorski J., 1949, *I. P. Pavlov*, „BMJ” 2, nr 4634: 944-950,
- Koskela M., Piepponen T. P., Andressoo J., Vöikar V., Airavaara M., 2018, *Towards developing a model to study alcohol drinking and craving in female mice housed in automated cages*, „Behavioural Brain Research” 352, nr August 2017: 116-124,

- Le Moal M., Simon H., 1991, *Mesocorticolimbic dopaminergic network: functional and regulatory roles*, „Physiological Reviews” 71, nr 1: 155-234,
- Leeman R. F., Heilig M., Cunningham C. L., Stephens D. N., Duka T., O'Malley S. S., 2010, *Ethanol consumption: How should we measure it? Achieving consilience between human and animal phenotypes*, „Addiction Biology” 15, nr 2: 109-124.
- Lin M., Fan B., Lui J. C. S., Chiu D.-M., 2007, *Stochastic analysis of file-swarming systems*, „Performance Evaluation” 64, nr 9-12: 856-875,
- Lindström B., Bellander M., Amodio D. M., Schultner D. T., Chang A., Tobler P. N., 2021, *A computational reward learning account of social media engagement*, „Nature Communications”: 1-10,
- Liu Z., Lin R., Luo M., 2020, *Reward Contributions to Serotonergic Functions*, „Annual Review of Neuroscience” 43: 141-162.
- Long A., Platt M., 2005, *Decision Making: The Virtue of Patience in Primates*, „Current Biology” 15, nr 21: R874-R876.
- Lottem E., Banerjee D., Vertechi P., Sarra D., Lohuis M. oude, Mainen Z. F., 2018, *Activation of serotonin neurons promotes active persistence in a probabilistic foraging task*, „Nature Communications” 9, nr 1: 1000,
- Loureiro M., Achargui R., Flakowski J., Van Zessen R., Stefanelli T., Pascoli V., Lüscher C., 2019, *Social transmission of food safety depends on synaptic plasticity in the prefrontal cortex*, „Science” 364, nr 6444: 991-995,
- Lüscher C., Ungless M. A., 2006, *The Mechanistic Classification of Addictive Drugs*, „PLoS Medicine” 3, nr 11: e437,
- MacKillop J., Amlung M. T., Few L. R., Ray L. A., Sweet L. H., Munafò M. R., 2011, *Delayed reward discounting and addictive behavior: a meta-analysis*, „Psychopharmacology” 216, nr 3: 305-321,
- Magnuson B. A., Carakostas M. C., Moore N. H., Poulos S. P., Renwick A. G., 2016, *Biological fate of low-calorie sweeteners*, „Nutrition Reviews” 74, nr 11: 670-689,
- Maia T. V., Frank M. J., 2011, *From reinforcement learning models to psychiatric and neurological disorders*, „Nature Neuroscience” 14, nr 2: 154-162,
- Makowiecki K., Hammond G., Rodger J., 2012, *Different Levels of Food Restriction Reveal Genotype-Specific Differences in Learning a Visual Discrimination Task*, „PLoS ONE” 7, nr 11: e48703,
- Mathis A., Mamidanna P., Cury K. M., Abe T., Murthy V. N., Mathis M. W., Bethge M., 2018, *DeepLabCut: markerless pose estimation of user-defined body parts with deep learning*, „Nature Neuroscience” 21, nr 9: 1281-1289,
- Matias S., Lottem E., Dugué G. P., Mainen Z. F., 2017, *Activity patterns of serotonin neurons underlying cognitive flexibility*, „eLife” 6: 1-24,
- Matsumoto I., 2013, *Gustatory Neural Pathways Revealed by Genetic Tracing from Taste Receptor Cells*, „Bioscience, Biotechnology, and Biochemistry” 77, nr 7: 1359-1362,
- Meder D., Herz D. M., Rowe J. B., Lehericy S., Siebner H. R., 2019, *The role of dopamine in the brain - lessons learned from Parkinson's disease*, „NeuroImage” 190, nr May 2018: 79-93,
- Mitchell J. M., Fields H. L., D'Esposito M., Boettiger C. A., 2005, *Impulsive responding in alcoholics*, „Alcoholism: Clinical and Experimental Research” 29, nr 12: 2158-2169.
- Miyazaki K., Miyazaki K. W., Yamanaka A., Tokuda T., Tanaka K. F., Doya K., 2018, *Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience*, „Nature Communications” 9, nr 1: 2048,

- Mohebi A., Pettibone J. R., Hamid A. A., Wong J.-M. T., Vinson L. T., Patriarchi T., ... Berke J. D., 2019, *Dissociable dopamine dynamics for learning and motivation*, „Nature” 570, nr 7759: 65-70,
- Morales I., Berridge K. C., 2020, *‘Liking’ and ‘wanting’ in eating and food reward: Brain mechanisms and clinical implications*, „Physiology & Behavior” 227, nr January: 113152,
- Murre J. M. J., Dros J., 2015, *Replication and Analysis of Ebbinghaus’ Forgetting Curve*, „PLOS ONE” 10, nr 7: e0120644,
- Neill J. C., Barnes S., Cook S., Grayson B., Idris N. F., McLean S. L., ... Harte M. K., 2010, *Animal models of cognitive dysfunction and negative symptoms of schizophrenia: Focus on NMDA receptor antagonism*, „Pharmacology & Therapeutics” 128, nr 3: 419-432,
- Niewoehner B., Single F. N., Hvalby Ø., Jensen V., Meyer zum Alten Borgloh S., Seeburg P. H., ... Bannerman D. M., 2007, *Impaired spatial working memory but spared spatial reference memory following functional loss of NMDA receptors in the dentate gyrus*, „European Journal of Neuroscience” 25, nr 3: 837-846,
- Nikiforuk A., 2012, *Selective blockade of 5-HT₇ receptors facilitates attentional set-shifting in stressed and control rats*, „Behavioural Brain Research” 226, nr 1: 118-123,
- Nikiforuk A., 2015, *Targeting the Serotonin 5-HT₇ Receptor in the Search for Treatments for CNS Disorders: Rationale and Progress to Date*, „CNS Drugs” 29, nr 4: 265-275,
- Nikiforuk A., 2016, *Od laboratorium do kliniki: ocena funkcji poznawczych w modelach zwierzęcych*, „Wszechświat” 117, nr 01-03: 32-38.
- Nikiforuk A., Kos T., Fijał K., Hołuj M., Rafa D., Popik P., 2013, *Effects of the Selective 5-HT₇ Receptor Antagonist SB-269970 and Amisulpride on Ketamine-Induced Schizophrenia-like Deficits in Rats*, „PLOS ONE” 8, nr 6: e66695,
- Nilsson S. R. O., Phillips B. U., Axelsson S. F. A., Alsiö J., 2019, *Serotonin and cognitive flexibility, The Serotonin System*, 133-154,
- Niv Y., Duff M. O., Dayan P., 2005, *Dopamine, uncertainty and TD learning*, „Behavioral and Brain Functions” 1, nr 1: 6,
- Noworyta-Sokolowska K., Kozub A., Jablonska J., Rodriguez Parkitna J., Drozd R., Rygula R., 2019, *Sensitivity to negative and positive feedback as a stable and enduring behavioural trait in rats*, „Psychopharmacology” 236, nr 8: 2389-2403,
- Palminteri S., Wyart V., Koechlin E., 2017, *The Importance of Falsification in Computational Cognitive Modeling*, „Trends in Cognitive Sciences” 21, nr 6: 425-433,
- Papp M., Moryl E., Willner P., 1996, *Pharmacological validation of the chronic mild stress model of depression*, „European Journal of Pharmacology” 296, nr 2: 129-136,
- Pearce J. M., Hall G., 1980, *A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli.*, „Psychological Review” 87, nr 6: 532-552,
- Pedersen M. L., Frank M. J., Biele G., 2017, *The drift diffusion model as the choice rule in reinforcement learning*, „Psychonomic Bulletin & Review” 24, nr 4: 1234-1251,
- Phillips B. U., Robbins T. W., 2020, *The role of central serotonin in impulsivity, compulsivity, and decision-making: comparative studies in experimental animals and humans*, t. 31, *Handbook of Behavioral Neuroscience*,
- Pollak Dorocic I., Fürth D., Xuan Y., Johansson Y., Pozzi L., Silberberg G., ... Meletis K., 2014, *A Whole-Brain Atlas of Inputs to Serotonergic Neurons of the Dorsal and Median Raphe Nuclei*, „Neuron” 83, nr 3: 663-678.
- Polskie Towarzystwo Diabetologiczne, 2020, *Zalecenia kliniczne dotyczące postępowania u chorych na cukrzycę 2020.*, t. 6, *Diabetologia Praktyczna*, .

- R Core Team, 2020, *R: A language and environment for statistical computing.*,
- Radwanska K., Kaczmarek L., 2012, *Characterization of an alcohol addiction-prone phenotype in mice*, „Addiction Biology” 17, nr 3: 601-612,
- Rajagopal L., Massey B. W., Michael E., Meltzer H. Y., 2016, *Serotonin (5-HT)_{1A} receptor agonism and 5-HT₇ receptor antagonism ameliorate the subchronic phencyclidine-induced deficit in executive functioning in mice*, „Psychopharmacology” 233, nr 4: 649-660,
- Reuber M. D., 1978, *Carcinogenicity of saccharin.*, „Environmental Health Perspectives” 25, nr August: 173-200,
- Rhodes J. S., Ford M. M., Yu C.-H., Brown L. L., Finn D. A., Garland T., Crabbe J. C., 2007, *Mouse inbred strain differences in ethanol drinking to intoxication*, „Genes, Brain and Behavior” 6, nr 1: 1-18,
- Robbins T. W., Cardinal R. N., 2019, *Computational psychopharmacology: a translational and pragmatic approach*, „Psychopharmacology” 236, nr 8: 2295-2305,
- Roberts A. J., Hedlund P. B., 2012, *The 5-HT₇ receptor in learning and memory*, „Hippocampus” 22, nr 4: 762-771,
- Roberts A. J., Krucker T., Levy C. L., Slanina K. A., Sutcliffe J. G., Hedlund P. B., 2004, *Mice lacking 5-HT₇ receptors show specific impairments in contextual learning*, „European Journal of Neuroscience” 19, nr 7: 1913-1922,
- Rossi M. A., Stuber G. D., 2018, *Overlapping Brain Circuits for Homeostatic and Hedonic Feeding*, „Cell Metabolism” 27, nr 1: 42-56,
- Rutledge R. B., Lazzaro S. C., Lau B., Myers C. E., Gluck M. A., Glimcher P. W., 2009, *Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson's Patients in a Dynamic Foraging Task*, „Journal of Neuroscience” 29, nr 48: 15104-15114,
- Schultz W., 2017, *Electrophysiological Correlates of Reward Processing in Dopamine Neurons*, *Decision Neuroscience*, 21-31,
- Schultz W., Dayan P., Montague P. R., 1997, *A neural substrate of prediction and reward.*, „Science” 275, nr 5306: 1593-1599.
- Schultz Wolfram, 2015, *Neuronal reward and decision signals: From theories to data*, „Physiological Reviews” 95, nr 3: 853-951.
- Simon S. A., De Araujo I. E., Gutierrez R., Nicolelis M. A. L., 2006, *The neural mechanisms of gustation: A distributed processing code*, „Nature Reviews Neuroscience” 7, nr 11: 890-901.
- Skinner B. F., 1950, *Are theories of learning necessary?*, „Psychological Review” 57, nr 4: 193-216.
- Smutek M., Turbasa M., Sikora M., Piechota M., Zajdel J., Przewlocki R., Parkitna J. R., 2014, *A Model of Alcohol Drinking under an Intermittent Access Schedule Using Group-Housed Mice*, „PLoS ONE” 9, nr 5: e96787,
- Stahl S. M., 2013, *Stahl's Essential Psychopharmacology: Neuroscientific Basis and Practical Application*, Fourth New York,
- Stephens D. N., Duka T., 2008, *Cognitive and emotional consequences of binge drinking: Role of amygdala and prefrontal cortex*, „Philosophical Transactions of the Royal Society B: Biological Sciences” 363, nr 1507: 3169-3179.
- Story G. W., Vlaev I., Seymour B., Darzi A., Dolan R. J., 2014, *Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective*, „Frontiers in Behavioral Neuroscience” 8, nr MAR: 1-20,
- Sutton R. S., Barto A. G., 1998, *Reinforcement Learning: An Introduction*, „IEEE Transactions on Neural Networks” 9, nr 5: 1054-1054,

- Sweatman T. W., Renwick A. G., 1979, *Saccharin Metabolism and Tumorigenicity*, „Science” 205, nr 4410: 1019-1020,
- Światowa Organizacja Zdrowia W., 2009, *Międzynarodowa Statystyczna Klasyfikacja Chorób i Problemów Zdrowotnych ICD-10*, ,
- Tao R., Auerbach S. B., 2000, *Regulation of serotonin release by GABA and excitatory amino acids*, „Journal of Psychopharmacology” 14, nr 2: 100-113,
- Terashima T., Shirakawa K., Maekawa M., Furukawa N., Yamaguchi S., Hori Y., 2012, *Differential expression of NMDA receptors in serotonergic and/or GABAergic neurons in the midbrain periaqueductal gray of the mouse*, „Neuroscience Letters” 528, nr 1: 55-60,
- Thorndike E. L., 1898, *Some Experiments on Animal Intelligence*, „Science” 7, nr 181: 818-824,
- Thorndike E. L., 1927, *The Law of Effect*, „The American Journal of Psychology” 39, nr 1/4: 212,
- Tolman E. C., 1948, *Cognitive maps in rats and men.*, „Psychological Review” 55, nr 4: 189-208,
- Volkow N. D., Wise R. A., 2005, *How can drug addiction help us understand obesity?*, „Nature Neuroscience” 8, nr 5: 555-560,
- Waltz J. A., 2017, *The neural underpinnings of cognitive flexibility and their disruption in psychotic illness*, „Neuroscience” 345: 203-217,
- Watkins C. J. C. H., Dayan P., 1992, *Technical Note: Q-Learning*, „Machine Learning” 8, nr 3: 279-292.
- Weber T., 2009, *Inducible gene manipulations in serotonergic neurons*, „Frontiers in Molecular Neuroscience” 2, nr 2: 1-8,
- Weber T., Schönig K., Tews B., Bartsch D., 2011, *Inducible Gene Manipulations in Brain Serotonergic Neurons of Transgenic Rats*, „PLoS ONE” 6, nr 11: e28283,
- Weiner J. L., Valenzuela C. F., 2006, *Ethanol modulation of GABAergic transmission: The view from the slice*, „Pharmacology & Therapeutics” 111, nr 3: 533-554,
- Wilson R. C., Collins A. G. E., 2019, *Ten simple rules for the computational modeling of behavioral data*, „eLife” 8, nr 6: 1-35,
- Wimmer G. E., Li J. K., Gorgolewski K. J., Poldrack R. A., 2018, *Reward Learning over Weeks Versus Minutes Increases the Neural Representation of Value in the Human Brain*, „The Journal of Neuroscience” 38, nr 35: 7649-7666,
- Winiarski M., Borowska J., Wołyniak R. M., Jędrzejewska-Szmek J., Kondrakiewicz L., Mankiewicz L., ... Knapska E., 2021, *Social learning about rewards – how information from others helps to adapt to changing environment*, „bioRxiv”: 2021.03.09.434563,
- Wise R. A., 2002, *Brain Reward Circuitry*, „Neuron” 36, nr 2: 229-240,
- Wise R. A., Bozarth M. A., 1987, *A psychomotor stimulant theory of addiction.*, „Psychological Review” 94, nr 4: 469-492,
- Wise R. A., McDevitt R. A., 2018, *Drive and Reinforcement Circuitry in the Brain: Origins, Neurotransmitters, and Projection Fields*, „Neuropsychopharmacology” 43, nr 4: 680-689,
- Wozniak P. A., Gorzelanczyk E. J., Murakowski J. A., 1995, *Two components of long-term memory*, „Acta Neurobiologiae Experimentalis” 55, nr 4: 301-305.
- Yoneyama N., Crabbe J. C., Ford M. M., Murillo A., Finn D. A., 2008, *Voluntary ethanol consumption in 22 inbred mouse strains*, „Alcohol” 42, nr 3: 149-160,
- Zimbardo P. G., Johnson R. L., McCann V., 2017, *Psychologia: Kluczowe koncepcje. Motywacja i uczenie się. Tom 2*, Warszawa.
- Zito K., Scheuss V., 2009, *NMDA Receptor Function and Physiological Modulation*, *Encyclopedia of Neuroscience*, 1157-1164,

9. Spis rycin

Rycina 2-1 Interakcja między uczniem i środowiskiem w uczeniu ze wzmocnieniem.....	25
Rycina 2-2 Pseudokod prostego algorytmu do rozwiązania problemu n-rękiego bandyty	26
Rycina 2-3 Prawdopodobieństwo wyboru danej opcji w zależności od różnicy w wartości oczekiwanej nagrody między dwoma opcjami i wartości współczynnika β	27
Rycina 2-4 Schemat blokowy algorytmu modelu z podwójnym tempem uczenia przy dwóch możliwych opcjach działania, zastosowany w niniejszych badaniach.	29
Rycina 2-5 Tempo rozpadu zależnie od wartości współczynnika S	32
Rycina 4-1 Wizualizacja wnętrza aparatu IntelliCage, wersja uproszczona.	37
Rycina 4-2 Aparat IntelliCage.....	38
Rycina 4-3 Wnętrze klatki do warunkowania instrumentalnego.....	41
Rycina 5-1 Właściwości nagradzające substancji.....	48
Rycina 5-2 Liczba prób uzyskania nagrody w etapie odwróceń..	49
Rycina 5-3 Preferencja opcji z większym prawdopodobieństwem nagrody.....	50
Rycina 5-4 Znormalizowana częstość interwału do kolejnego działania.	52
Rycina 5-5 Współczynniki regresji logistycznej.....	53
Rycina 5-6 Prawdopodobieństwo działania „zostań” na podstawie regresji logistycznej.....	54
Rycina 5-7 Ułamek poprawnych przewidywań modeli zastosowanych w badaniach.....	56
Rycina 5-8 Prawdopodobieństwo warunkowe modelu na podstawie wagi Akaike.....	57
Rycina 5-9 Ranking modeli na podstawie odsetka grupy z największym dopasowaniem.....	58
Rycina 5-10 Preferencja substancji i liczba prób uzyskania nagrody w czasie doświadczenia....	65
Rycina 5-11 Preferencja opcji o wyższym prawdopodobieństwie nagrody.....	66
Rycina 5-12 Efekt mutacji na preferencję sacharyny..	67
Rycina 5-13 Liczba prób uzyskania nagrody w etapie odwróceń.....	68
Rycina 5-14 Preferencja opcji o wyższym prawdopodobieństwie nagrody.	70
Rycina 5-15 Rozkład czasu mijającego do kolejnej próby uzyskania nagrody w 2-minutowych przedziałach.	71
Rycina 5-16 Współczynniki regresji logistycznej.	72
Rycina 5-17 Ułamek poprawnych przewidywań modeli zastosowanych w badaniach.	74
Rycina 5-18 Prawdopodobieństwo warunkowe modelu na podstawie wagi Akaike.....	75
Rycina 5-19 Ranking modeli na podstawie odsetka grupy z największym dopasowaniem.....	76
Rycina 5-20 Preferencja sacharyny.....	82
Rycina 5-21 Preferencja opcji o wyższym prawdopodobieństwie uzyskania nagrody.	83
Rycina 5-22 Działanie w sytuacji zróżnicowanego prawdopodobieństwa nagrody.	84
Rycina 5-23 Opóźnienie między odpowiedziami instrumentalnymi.....	85
Rycina 5-24 Dopasowanie modelu do danych.	86
Rycina 5-25 Optymalne wartości parametrów α modelu „podwójnie mieszanego”.....	87

10. Spis tabel

Tabela 2—1 Sposób aktualizacji wartości oczekiwanej nagrody w modelach zastosowanych w niniejszych badaniach.....	28
Tabela 4—1 Zestawienie doświadczeń ze schematem odwróceń włączającym fazy pośrednie, o równym prawdopodobieństwie uzyskania nagrody.....	38
Tabela 4—2 Zestawienie doświadczeń ze skróconym schematem, bez faz pośrednich o równym prawdopodobieństwie uzyskania nagrody.	40
Tabela 4—3 Fazy testu z probabilistycznym odwróceniem w klatce Skinnera.....	42
Tabela 4—4 Zestawienie doświadczeń z zastosowaniem klatki do warunkowania instrumentalnego oraz pochodzących z innych testów.....	43
Tabela 4—5 Wartości parametrów modeli.....	46
Tabela 5—1 Optymalne wartości parametrów modeli.....	58
Tabela 5—2 Różnice w parametrach α w modelu "podwójnie mieszanym".....	60
Tabela 5—3 Statystyka dla wartości parametrów modeli.....	62
Tabela 5—4 Wartości optymalne parametrów modeli.....	77
Tabela 5—5 Różnice w parametrach α w modelu "podwójnie mieszanym".....	79
Tabela 5—6 Znamienność różnic wartości parametrów.....	80
Tabela 5—7 Wartości optymalne parametrów modeli.....	87