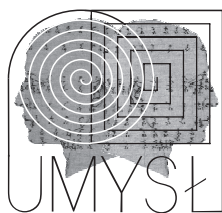


UMYSŁ  
KONFABULUJĄCY

## KOMITET REDAKCYJNY SERII

Andrzej Klawiter  
Zbysław Muszyński  
Robert Piłat  
Robert Poczobut  
Urszula Żegleń



---

PRACE Z FILOZOFII I KOGNITYWISTYKI

Magdalena Reuter

# UMYSŁ KONFABULUJĄCY

Analiza kognitywistyczna



Poznań 2014

RECENZENCI

prof. zw. dr hab. Renata Ziemińska  
prof. zw. dr hab. Paweł Zeidler

Wydanie I

PROJEKT OKŁADKI

Marta i Zdzisław Kwiatkowsy

REDAKCJA

Adriana Staniszevska

KOREKTA

Adriana Staniszevska  
Michał Staniszevski

© Copyright by Magdalena Reuter 2014

Publikacja finansowana przez Instytut Psychologii,  
dziekana Wydziału Nauk Społecznych  
i rektora Uniwersytetu im. Adama Mickiewicza w Poznaniu

**ISBN 978-83-62243-73-0**

WYDAWNICTWO NAUKOWE WYDZIAŁU NAUK SPOŁECZNYCH  
UNIwersytetu im. Adama Mickiewicza w Poznaniu  
60-569 Poznań, ul. Szamarzewskiego 89c

DRUK

Zakład Graficzny UAM  
61-712 Poznań, ul. Wieniawskiego 1

# Spis treści

Wstęp .....	7
-------------	---

## CZĘŚĆ I: ZAGADNIENIE KONFABULACJI – ANALIZA POJĘCIOWA

Rozdział I. Filozoficzny status błędnych przekonań .....	13
1. Zagadnienie przekonań w epistemologii, filozofii umysłu i kognitywistyce	13
2. Pojęcie błędnych przekonań według Alvina Goldmana .....	14
3. Klasyfikacja błędnych przekonań według Ryana McKaya i Daniela C. Dennetta .....	19
Rozdział II. Konfabulacja w ujęciu Williama Hirsteina .....	23
1. Konfabulacje normalne a patologiczne .....	23
1.1. Pamięciowe ujęcie konfabulacji. ....	25
1.1.1. Krytyka pamięciowego ujęcia Hirsteina .....	28
1.2. Epistemiczne ujęcie konfabulacji .....	32
1.2.1. Dwie fazy konfabulowania według Hirsteina .....	35
1.3. Kategoryzacje konfabulacji klinicznych oraz normalnych .....	39
Rozdział III. Konfabulacja: pojęcie, odmiany, sposoby wyjaśniania .....	43
1. Trudności definicyjne związane z terminem „konfabulacja” .....	43
2. Rodzaje konfabulacji. Zarys dotychczasowych podziałów .....	44
3. Podziały konfabulacji naturalnych .....	46
3.1. Podział konfabulacji ze względu na rodzaj procesu poznawczego ....	46
3.2. Podział konfabulacji ze względu na lokalizację czasową .....	47
3.3. Podział konfabulacji ze względu na treść .....	49
3.4. Podział konfabulacji ze względu na formę .....	50
3.5. Podział konfabulacji ze względu na typ procesu konfabulowania ....	50
4. Mechanizmy konfabulowania i procesy konfabulacyjne .....	52
4.1. Mechanizmy konfabulowania .....	52
4.2. Charakterystyka procesu konfabulowania .....	54
4.2.1. Modele procesów konfabulowania .....	54

5. Konfabulacja a pokrewne fenomeny epistemiczne, czyli co jest, a co nie jest konfabulacją . . . . .	55
5.1. Mówienie prawdy – konfabulacja a kłamstwo i fikcyjność . . . . .	55
5.2. Pewność przypisywana przekonaniom. Konfabulacja a samooszustwo – podobieństwa i różnice . . . . .	58

## CZĘŚĆ II: ZAGADNIENIE KONFABULACJI – ZASTOSOWANIA

Rozdział IV. Konfabulacje pamięciowe – badanie fałszywych wspomnień . . . . .	69
1. Mechanizmy powstawania fałszywych wspomnień . . . . .	69
2. Jak odróżnić fałszywe wspomnienia od prawdziwych? . . . . .	76
3. Podziały błędów pamięci a trzy tryby konfabulowania . . . . .	78
Rozdział V. Konfabulacje socjocentryczne a teoria Mind-Reading . . . . .	81
1. Klasyczna teoria Mind-Reading . . . . .	81
2. Teoria Mind-Reading a konfabulacje . . . . .	83
3. Zespoły błędnej identyfikacji a teoria Mind-Reading według Williama Hirsteina . . . . .	84
Rozdział VI. Podmiotowe i świadomościowe determinanty konfabulacji . . . . .	91
1. Konfabulacje a podmiotowość . . . . .	91
1.1. Koncepcja pięciu rodzajów „ja” według Ulrica Neissera . . . . .	92
1.1.1. Koncepcja Neissera a konfabulacje . . . . .	99
1.2. Jaźń jako środek narratywnej ciężkości według Daniela C. Dennetta . . . . .	104
2. Świadomość podmiotu konfabulującego . . . . .	108
2.1. Heterofenomenologia Dennetta w kontekście zagadnienia konfabulacji . . . . .	108
2.1.1. Orwellowski i stalinowski model świadomości a teatr kartezyjański . . . . .	111
2.2. Wykonawcze teorie konfabulacji pamięciowych a paradoks homunkulusa . . . . .	115
2.2.1. Świadomość czasu, świadomość wiedzy i ich znaczenie dla zagadnienia konfabulacji pamięciowych . . . . .	118
2.3. Paradoks homunkulusa a epistemiczna teoria konfabulacji Williama Hirsteina . . . . .	122
Rozdział VII. Konfabulacja jako egzaptacja . . . . .	125
1. Ewolucja błędnych przekonań a problem wartości przystosowawczej wiarygodnej wiedzy . . . . .	125
1.1. Krytyka radykalnego adaptacjonizmu Stephena J. Goulda . . . . .	127
1.2. Samooszustwo życzeniowe jako pendentyw . . . . .	130
1.3. Iluzje poznawcze i konfabulacje – adaptacje, pendentywy czy egzaptacje? . . . . .	131
Zakończenie . . . . .	137
Bibliografia . . . . .	141

# Wstęp

Istotnym elementem życia jest to, że ludzie opowiadają historie o sobie samych, innych osobach i świecie, który ich otacza. W narracjach tych często pojawiają się luki w wiedzy, którą jednostka dysponuje. Bywają one uzupełniane informacjami bez pokrycia, zgadywaniem bądź konfabulacjami, czyli błędnie ugruntowanymi przekonaniem. Od wielu lat fenomen konfabulacji stanowi jedną z intrygujących zagadek badawczych. Początkowo traktowano go jako rodzaj ułomności ludzkiej i próbowano wyjaśniać poprzez odwołania do przypadków patologii neuronalnych. Obecnie jest on dość dobrze rozpoznany zjawiskiem, które badają nie tylko patologowie, neurobiologowie, ale również psychologowie, epistemologowie i kognitywiści.

Epistemologia i kognitywistyka okazały się inspiracją do podjęcia zadania badawczego polegającego na ukazaniu bardziej złożonego obrazu konfabulacji niż dotąd proponowany. Refleksja badawcza dotycząca konfabulacji stała się celem książki *Umysł konfabulujący. Analiza kognitywistyczna*. Została ona przeprowadzona na podstawie najnowszych wyników badań z zakresu psychologii, neuronauki oraz filozofii umysłu.

Zadaniu temu służy również struktura książki. Część I zatytułowana „Zagadnienie konfabulacji – analiza pojęciowa” przedstawia badania nad konfabulacją z perspektywy współczesnych badań epistemologicznych oraz wybrane wątki kognitywistyczne. Problematykę tę omawiam w trzech pierwszych rozdziałach, które prezentują kwestie niezbędne do pełnego zrozumienia tytułowego zagadnienia. Rozpaczynam od omówienia problemu błędnych przekonań, które opieram na stanowisku Alvina Goldmana, jako epistemologicznie dogodnym punkcie wyjścia do analizy fenomenu konfabulacji. Jako rozszerzenie tego ujęcia prezentuję klasyfikację błędnych przekonań Daniela C. Dennetta i Ryana McKaya (rozdział I). Ważniejsze jednak jest stanowisko Williama Hirsteina, który badał zjawisko konfabulacji, ale traktował je jako szczególny wariant zachowań patologicznych, co – w moim przekonaniu – nie wystarczy

do pełnego zrozumienia tego rodzaju ludzkich zachowań (rozdział II). Hirstein nieprecyzyjnie określił jednak właściwości i kategorie konfabulacji, toteż w toku analizy epistemologicznej nasuwa się spostrzeżenie, że przypadki konfabulacji nie ograniczają się do zachowań patologicznych – konfabulują bowiem także ludzie bez dysfunkcji procesów poznawczych. Poprzez wstępną krytykę stanowiska Hirsteina proponuję zatem poszerzenie pojmowania zjawiska konfabulacji o tzw. konfabulacje normalne. Propozycji tej towarzyszy również próba ich wstępnej kategoryzacji. Umożliwia ona nie tylko dokładniejszą charakterystykę pojmowania konfabulacji, ale daje podstawy do wyróżnienia ich odmian oraz sposobów ich wyjaśniania (rozdział III). Zaproponowana typologia przypadków naturalnego konfabulowania pozwala na wyróżnienie pięciu typów konfabulacji ze względu na: rodzaj procesu poznawczego, lokalizację czasową, przedmiot konfabulacji, formę konfabulacji oraz typ procesu konfabulowania. Omawiam również proces i ściśle z nim powiązane mechanizmy konfabulowania oraz zestawiam konfabulację z takimi fenomenami epistemicznymi, jak kłamstwo, fikcjotwórstwo i samooszustwo.

Część II zatytułowana „Zagadnienie konfabulacji – zastosowania” w pierwszej kolejności prezentuje konfabulacje pamięciowe jako szczególnie istotne w generowaniu treści konfabulacyjnych. Ze względu na przypadki błędów pamięciowych w zdrowych systemach poznawczych można wyróżnić trzy tryby konfabulowania: wypełniający, scalający i addytywny (rozdział IV). Istotnym uzupełnieniem tych treści jest rozdział V, w którym omawiam konfabulacje socjocentryczne, jakie można zestawić z teorią Mind-Reading Barona-Cohena, czyli teorią „czytania w umysłach”. Szczególnie ważne wydaje się jednak podmiotowe i świadomościowe uwarunkowanie zjawiska konfabulacji. Omawiam je w rozdziale VI, gdzie odwołuję się m.in. do poglądów Neissera (koncepcji pięciu rodzajów „ja”) oraz Dennetta (idei jaźni jako środka „narratywnej ciężkości” oraz orwellowskiego i stalinowskiego modelu świadomości). Ponadto przedstawiam zagadnienie statusu argumentu homunkulusa wobec krytyk wykonawczych teorii pamięciowych, które opisuje Gianfranco Dalla Barba. Pracę zamyka kwestia biologiczno-ewolucyjnego statusu fenomenu konfabulacji, którą traktuję jako szczególny przypadek egzaptacji, czyli formy wytworzonej ewolucyjnie w toku rozwoju filogenetycznego gatunku ludzkiego (rozdział VII). Zaprezentowane argumenty potwierdzają, że działanie umysłu ludzkiego – przy udziale czynników kulturowych – zależy od tak złożonych uwarunkowań, że nie sposób objaśnić wszystkich fenomenów funkcjonowania ludzkiego umysłu za pomocą klasycznych pojęć teorii ewolucji, takich jak adaptacja. Wydaje się więc, że konfabulacyjne zdolności naszego umysłu są raczej wstępnym, bo egzaptacyjnym dopasowaniem do działania w środowisku kulturowym.

W zakończeniu książki podsumowuję uzyskane wyniki badań, a w bibliografii przedstawiam najnowsze prace, w których poruszono zagadnienie konfabulacji oraz tematy nawiązujące do rozumienia tego zjawiska. Problematyka ta znajduje się jeszcze na etapie analiz empirycznych, jak wskazują na to choćby badania Hirsteina.

Omówionej w pracy problematyce w polskiej literaturze nie poświęcono dotąd szerzej zakrojonych studiów. Jednak rosnące zapotrzebowanie na humanistyczno-biologiczną refleksję nad tym zjawiskiem sprawia, że w ostatnim czasie intensywnie rozwijane są właśnie badania epistemologiczne i kognitywistyczne. Można nawet zaryzykować stwierdzenie, iż wyprzedzają one w tym zakresie usystematyzowane badania eksperymentalne. Książka niniejsza mieści się w nurcie teoretycznym (epistemologiczno-kognitywistycznym).

Pozostaje mi jeszcze podziękować wszystkim osobom, które przyczyniły się do powstania tej pracy. Dziękuję przede wszystkim promotorowi pracy doktorskiej, na podstawie której powstała ta książka, Profesorowi Krzysztofowi Łastowskiemu, za merytoryczną pomoc, inspirujące dyskusje, a także krytyczne uwagi i wskazówki. Dziękuję też recenzentom pracy doktorskiej Profesor Renacie Ziemińskiej i Profesorowi Pawłowi Zeidlerowi za krytyczne uwagi oraz interesujące spostrzeżenia, które zainspirowały mnie do nowych przemyśleń w trakcie pisania tej książki. Wyrazy podziękowania składam również Profesorowi Andrzejowi Klawiterowi i Doktorowi Piotrowi Przybyszowi za podsuniecie ciekawych pomysłów oraz konstruktywną krytykę prezentowanych przeze mnie zagadnień na seminariach kognitywistycznych. Wyrazy wdzięczności składam także na ręce Koleżanek i Kolegów z Zakładu Epistemologii i Kognitywistyki za inspirujące debaty oraz intelektualną atmosferę, która przyczyniła się do powstania wielu idei tej pracy.



CZĘŚĆ I

# ZAGADNIENIE KONFABULACJI – ANALIZA POJĘCIOWA



## Filozoficzny status błędnych przekonań

### 1. Zagadnienie przekonań w epistemologii, filozofii umysłu i kognitywistyce

Formowanie przekonań jest jedną z podstawowych cech ludzkiego umysłu, dlatego problem przekonań znajduje się w polu zainteresowań zarówno epistemologii, filozofii umysłu, jak i kognitywistyki. Każda z tych dyscyplin zajmuje się jednak nieco innym aspektem tego zagadnienia.

Przekonania to postawy, jakie przyjmujemy, kiedy uważamy coś za prawdziwe lub kiedy uznajemy, że coś za czymś przemawia. Jednak wiele kwestii, co do których jesteśmy przekonani, ma charakter powszedni, codzienny i nie potrzebujemy przypominać sobie danej informacji, aby stwierdzić, że posiadamy jakieś przekonanie. Na przykład nie reflektujemy na co dzień nad tym, że posiadamy głowę czy ręce, że filiżanka, z której pijemy herbatę, jest z porcelany, że blok mieszkalny posiada klatkę schodową, że żyjemy w XXI wieku czy że inne osoby także reprezentują postawy propozycjonalne, w tym przekonania. Niemniej nawet wtedy, gdy nie skupiamy uwagi na wszystkich tych aspektach rzeczywistości w jednym czasie, można powiedzieć, że jesteśmy co do nich przekonani.

Epistemologia, jako dział filozofii zajmujący się naturą, wartościami i granicami ludzkiego poznania, koncentruje się na opisie różnic między przekonaniami a wiedzą oraz kwestią uzasadniania przekonań. Pojęcie wiedzy jest przeciwstawiane pojęciu przekonania. Podczas gdy wiedza jest niezawodna, pewna i wymuszona przez „obiektywną” rzeczywistość, przekonania bywają zawodne, niepewne i subiektywne (Marciszewski 1972). Współczesne dyskusje epistemologiczne odwołują się do standardowej koncepcji wiedzy jako prawdziwego, uzasadnionego przekonania, która wywodzi się od Platona. W koncepcji tej wiedza traktowana jest jako jeden z rodzajów przekonania – takiego, które jest prawdziwe i wystarczająco dobrze uzasadnione. Kontrprzykłady dla tej

definicji podał Edmund Gettier w artykule pt. *Is Justified True Belief Knowledge?* (1963), które stały się przedmiotem szerokiej debaty w epistemologii.

Z kolei współczesne dyskusje toczone w ramach analitycznej filozofii umysłu na temat przekonań są częścią bardziej ogólnej debaty poświęconej postawom propozycjonalnym. Postawa propozycjonalna to stan mentalny posiadania jakiejś postawy, pozycji, opinii na temat twierdzenia lub potencjalnego stanu rzeczy, w którym to twierdzenie jest prawdziwe. Stan mentalny tego rodzaju jest kanonicznie wyrażany w formie „S A, że P”, gdzie S to podmiot posiadający stan mentalny, A – postawa, a P jest zdaniem wyrażającym twierdzenie. Na przykład Jan [podmiot] wątpi [postawa], że poza Układem Słonecznym istnieje życie [zdanie wyrażające twierdzenie]; Ewa [podmiot] ma nadzieję [postawa], że poza Układem Słonecznym istnieje życie [zdanie wyrażające twierdzenie]; Michał [podmiot] jest przekonany [postawa], że poza Układem Słonecznym istnieje życie; Wiktor [podmiot] obawia się, że poza Układem Słonecznym istnieje życie. W co wątpi lub czego obawia się jedna osoba, inna może mieć nadzieję, wierzyć, zamierzać lub być przekonana. A zatem postawy propozycjonalne to inaczej różne stany mentalne, jakie podmioty mogą mieć wobec tego samego twierdzenia. Jedną z nich jest postawa (stan mentalny) posiadania przekonania. Dla analitycznej filozofii umysłu istotne są przede wszystkim zagadnienia koncentrujące się wokół relacji umysł – ciało, w którym najistotniejsza jest odpowiedź na pytanie, czy i w jaki sposób organizm biologiczny może nabywać i podtrzymywać przekonania. Interesująca dla tej dyscypliny badawczej jest więc kwestia statusu przekonań.

Z kolei psychologia i neuropsychologia poznawcza zwracają uwagę na mechanizmy formowania przekonań przez ludzki system poznawczy oraz funkcje, jakie pełni podtrzymywanie różnego rodzaju przekonań przez podmiot. Psychologia zorientowana ewolucyjnie próbuje zaś wskazać na ewolucyjne mechanizmy formowania poprawnych bądź błędnych przekonań.

## 2. Pojęcie błędnych przekonań według Alvina Goldmana

Zagadnienie błędnie ugruntowanych przekonań porusza Alvin Goldman w ramach koncepcji reliabilizmu procesualnego. Filozof ten wyznacza kryterium dla błędnie ugruntowanego przekonania.

Alvin Goldman jest krytycznym kontynuatorem programu epistemologii znaturalizowanej Willarda Van Ormana Quine'a. Jego stanowisko nazywane jest naturalizmem umiarkowanym, sytuującym się między naturalizmem radykalnym a tradycyjną epistemologią (Ziemińska 2002, s. 80). Epistemika Goldmana to program epistemologii interdyscyplinarnej o podłożu naturalistycznym.

W przeciwieństwie do radykalnych naturalistów, takich jak Paul i Patricia Churchlandowie czy Stephen Stich, Goldman nie opowiada się za wyeliminowaniem epistemologii na rzecz psychologii poznawczej, neuropsychologii i innych nauk o poznaniu. Zamiast zniesienia tradycyjnej ontologii przekonań i sądów oraz związanych z nimi prawdziwościowych wartości epistemicznych filozof ten postuluje, by uznać epistemologię za odrębną dyscyplinę, która będzie jednak wykorzystywać dane nauk empirycznych na temat procesów poznawczych.

Epistemikę Goldmana można podzielić na indywidualną i społeczną (Ziemińska 2002, s. 81). W książce *Epistemology and Cognition* (1986) filozof ten zajmuje się epistemologią indywidualną, która powinna jego zdaniem korzystać z wyników badań kognitywistyki. Społeczna epistemologia powinna z kolei korzystać z nauk społecznych i humanistycznych.

Epistemologia indywidualna dzieli się na prymarną i sekundarną. Prymarna epistemologia, zwana inaczej ogólną, dotyczy podstawowej architektury ludzkiego poznania i bada procesy, struktury i mechanizmy poznawcze. Epistemologia sekundarna, zwana też szczegółową, dotyczy metod, algorytmów i heurystyk używanych w poszczególnych dziedzinach wiedzy, takich jak matematyka czy fizyka (Ziemińska 2002, s. 81).

Według Goldmana epistemologia nie jest gałęzią psychologii czy innej nauki o poznaniu. Dlatego też nie jest on eliminatywistą. Zadanie epistemologii stanowi filozoficzna analiza pojęć epistemicznych. W pierwszej kolejności w książce *Epistemology and Cognition* Goldman postuluje, by przyjrzeć się bliżej pojęciom epistemicznym, takim jak wiedza czy uzasadnianie. W drugiej kolejności zastanawia się nad kwestią uzasadniania przekonań w świetle wiedzy z zakresu psychologii poznawczej. Jednak jego zdaniem radykalne odróżnienie psychologii i epistemologii jest niemożliwe. Epistemologia bowiem, oprócz badania analizy pojęć epistemicznych, powinna postawić sobie za cel badanie genezy przekonań. Tego rodzaju badania powinny być przeprowadzane za pomocą psychologii poznawczej i innych nauk wchodzących w skład kognitywistyki, takich jak: psychologia poznawcza, psychologia rozwojowa, teoria sztucznej inteligencji, antropologia poznawcza, neuronauka, lingwistyka i filozofia (Goldman 1993).

Według Goldmana kognitywistyka ma ogromne znaczenie dla epistemologii. Jednym z zadań epistemologii jest bowiem pytanie o źródła ludzkiej wiedzy i racjonalności. Aby na nie odpowiedzieć, należy posiłkować się naukami kognitywnymi. To one odpowiadają na pytanie o procesy i metody prowadzące do wiedzy i racjonalności oraz o poznawcze ograniczenia w tym zakresie. Czysta filozoficzna analiza językowa wystarczy jedynie w przypadku odpowiedzi na podstawowe pytania: czym jest wiedza i racjonalność. Jednak pytania o źródła wiedzy i racjonalności powinny być rozpatrywane w kontekście wyników badań nauk o poznaniu.

Filozof może odnosić się do nauk kognitywnych na trzy sposoby: może wnieść swój wkład do tych nauk, być ich krytykiem bądź korzystać z wyników tych nauk. Wkładem filozofii w kognitywistykę jest dostarczanie narzędzi, takich jak pojęcia semantyki i logiki, które wykorzystywane są w lingwistyce i teorii sztucznej inteligencji – predykcja, kwantyfikacja, referencja, postawy propozycyjalne. Krytyka metodologiczna z kolei obliuguje filozofa do refleksji nad podstawowymi pojęciami i założeniami nauk kognitywnych. Jednym z takich pojęć są postawy propozycyjalne, np. przekonania, obawy czy pragnienia. Filozof zajmujący się krytyką powinien zastanowić się nad statusem postaw propozycyjalnych w kognitywistyce – czy są one właściwymi dla niej pojęciami, czy też należy je odrzucić. Jednak zdaniem Goldmana najważniejszą rolą filozofa w kognitywistyce jest korzystanie z wyników jej badań w celu rozwiązywania własnych problemów, np. dla ustalenia norm epistemicznych filozof może skorzystać z empirycznego opisu procesów poznawczych (Ziemińska 2002, s. 93).

Jako istotne dla epistemologii Goldman uważa np. dane empiryczne z zakresu percepcji przestrzennej głębi oraz problem rozpoznawania przedmiotów na podstawie częściowego ich widzenia w niekorzystnych warunkach. Innym przykładem są wyniki badań z kognitywistyki nad samozłudzeniami. Kwestia ta jest istotna dla rozważań nad zagadnieniem konfabulacji. Samozłudzenia, czy też konfabulacje egocentryczne, to fałszywe przekonania na własny temat, które polegają na tym, że ludzie uważają się pod pewnymi względami za lepszych od przeciętnej. Samozłudzenia wyjaśniane są często przez odwołanie się do pojęcia „schematu »ja«”, czyli mentalnej struktury ludzkiego umysłu kierującej się wyborem i przechowywaniem informacji na temat samego siebie. Percypowanie samego siebie zależy od ukrytego pozytywnego schematu „ja”. Przywoływanie z pamięci zdarzeń i wartościowań odbywa się na podstawie teorii „ja”, która przywołuje pozytywne cechy samego siebie. Takie tendencje do pozytywnej samooceny można uznać za irracjonalne, niemniej z ewolucyjnego punktu widzenia są one korzystne, gdyż motywują ludzi do działań zmierzających do osiągnięcia sukcesów osobistych i zawodowych.

Konsekwencją naturalizmu umiarkowanego Goldmana jest eksternalizm epistemiczny w uzasadnianiu przekonań, czyli stanowisko, zgodnie z którym podmiot nie musi mieć dostępu do czynników uzasadniających jego przekonanie. Innymi słowy, prawdziwe przekonanie może należeć do wiedzy dzięki czynnikom, które nie są introspekcyjnie dostępne podmiotowi, a on sam może nie zdawać sobie sprawy z tego, jakie procesy czy relacje uzasadniają jego przekonanie (Ziemińska 2002).

Eksternalistyczna teoria uzasadnionych przekonań opiera się na stanowisku kauzalistycznym. Uznaje ono, że teoria uzasadnionych przekonań powinna

odwoływać się do procesów, które inicjują i podtrzymują te przekonania. Pod tym względem internalistyczne teorie uzasadniania są błędne, ponieważ pomijają element kauzalny (czyli procesy, które powodują i podtrzymują przekonania).

Goldman proponuje reliabilistyczne pojęcie uzasadniania przekonań odnoszące się do procesów je podtrzymujących. Reliabilizm jest jedną z wersji eksternalizmu z uwagi na to, że własność reliabilności nie jest dostępna introspekcyjnie.

Status uzasadnienia przekonania jest funkcją reliabilności procesu czy procesów, które go wytwarzają, a reliabilność procesu jest tendencją do wytwarzania prawdziwych przekonań (Ziemińska 2002, s. 171).

Procesami reliabilnymi są typowe procesy percepcji, pamięci, poprawnego rozumowania, zaś niereliable są: zgadywanie, poleganie na ocenie emocjonalnej, myślenie życzeniowe, uogólnienie. Do tych ostatnich zaliczam także konfabulacje.

Czym jest proces formowania przekonań? Zdaniem Goldmana jest to procedura, która pewne dane wejściowe przetwarza na dane wyjściowe. Danymi wyjściowymi w tym procesie są przekonania, czyli uznawanie jakichś sądów w jakimś czasie. Reliabilność wyznaczana jest przez własności statystyczne, takie jak wytwarzanie prawdziwych przekonań w 50 czy 90%. Reliabilność jest stopniowalna, podobnie jak uzasadnianie. Proces dochodzenia do przekonań jest typem, ponieważ tylko typy, a nie egzemplarze, posiadają własności statystyczne.

Zdaniem Goldmana proces formowania przekonań jest wiarygodny wówczas, gdy prowadzi do wytworzenia przekonania prawdziwego. Uzasadnienie dla przekonania zapewnia więc reliabilny proces poznawczy. Goldman wyróżnia procesy poznawcze niezależne od przekonań (których danymi wyjściowymi są stany niedoksastyczne) oraz procesy poznawcze zależne od przekonań (których danymi wyjściowymi są inne przekonania). Kryterium uzasadniania, stosownie do tego rozróżnienia, przedstawia się następująco:

- (1) Jeżeli przekonanie S-a, że p w czasie t pochodzi („bezpośrednio”) z niezależnego od przekonań procesu, który jest (bezwzględnie) reliabilny, to S-a przekonanie, że p w t, jest uzasadnione.
- (2) Jeżeli przekonanie S-a, że p w czasie t pochodzi („pośrednio”) z zależnego od przekonań procesu, który jest (przynajmniej) warunkowo reliabilny, i jeżeli przekonania (jeżeli takie są), którymi posługuje się ten proces, wytwarzając przekonanie S-a, że p w t, same są uzasadnione, to S-a przekonanie, że p w t, jest uzasadnione (Goldman 1979).

Dla konfabulacji istotny będzie warunek (1), ponieważ uważa się, iż przynajmniej część z nich powstaje ze stanów niedoksastycznych. Rozwinięciem warunku (1) jest sformułowanie:

(1a) Jeżeli S-a przekonanie, że p w czasie t pochodzi z reliabilnego kognitywnego procesu, i nie ma żadnego innego reliabilnego lub warunkowo reliabilnego procesu dostępnego S-owi, który byłby używany przez S-a dodatkowo z procesem aktualnie używanym i prowadziłyby do przekonania, że nie-p, to S-a przekonanie, że p w t, jest uzasadnione (Goldman 1998).

Przed wszystkim powstaje jednak pytanie, jak ustalić niereliabilność procesów poznawczych. Goldman w swojej późniejszej pracy rozwinął reliabilizm procesualny, nawiązując do epistemologii aretycznej (Ziemińska 2002, s. 201). Postulowany przez filozofa reliabilizm aretyczny odwołuje się do sprawności intelektualnych oraz intelektualnych wad. W tym ujęciu przekonanie jest uzasadnione, gdy pochodzi z intelektualnej sprawności, zaś nieuzasadnione – gdy pochodzi z intelektualnej wady (Ziemińska 2002, s. 201). Intelektualne sprawności posiadają procesy poznawcze, które uznajemy za rzetelne, intelektualne wady zaś – te procesy, które uznawane są za nierzetelne. Lista intelektualnych sprawności i wad ustalana jest przez osoby oceniające epistemiczną rzetelność procesów poznawczych (*epistemic evaluators*) (Ziemińska 2002, s. 202). Zatem o tym, które procesy poznawcze są intelektualnie sprawne, a które nie, decydują eksperci, a nie faktyczna reliabilność.

Zdaniem Williama Hirsteina pacjenci, którzy konfabulują, naruszają kryterium (1) i (1a). W każdym z podanych przez Hirsteina zespołów neuropsychologicznych, w których występują konfabulacje, procesy poznawcze nie są wiarygodne, aby wytworzyć prawdziwe uzasadnione przekonanie. W przypadku zespołu Korsakowa nie można polegać na procesie wydobywania informacji z pamięci; w zespole Antona procesy percepcyjne są niewiarygodne; w przypadku anozognozji przy paraliżu lewostronnym procesy tworzące reprezentację lewej strony ciała nie są rzetelne etc. Zatem warunek (1) nie zostaje w przypadku tych rodzajów konfabulacji spełniony. Jednak w procesie tworzenia się konfabulacji ważne jest nie tylko uszkodzenie systemu poznawczego, ale i uszkodzenie procesów sprawdzających. Dlatego konfabulacje naruszają także warunek (1a).

W przypadku konfabulacji klinicznych przekonania pacjentów są skutkiem nieprawidłowo funkcjonujących procesów formowania przekonań w wyniku uszkodzeń różnych obszarów mózgu. Wśród neuropsychologów i epistemologów nie ma więc rozbieżności co do tego, że przekonania tych pacjentów są błędnie ugruntowane. Powstaje oczywiście wiele pytań dotyczących interpretacji badań oraz danych anatomicznych i fizjologicznych w każdym z zespołów, w których występują konfabulacje kliniczne.

Termin „konfabulacja” stosuje się nie tylko w przypadkach klinicznych, lecz niekiedy także do opisu błędnych przekonań u osób zdrowych. Konfabulacjami nazywa się np. przekonania oparte na błędnym wydobyciu informacji z pamięci. Badania Elizabeth Loftus nad fałszywą treścią pamięci wykazują, że ludzka pamięć jest zawodna i często poddaje się złudzeniom (Loftus 2002). Podobne wnioski płyną z badań Daniela Schactera, który wyróżnił różne rodzaje błędów przekształcenia pamięci: błędną atrybucję źródła, tendencyjność, sugestywność i uporczywość pamięci (Schacter 2003). Konfabulacje spotykane są też często u małych dzieci, które myślą rzeczywistość z wyobraźnią oraz są bardziej podatne na zafałszowania pamięci pod wpływem sugestii (Schacter 2003, s. 204).

Czy w przypadkach konfabulacji u osób zdrowych warunki konfabulowania będą takie same jak te podane przez Hirsteina? Czym różnią się nierzetelne procesy poznawcze wynikające z uszkodzeń mózgu od nierzetelnych procesów u osób zdrowych? Wydaje się, że potrzebna jest nowa epistemiczna charakterystyka, która uwzględniłaby konfabulacje w systemach poznawczych osób zdrowych.

### 3. Klasyfikacja błędnych przekonań według Ryana McKaya i Daniela C. Dennetta

W artykule pt. *The evolution of misbelief* Ryan McKay i Daniel C. Dennett (2009) charakteryzują rodzaje błędnych przekonań z punktu widzenia ewolucji umysłu. Przedstawiona przez nich definicja przekonań jest dość szeroka i obejmuje „stany funkcjonalne organizmu, które implementują lub ucieleśniają aprobatę organizmu dla jakiegoś stanu rzeczy jako aktualnego” (McKay i Dennett 2009, s. 493). Błędne przekonanie to zatem „stan funkcjonalny organizmu aprobujący jakiś stan rzeczy, który nie występuje” (McKay i Dennett 2009, s. 493). Zastanawiając się nad rodzajami błędnych przekonań, za McKayem i Dennettem uznają, że istnieją dwa typy błędnych przekonań: nietolerowalne błędne przekonania, powstające w wyniku dysfunkcji systemu formowania przekonań oraz tolerowalne błędne przekonania, powstające w poprawnie funkcjonujących systemach poznawczych. Ta druga grupa dzieli się na błędne przekonania, które są adaptacjami do środowiska fizycznego bądź społecznego, oraz błędne przekonania będące pendentami (por. rozdział VII), czyli skutkami ubocznymi poprawnie funkcjonującego systemu formowania przekonań, który ma jednak pewne ewolucyjne ograniczenia.

McKay i Dennett uważają, że ewolucja uprzywilejowała nabywanie i podtrzymywanie przekonań prawdziwych, gdyż przybliżają one rzeczywistość, w oparciu o którą działa osoba posiadająca przekonania. Nasze przekonania na temat świata, innych osób i siebie samych pozwalają na efektywne działanie

w środowisku fizycznym i społecznym. Są one narzędziami wytwarzanymi przez system poznawczy człowieka ukształtowany w drodze ewolucji. Nabywanie przekonań prawdziwych i rzetelnie ugruntowanych jest preferowane przez dobór naturalny. Jednak, jeśli założenie to jest zgodne z poglądem przedstawionym wyżej i ewolucja rzeczywiście preferuje posiadanie przekonań rzetelnie ugruntowanych, to powstaje pytanie, co dzieje się w przypadku licznych przykładów posiadania przekonań błędnie ugruntowanych.

Zdaniem McKaya i Dennetta większość z nas posiada niekiedy błędnie ugruntowane przekonania. Z dużym prawdopodobieństwem można stwierdzić, że każdy z nas został kiedyś wprowadzony w błąd lub oszukany. Niekiedy również oszukujemy samych siebie, aby zmniejszyć dysonans poznawczy. W naszych poprawnie funkcjonujących systemach poznawczych pojawiają się też czasem konfabulacje na temat siebie, innych bądź świata. Zasadnicze pytanie, jakie się tu pojawia, dotyczy ewolucyjnych korzeni błędnie ugruntowanych przekonań. Czy są one skutkiem dysfunkcji systemu poznawczego, czy raczej efektem ubocznym sprawnie działającego systemu? To pytanie prowadzi autorów do dokonania taksonomii błędnie ugruntowanych przekonań. Wyróżniają oni dwa ich typy: (1) błędne przekonania będące skutkiem dysfunkcji procesu formowania przekonań; (2) błędne przekonania będące wynikiem sprawnie działającego systemu formowania przekonań.

McKay i Dennett proponują, aby w pierwszej kolejności przyjrzeć się systemom zaprojektowanym przez ludzi (McKay i Dennett 2009, s. 494). Istnieją bowiem dwa typy niesprawności artefaktów stworzonych przez człowieka: (1) zawinione ograniczenia projektu (*culpable design limitations*) i (2) wybacalne ograniczenia projektu (*forgivable design limitations*). Te pierwsze występują wówczas, gdy zaprojektowany artefakt ma uchybienia konstrukcyjne. Przykładem może być drogi zegarek, który spóźnia się każdego dnia o kilka sekund. Tolerowalnym i wybaczalnym ograniczeniem systemu jest natomiast sytuacja, w której zegarek ten spóźnia się o sekundę każdego dnia (McKay i Dennett 2009, s. 494).

Następnie należy przyjrzeć się systemom poznawczym stworzonym przez ewolucję i ich ograniczeniom. Zawinione ograniczenia projektu to dysfunkcje systemu poznawczego, natomiast tolerowalne ograniczenia projektu to takie, które mieszczą się w ramach wytworzonego przez ewolucję systemu. Analogia z powyższym przykładem może być jedynie częściowa, gdyż ewolucja to nie intencjonalny twórca, który wytwarza artefakty w sposób celowy, lecz raczej – używając terminologii Richarda Dawkinsa – „ślepy zegarmistrz” (McKay i Dennett 2009, s. 495).

Pierwsza kategoria to błędy, które stanowią efekt dysfunkcji doksastycznej. Błędne przekonania są tu skutkiem zaburzenia procesów formowania przeoko-

nań. Procesy formowania przekonań zdefiniowane są jako system przetwarzania informacji, który ma na wejściu jakieś dane percepcyjne, a na wyjściu – przekonania. Przykładem takich dysfunkcji są urojenia występujące w różnych zespołach neuropsychologicznych, m.in. w schizofrenii (McKay i Dennett 2009, s. 496).

Urojenia to błędne przekonania *par excellence* – fałszywe przekonania, które są podtrzymywane z silnym przekonaniem bez względu na dowody przeczące oraz próby innych osób odwiedzenia od nich osoby z urojeniami (McKay i Dennett 2009, s. 496).

Urojenia wahają się od bardzo dziwnych i nieprawdopodobnych do nieprawdziwych, ale prawdopodobnych. Do tej kategorii oprócz urojeń zaliczam konfabulacje i samooszustwa patologiczne (por. rozdział II).

Z kolei tolerowalne błędne przekonania to takie, które są wpisane w poprawnie funkcjonujący system poznawczy. Zdaniem McKaya i Dennetta nawet jeśli ewolucja byłaby perfekcyjnym procesem, to i tak istniałyby ograniczenia, które powodowałyby odstępstwa od poprawnie ugruntowanych przekonań. Jeśli jednak jest tak, że ewolucja ma ograniczenia, to czy świadczy to o adaptacyjności owych odstępstw? McKay i Dennett uważają, że to, iż poprawnie funkcjonujące systemy poznawcze raz na jakiś czas wytwarzają błędne przekonania, nie pozostaje w sprzeczności z podejściem ewolucyjnym. Ewolucja bowiem nie wytwarza optymalnie zaprojektowanych systemów (McKay i Dennett 2009, s. 11). Niektóre aspekty ludzkiego umysłu są bowiem skutkami ubocznymi cech adaptacyjnych i można je zakwalifikować jako pendentywy bądź egzaptacje w terminologii Stephena J. Goulda i Richarda C. Lewontina (1979) (por. rozdział VII).

Zaproponowany przez McKaya i Dennetta podział błędnych przekonań na będące wynikiem dysfunkcji procesów formowania przekonań oraz będące wynikiem poprawnie funkcjonujących procesów poznawczych można rozwinąć i do pierwszej grupy błędów zaliczyć urojenia, konfabulacje patologiczne i samooszustwa patologiczne, a do drugiej – konfabulacje, samooszustwa normalne i iluzje poznawcze. Rodzaje błędnych przekonań przedstawia tabela 1.

Tabela 1. Podział błędnych przekonań

Błędne przekonania jako efekt dysfunkcji procesów formowania wiarygodnych przekonań	Błędne przekonania jako efekt poprawnie funkcjonujących procesów poznawczych
– konfabulacje patologiczne	– konfabulacje normalne
– samooszustwa patologiczne	– samooszustwa normalne
– urojenia	– iluzje poznawcze

Źródło: opracowanie własne na podstawie: McKay i Dennett 2009.

Zaproponowany podział obrazuje dwa typy zjawisk: błędne przekonania będące wynikiem uszkodzeń mózgu oraz błędne przekonania w poprawnie funkcjonujących systemach poznawczych. Te pierwsze to zawinione ograniczenia projektu, a drugie to tolerowalne błędne przekonania.

## Konfabulacja w ujęciu Williama Hirsteina

### 1. Konfabulacje normalne a patologiczne

W filozofii i naukach kognitywnych zagadnienie konfabulacji pojawiało się dotychczas sporadycznie. Jedną z pierwszych monografii z tej dyscypliny badawczej jest książka Williama Hirsteina *Brain fiction. Self-deception and the riddle of confabulation* (2005). Praca ta jest próbą całościowego ujęcia zagadnienia konfabulacji z dwóch perspektyw: epistemologii i neuronauki (Reuter 2011). Hirstein skupia się w niej wyłącznie na konfabulacjach klinicznych, analizując różne zespoły neuropsychologiczne. Lista ta zaczyna się od konfabulacji pamięciowych w zespole Korsakowa. Hirstein zauważa, że pojęcie konfabulacji pierwotnie było używane w bardzo wąskim sensie – wyłącznie do określania zaburzeń pamięci w tym zespole oraz rekompensowania luk pamięciowych przez fikcyjne opowieści na temat przeszłości. Pacjent nie zdaje sobie sprawy z tego, że jego opowieści nie są oparte na danych pamięciowych, lecz zmyślane. W neuropsychologii po raz pierwszy pojęcie konfabulacji zostało użyte przez niemieckiego psychiatrę Karla Bonhoeffera w odniesieniu do pacjentów z zespołem Korsakowa (Hirstein 2005, s. 7). Pacjenci ci wypełniają luki w pamięci, tworząc często spójne historie na temat swojej przeszłości.

Wąskie podejście do zjawiska konfabulacji stanowi pewien kłopot zarówno dla neuropsychologii, jak i rozważań epistemologicznych. Czy konfabulować można wyłącznie w obrębie pamięci, czy także w zakresie innych procesów poznawczych, takich jak percepcja wzrokowa? Hirstein uważa, że termin „konfabulacja” należy rozszerzyć na inne zespoły neuropsychologiczne, w których dochodzi do wypełniania luk w percepcji wzrokowej (np. w zespole Antona), konfabulacji na temat sprawności własnego ciała (np. w anozognozji przy hemiplegii), na temat identyfikacji innych osób (np. w zespole błędnej identyfikacji, tj. zespole Capgrasa czy Fregoliego) oraz na temat uzasadniania własnych emocji i działań (np. w zespole postkomisurotomycznym). Zestawie-

nie tych przypadków i zbadanie ich podstaw neuronalnych pozwala na ujęcie konfabulacji w szerszej perspektywie badawczej.

Okazuje się, że mimo różnych uszkodzeń mózgu w każdym z tych zespołów można dopatrzeć się podobieństw, które zdaniem Hirsteina pozwalają na ujęcie ich w jednej kategorii konfabulacji. Podobieństwa te można zauważyć na poziomie uszkodzeń obszarów neuronalnych. We wszystkich przypadkach patologicznego konfabulowania występują bowiem uszkodzenia kory oczodołowo-czołowej lub jej połączeń z obszarem mózgu odpowiedzialnym za przetwarzanie danych pamięciowych bądź percepcyjnych. Uważa się, że obszar ten jest odpowiedzialny za monitorowanie dochodzących do niego danych, a jego uszkodzenie bądź uszkodzenie połączeń prowadzących do niego uniemożliwia sprawdzenie, czy dana informacja jest prawidłowa, czy nie. Dlatego też pacjenci mają błędne przekonania ugruntowane na nieprawidłowo funkcjonujących procesach poznawczych.

Tabela 2. Rodzaje konfabulacji patologicznych według Williama Hirsteina

Zakres konfabulowania	Rodzaj zespołu neuropsychologicznego
– percepcja wzrokowa	– zespół Antona
– percepcja ciała	– anozognozja przy hemiplegii
– pamięć	– zespół Korsakowa
– rozpoznawanie umysłów innych	– zespoły błędnej identyfikacji
– podejmowanie decyzji	– uzasadnianie decyzji przez pacjentów po komisurotomii

Źródło: Hirstein 2005.

Tym, co wspólne we wszystkich wymienionych przypadkach zespołów, w których występują konfabulacje, jest zatem obecność dwóch typów zaburzeń:

- specyficznych dla danego zespołu uszkodzeń mózgu,
- uszkodzeń systemu sprawdzającego reprezentacje, związanych z daną dziedziną wiedzy czy to pamięciowej, czy percepcyjnej, które są podstawą tworzenia przekonań.

Warunkiem wytworzenia konfabulacji jest więc dysfunkcja dwóch współdziałających ze sobą systemów: epistemicznego i sprawdzającego. Wewnętrzny dialog między procesami tworzenia wypowiedzi a procesami sprawdzającymi dane, których ta wypowiedź dotyczy, jest jednym z podstawowych składników naszego życia mentalnego. To jednak, zdaniem Hirsteina, do wytworzenia konfabulacji nie wystarczy. Trzecim ważnym czynnikiem konfabulowania jest tendencja do wytwarzania przekonań, które zapełniają lukę w systemie poznawczym.

Mimo iż Hirstein w swojej książce zajmuje się wyłącznie przypadkami patologicznymi, słusznie zauważa, że zakres terminu „konfabulacja” można rozszerzyć na opisy przypadków nieświadomego uzupełniania luk w systemie poznawczym zdrowych osób. Filozof podkreśla, że konfabulacja nie jest wyłącznie dziwnym, rzadko występującym fenomenem neurologicznym, którego zbadanie ma znaczenie wyłącznie dla neurologów i psychiatrów. Konfabulowanie jest zjawiskiem na tyle często występującym w ludzkiej populacji, że wyjaśnienie mechanizmów jego działania może mieć duże znaczenie dla innych dyscyplin badawczych, takich jak psychologia, nauki polityczne czy ekonomia.

Hirstein nie podejmuje zagadnienia konfabulacji w poprawnie funkcjonujących systemach poznawczych, wymienia jedynie kilka przypadków konfabulacji normalnych, takich jak: konfabulacje u dzieci, konfabulacje dotyczące uzasadnień wyborów konsumenckich, fałszywa pamięć, konfabulacje dotyczące podejmowania decyzji. Nie analizuje natomiast poszczególnych przypadków tego rodzaju konfabulowania ani nie precyzuje warunków i mechanizmów ich powstawania. Co więcej, nie wyznacza granicy między konfabulacjami patologicznymi a naturalnymi. Jego definicja konfabulacji stosuje się zarówno do pierwszego, jak i drugiego typu. Tymczasem nie jest jasne, czy warunki konfabulowania podane przez niego można zastosować do przypadków konfabulacji w zdrowych systemach poznawczych.

### 1.1. Pamięciowe ujęcie konfabulacji

Najbardziej powszechnym ujęciem zagadnienia konfabulacji jest ujęcie klasyczne, zwane inaczej pamięciowym. W ujęciu tym konfabulacje to opowiadane przez pacjentów fikcyjne historie, tworzone w celu wypełnienia luk w pamięci. Tę wąską definicję konfabulacji wspiera oficjalny *Diagnostic and Manual of Mental Disorders* (DSM IV), wydany przez Amerykańskie Towarzystwo Psychiatryczne, według którego konfabulacja to „deklamacja wyobrażonych zdarzeń wypełniających luki w pamięci” (za: Hirstein 2005, ss. 2-3).

Ujęcie to ma swe źródła w historii psychiatrii. Termin „zespół Korsakowa” zaproponował w 1897 r. Friedrich Jolly<sup>1</sup> (za: Schnider 2008, s. 9). W 1887 r. w swoim doktoracie psychiatra Siergiej Korsakow opisał 20 pacjentów cierpiących na chorobę alkoholową, u których zauważył bardzo podobne symptomy chorobowe. Krótco potem, w 1889 r., podsumował swoje obserwacje i opisał charakterystyczne objawy:

---

<sup>1</sup> Zespół Korsakowa występuje często u długoletnich alkoholików, ale nie tylko. Jest on skutkiem braku witaminy B1, zaobserwowano go więc także u anorektyków, u osób otrzymujących długotrwałe dożylnie odżywianie oraz u osób cierpiących na chroniczne uszkodzenia wątroby (Hirstein 2005, s. 50).

- bieżące percepcje znikają z pamięci prawie natychmiast, mimo to ciąg zdarzeń istnieje nieświadomie i nadal ma wpływ na łańcuch wyobrażeń pacjenta,
- pamięć odległych rzeczy, które zdarzyły się przed chorobą, jest czasami zachowana perfekcyjnie,
- różne rodzaje pamięci nie są wygaszane podobnie łatwo, pamięć nawyków trwa przez dłuższy czas (Schnider 2008, s. 12).

Jak zauważa Armin Schnider, sam Korsakow nazwał konfabulacje pseudoreminiscencjami. W 1891 r. w artykule *Illusions of memory (pseudo-remiscences) in polyneuritic psychosis* opisał je jako „sytuację, w której pacjent wymyśla sobie zdarzenie, którego w rzeczywistości nie doświadczył, ale które tylko przyszło mu do głowy, tak jakby wydarzyło się jemu samemu” (Schnider 2008, s. 12)<sup>2</sup>.

Na początku XX wieku Karl Bonhoeffer zdefiniował psychozę Korsakowa, opisując jej objawy w następujący sposób: „ostry defekt uczenia się, utrata pamięci bieżącej, utrata orientacji oraz silna tendencja do konfabulowania” (za: Schnider 2008, s. 9). Zdaniem Bonhoeffera konfabulacje powstają, kiedy pacjent tworzy fikcyjne narracje w celu wypełnienia luk w pamięci.

Obecnie wyróżnia się cztery składowe zespołu amnestycznego Korsakowa:

- luki pamięciowe zdarzeń bieżących (amnezja następcza),
- obecność amnezji wstecznej,
- konfabulacje (tj. zmyślenia),
- pseudoreminiscencje (zniekształcenia pamięciowe oparte na faktach, ale znacznie od nich odbiegające) (Herzyk 2005, s. 252).

Według Anny Herzyk konfabulacje w zespole amnestycznym Korsakowa „pojawiają się tylko w wybranych wariantach zespołu, gdy patologia rozszerza się na obszary podkorowo-czołowe i zaczynają pogłębiać się zaburzenia wglądu i samooceny. Nie znaleziono bezpośrednich korelacji między głębokością zaburzeń pamięci a obecnością lub brakiem konfabulacji. Wypełnienie luk pamięciowych zmyślonymi wspomnieniami można traktować jako zachowania kompensacyjne, żeby ukryć trudności” (Herzyk 2005, s. 252).

Konfabulacje w zespole Korsakowa powstają prawdopodobnie w wyniku działania dwóch oddzielnych mechanizmów, niesprawnych z powodu uszkodzeń mózgu:

---

<sup>2</sup> Pacjenci z zespołem Korsakowa cierpią zarówno na amnezję następczą, jak i wsteczną. Ponieważ nie są zdolni do tworzenia nowych wspomnień, często nie potrafią ocenić czasu, jaki spędzili w szpitalu, a nawet swojego wieku. Niekiedy nawet nie są świadomi tego, gdzie się aktualnie znajdują. Podobna dezorientacja występuje u osób z innymi rodzajami amnezji. Jednakże uszkodzenia mózgu w amnezji Korsakowa znacznie różnią się od innych przypadków utraty pamięci, co zdaniem Hirsteina wyjaśnia, dlaczego nie wszyscy pacjenci amnestyczni konfabulują (Hirstein 2005, s. 65).

– pacjenci mają trudności z odtwarzaniem informacji z pamięci w wyniku uszkodzeń w obszarach skroniowych (pod tym względem są podobni do osób z amnezją skroniową),

– pacjenci mają kłopoty z uświadomieniem sobie, iż wspomnienia, które przywołują, są wytworzoną przez nich fikcją. Problemy te są bezpośrednim skutkiem niepoprawnego działania procesów wykonawczych w uszkodzonych płatach czołowych (Hirstein 2005, s. 61).

Poglądy neuropsychologów na temat tego, która z wymienionych faz jest ważniejsza i kluczowa dla powstawania konfabulacji w zespole Korsakowa, znacznie się różnią. Dlatego można wyodrębnić dwie grupy teorii: teorie odtwarzania i teorie wykonawcze (Hirstein 2005, s. 44). Teorie odtwarzania odwołują się bezpośrednio do poglądów Korsakowa, który uważał, iż pseudoreminiscencje powstają w wyniku nieprawidłowego wydobycia informacji z pamięci (Schnider 2008, s. 12). Konfabulacje wedle tego pierwotnego podejścia spowodowane są przez deficyt strategicznego odtwarzania wspomnień, co skutkuje utratą sensu porządku czasowego wspomnień i ich źródła, tj. miejsca i czasu, jaki reprezentują. Teorie wykonawcze kładą zaś nacisk na kluczową przyczynę konfabulacji – uszkodzenie płatów czołowych odpowiedzialnych za procesy wykonawcze (Hirstein 2005, s. 61). Nie zaprzeczają one, iż powodem konfabulacji pamięciowych jest jakiś rodzaj amnezji wywołanej uszkodzeniami obszarów skroniowych, jednak bez dodatkowego czynnika, jakim jest niezdolność systemu poznawczego do wychwycenia niepoprawnych reprezentacji, konfabulacje na temat przeszłych zdarzeń w ogóle by nie powstały. Osoby, które mają zwykłe deficyty pamięci spowodowane amnezją skroniową, twierdzą po prostu, że nie pamiętają zdarzeń, nie tworząc rozbudowanych konfabulacji na temat przeszłości.

Wśród teorii wykonawczych wiążących konfabulacje w syndromie Korsakowa z uszkodzeniami płatów czołowych warto odnotowania są dwa stanowiska. Pierwsze to podejście Marcii Johnson, która sądzi, że konfabulacje spowodowane są deficytem funkcji wykonawczych związanych z monitorowaniem rzeczywistości (za: Hirstein 2005, s. 62). Monitorowanie rzeczywistości to zdolność odróżniania realnych zdarzeń od zdarzeń jedynie wyobrażonych, nabywana w trakcie rozwoju ontogenetycznego. Dla poparcia tego stanowiska przywołuje się przykłady konfabulacji u małych dzieci, których kora czołowa nie jest jeszcze wystarczająco rozwinięta, by mogły one odróżniać realne zdarzenia od wyobrażonych. Z wiekiem dzieci zaczynają jednak sprawniej rozpoznawać prawdziwe wspomnienia i odróżniać je od wyobrażeń.

Drugie stanowisko w ramach teorii wykonawczych to podejście zakładające, iż konfabulacje pamięciowe powstają w wyniku niesprawnie funkcjonujących procesów hamowania wspomnień nieodpowiednich do kontekstu. Zdaniem

Schnidera kora oczodołowo-czołowa wraz z jej limbicznymi połączeniami pełni funkcję inhibicyjną względem przypomnień nieodpowiednich do wykonania bieżącego zadania (za: Hirstein 2005, s. 63).

### 1.1.1. Krytyka pamięciowego ujęcia Hirsteina

Cechą charakterystyczną klasycznego podejścia do konfabulacji jest to, że jej powstawanie wiąże się z „wypełnianiem luk” w systemie pamięci. Pojawiają się tu jednak dwa istotne pytania: W jaki sposób konfabulacja wypełnia luki w pamięci? oraz: Czy konfabulowanie rzeczywiście polega na wypełnianiu luk wyłącznie w pamięci?

Odnosnie do pierwszego z tych pytań należy zauważyć, że z klasycznego podejścia do konfabulacji wynika, iż wypełnianie luk odzwierciedla ogólną tendencję ludzkiego umysłu do tworzenia spójnej, kompletnej reprezentacji świata, występującej nawet na bardzo wczesnych etapach przetwarzania informacji. Często przywołuje się w tym kontekście porównanie z wypełnianiem plamki ślepego oka. Plamka ślepa powstaje na tarczy nerwu wzrokowego, czyli w miejscu na siatkówce oka, w którym aksony komórek zwojowych opuszczają siatkówkę, tworząc nerw wzrokowy (Tessier-Lavigne 2000; Tovée 2003). Miejsce to, całkowicie pozbawione fotoreceptorów, jest niewrażliwe na światło, a zatem brakuje mu informacji wzrokowej o fragmencie obrazu. Luki w polu widzenia obu oczu nie pokrywają się, ponieważ plamki ślepe w gałkach ocznych umieszczone są symetrycznie względem płaszczyzny ciała. Zatem to, czego nie widzi jedno oko, widzi oko drugie. Nawet w sytuacji, gdy jedno oko jest zasłonięte, luka w polu widzenia uzupełniana jest odpowiednimi jakościami sensorycznymi i powstaje wrażenie, że pole widzenia jest pełne. Mózg nie ignoruje plamki ślepej, lecz niejako dorysowuje brakujący element, wykorzystując do tego jeden z procesów, takich jak wypełnienie, interpolacja lub uzupełnienie (Koch 2008, s. 67). Dzięki temu powstaje spójna reprezentacja sceny wzrokowej.

Można więc stwierdzić, że w tej sytuacji mózg dokonuje uzupełnienia. Todd E. Feinberg i David M. Roane nazywają to zjawisko „wypełnieniem konfabulacyjnym plamki ślepej” (za: Hirstein 2005, s. 188), podkreślając, że ma ono miejsce w zdrowych systemach poznawczych. Zatem po pierwsze, o konfabulacjach można mówić już na najwcześniejszych etapach przetwarzania informacji (np. informacji wzrokowej); po drugie, konfabulacje dotyczą także ludzi zdrowych oraz nie muszą mieć wyłącznie formy językowej.

Porównanie z plamką ślepą jest o tyle chybione, że np. „luka w pamięci” w przypadku osób z zespołem Korsakowa nie zostaje w ten sam właściwy sposób wypełniona, jak ma to miejsce w przykładzie zaczerpniętym z naturalnej percepcji wzrokowej. W przypadku osób konfabulujących reprezentacja świata

nie jest adekwatna<sup>3</sup>. Konfabulacje, choć często wewnętrznie spójne i kompletne, nie stanowią odzwierciedlenia tego, co rzeczywiście ma lub miało miejsce. W tych przypadkach luka w pamięci czy też w percepcji wzrokowej wprawdzie jest wypełniana, ale reprezentacjami fałszywymi, niezgodnymi ze stanem faktycznym przeszłych zdarzeń w życiu pacjenta czy też aktualnym stanem świata.

Ponadto w przypadku zespołu Korsakowa trudno jest mówić o wypełnianiu luki w pamięci epizodycznej bieżących zdarzeń. Pacjenci ci nie tyle mają luki w pamięci, ile raczej nie posiadają żadnych danych pamięciowych. Jest więc tak, że miejsce to wypełniane jest „raczej błotem niż zaprawą murarską” (Hirstein 2005, s. 17).

Głównym punktem krytyki podejścia pamięciowego nie jest jednak problem związany z wypełnianiem luki w pamięci, ale pytania bardziej podstawowe: Czy termin „konfabulacja” należy zastrzec wyłącznie do opowiadania przez pacjentów historii na temat własnej przeszłości? Czy konfabulacje są wyłącznie związane z deficytami pamięci, czy także z deficytami innych systemów poznawczych?

Należy zwrócić uwagę, że wraz z rozwojem psychiatrii termin „konfabulacja” zaczął być stosowany szerzej – do określenia zaburzeń u pacjentów, którzy nie posiadali żadnych deficytów pamięciowych. Zdaniem niektórych badaczy konfabulacje występują także w takich zespołach, jak: anozognozja przy hemiplegii, zespół Antona, zespoły dezidentyfikacji (tj. zespół Capgrasa, Fregolioego, Cotarda), zespół postkomisurotomiczny (Hirstein 2005).

W każdym z tych zespołów u pacjentów nie obserwuje się ubytków pamięci. Anozognozja przy paraliżu lewostronnym dotyczy konfabulacji na temat funkcjonowania lewej strony własnego ciała. Zaburzenie to występuje na skutek uszkodzenia okolic ciemieniowych prawej półkuli mózgu (Grabowska 2001, s. 35). Pacjenci cierpiący na to zaburzenie ignorują dolegliwości związane z paraliżem bądź niedowładem lewej strony ciała. Zachowują się tak, jak gdyby nie wiedzieli o własnych niedyspozycjach, twierdząc, że lewa strona ciała funkcjonuje w poprawny, niezmienny sposób. Potrafią nawet opisywać odczucia związane z podnoszeniem lewej ręki. Oto jeden z wywiadów, jaki przeprowadził Vilayanur S. Ramachandran ze swoją pacjentką cierpiącą na anozognozję:

Lekarz: I jak się teraz Pani czuje?

Pacjentka: Dobrze.

L: Czy może pani chodzić?

---

<sup>3</sup> Podobnie dzieje się w przypadku pacjentów z zespołem Antona, którzy będąc pozbawieni percepcji wzrokowej, konfabulują na temat tego, co rzekomo widzą. Można tu tylko umownie mówić o „wypełnieniu luki”, gdyż pacjenci ci nie mają żadnych danych percepcyjnych, a zatem tworzą od podstaw konfabulacje na ich temat.

P: Oczywiście, że mogę [...]

L: A co z rękami? Proszę podnieść ręce. Czy może Pani nimi ruszać? [...]

P: Oczywiście, że mogę ruszać swoimi rękami.

L: Czy może Pani podnieść prawą rękę?

P: Tak.

L: Czy może Pani podnieść lewą rękę?

P: Tak, mogę podnieść lewą rękę.

L: Czy obie ręce są równie silne?

P: Tak, obie są równie silne (Ramachandran i Blakeslee 1999, s. 128).

Ramachandran poprosił również pacjentkę, aby podniosła ręce i zaklaskała nimi. Kobieta twierdziła, że widzi obie swoje ręce klaszczące, a nawet słyszy głos klaskania, mimo iż podnosiła tylko prawą rękę.

Zespół Antona to z kolei zaburzenie polegające na ignorowaniu własnej ślepoty przez osoby, które cierpią na ślepotę korową<sup>4</sup>. Pacjenci nie reagują na bodźce wzrokowe, nie różnicują kolorów, kształtów i tekstury, ale opisują niekiedy z detalami to, co rzekomo widzą przed swoimi oczami. Osoby z zespołem Antona nie zdają sobie sprawy z tego, że nie widzą, w związku z czym zaprzeczają problemom z widzeniem, kiedy lekarz próbuje im je wykazać, np. „symulują czynność cichego czytania gazety, a poproszeni o głośne czytanie, odpowiadają, że boli ich głowa, że nie są w nastroju, że zapomnieli okularów lub że w pokoju jest zbyt ciemno” (Herzyk 2001, s. 46). Tego typu fikcyjne historie na temat „widzianych” przed oczami scen wzrokowych w literaturze neuropsychologicznej nazywa się również konfabulacjami.

Termin „konfabulacja” stosuje się także do określenia zaburzeń u osób z różnymi zespołami błędnej identyfikacji. Najbardziej charakterystycznym z nich jest zespół Capgrasa, czyli przekonanie, że osoby bliskie zostały zastąpione przez dublerów. Uważa się, że przyczyną tego zaburzenia jest dysocjacja między rozpoznaniem kognitywnym a afektywnym – twarze osób bliskich są nadal rozpoznawane, ale nie ma reakcji emocjonalnych, np. widocznych zmian oporności skóry (Ramachandran i Blakeslee 1999, s. 167). Osoby z zespołem Capgrasa twierdzą, że osoby bliskie przypominają im wyglądem ojca, matkę czy też żonę, ale według nich nie mogą być to osoby bliskie. Powodem tego jest to, iż uszkodzone zostały obszary odpowiedzialne za emocje towarzyszące widzianej twarzy. Oto fragment wywiadu z pacjentem z zespołem Capgrasa przeprowadzony przez Ramachandrana:

---

<sup>4</sup> Zaburzenie dotyczące ignorowania własnych dolegliwości, związane z przekonaniem, że nic się nie zmieniło, nosi w neuropsychologii szerszą nazwę: „anozognozja”. Zespół Antona jest więc jednym z przypadków anozognozji. Innym przypadkiem jest opisywana wyżej anozognozja przy paraliżu lewostronnym.

Lekarz: Arturze, kto przywiózł cię do szpitala?

Pacjent: Ten człowiek, który jest w poczekalni. To starszy pan, który się mną zajmuje.

L: Masz na myśli swojego ojca?

P: Nie, panie doktorze, ten człowiek nie jest moim ojcem. On tylko wygląda jak mój ojciec. Przypuszczam, że może być oszustem, ale nie wydaje mi się, by działał na moją szkodę (Ramachandran i Blakeslee 1999, s. 159).

Uważa się, że w poprawnie funkcjonującym mózgu informacje o widzianej twarzy przekazywane są do płata skroniowego mózgu, odpowiedzialnego za rozpoznanie. Informacja trafia dalej do ośrodków limbicznych, które dokonują interpretacji emocjonalnej. Postrzegana twarz bliskiej osoby jest więc ściśle powiązana z emocjami, które są aktywizowane zawsze, kiedy ją widzimy lub o niej myślimy. W przypadku zespołu Capgrasa zdolność ta jest upośledzona. Pacjenci mają uszkodzone połączenia układu limbicznego z ośrodkiem odpowiedzialnym za rozpoznawanie twarzy, tzw. zakrętem wrzecionowatym (Ramachandran i Blakeslee 1999, s. 167). W wyniku tego pojawia się u nich błędne przekonanie, że ich bliscy zostali zastąpieni przez dublerów. Tego rodzaju błędne przekonania również nazywa się konfabulacjami (Hirstein 2005, s. 101).

Innym zespołem neuropsychologicznym, do opisu którego stosuje się termin „konfabulacja”, jest zespół postkomisurotomiczny. W latach 60., po wykonaniu jednej z pierwszych operacji przecięcia spoidła wielkiego mózgu łączącego dwie półkule, Roger Sperry i Michael Gazzaniga odkryli, że lewa półkula mózgu, która zawiaduje szeregiem funkcji językowych, ma skłonność do konfabulowania na temat działań wykonywanych przez lewą rękę kontrolowaną przez prawą półkulę. Aby wywołać konfabulacje u pacjentów, badacze stworzyli wiele testów eksperymentalnych, w których środkiem do osiągnięcia celu było lateralne prezentowanie informacji wzrokowej za pomocą tachistoskopu<sup>5</sup>. Ekspozycje bodźców były bardzo krótkie – od 100 do 200 milisekundy, tak by badany wpatrujący się w kropkę na środku ekranu nie zdążył przenieść wzroku z punktu fiksacji na bodziec<sup>6</sup>. Następnie pytano badanych, co widzieli po lewej stronie pola widzenia. Za każdym razem byli nieświadomi wyświetlanych po lewej stronie pola widzenia obrazków lub słów. Niemniej gdy proszono ich, aby lewą ręką (za którą odpowiada prawa półkula)

<sup>5</sup> Obecnie do prezentacji lateralnej stosuje się komputery (Gazzaniga 2000, s. 1296).

<sup>6</sup> Na zapoczątkowanie ruchu sakkadowego gałki ocznej potrzeba 200 milisekund. Procedura stosowana w eksperymencie gwarantuje więc kierowanie informacji tylko do jednej półkuli. Przy operacji komisurotomii skrzyżowanie nerwów wzrokowych jest zachowane, toteż informacja z prawego pola widzenia obu oczu trafia do lewej półkuli, a z lewego pola widzenia obu oczu – do półkuli prawej. Ponieważ jednak półkule są rozdzielone i wymiana informacji między nimi nie zachodzi, w opisywanym eksperymencie lewa półkula widzi to, czego nie widzi prawa, i na odwrót.

wybrali jeden z wielu przedmiotów ukrytych pod ekranem, wybierali przedmiot, którego reprezentacja wizualna pojawiała się wcześniej po lewej stronie pola widzenia (Springer i Deutsch 1998, s. 43). Kiedy pytano ich, dlaczego wybrali ten właśnie przedmiot, wymyślali zgrabne historyjki (tj. konfabulowali) w celu uzasadnienia własnych wyborów. Ponadto, kiedy proszono ich o narysowanie lewą ręką obrazka, rysowali to, co wcześniej znajdowało się po lewej stronie. Zapytani, dlaczego namalowali taki a nie inny rysunek, wymyślali opowieść, aby uzasadnić własne działanie<sup>7</sup>.

Pacjenci po komisurotomii konfabulują nie tylko w zakresie podejmowanych przez siebie decyzji i działań, lecz także w zakresie uzasadniania własnych emocji. Jednej z pacjentek eksponowano po lewej stronie pola widzenia zdjęcie nagiej kobiety. Pacjentka, rumieniąc się, zaczęła się śmiać. Zapytana o przyczynę takiej reakcji odpowiedziała: „Ależ wy macie urządzenia” (za: Springer i Deutsch 1998, s. 43).

Jak wynika z powyższej argumentacji, termin „konfabulacja” stosuje się do wielu różnych zespołów neuropsychologicznych, nie tylko tych związanych z deficytami pamięci. Należałoby zatem traktować ten termin szerzej, stosując go do większej klasy zjawisk. Pomimo tego, że na poziomie neuroanatomicznym opisane uszkodzenia mózgu różnią się pod wieloma względami, warto się zastanowić, czy mają one jakiś wspólny mianownik na poziomie opisu funkcjonalnego. Co zatem sprawia, że uprawnione jest użycie w każdym z tych przypadków terminu „konfabulacja”?

## 1.2. Epistemiczne ujęcie konfabulacji

Ujęcie epistemiczne stara się uporać z problemami, jakie wynikają z ujęcia pamięciowego. Proponuje tym samym rozszerzenie znaczenia konfabulacji. Po pierwsze, znosi warunek pamięciowy ujęcia klasycznego. Konfabulacja nie jest wyłącznie związana z deficytem pamięci bądź z błędnym wydobywaniem informacji z pamięci. Oprócz pamięci, jak ma to miejsce w zespole Korsakowa, może ona także dotyczyć:

- aktualnej percepcji wzrokowej (zespół Antona),
- aktualnej percepcji własnego ciała (anozognozja przy hemiplegii), rozpoznawania umysłów innych (zespół Capgrasa),

---

<sup>7</sup> Gazzaniga uznał, że świadoma i językowa lewa półkula tworzy racjonalne wyjaśnienia dla działań podmiotu, aby zachować ciągłość narracji wewnętrznej (Gazzaniga 2000, s. 1316). Tego rodzaju narracja jest niezwykle istotna dla naszego świadomego „ja”, w związku z czym gdy tylko pojawia się jakaś luka, nasz system poznawczy natychmiast wypełnia ją odpowiednim rodzajem przekonań. Choć nie są to przekonania prawdziwe, często udaje się dzięki nim utrzymać spójność naszego systemu przekonań oraz naszego „ja”.

- uzasadniania celów własnych działań oraz podjętych decyzji (zespół postkomisurotomiczny),
- uzasadniania własnych emocji (zespół postkomisurotomiczny, pacjenci z heminegлекtem).

Ujęcie epistemiczne znosi także warunek werbalny ujęcia językowego. Konfabulacja może mieć postać nie tylko narracji językowej, lecz także reprezentacji niewerbalnej. Hirstein proponuje, aby konfabulacjami nazywać błędnie ugruntowane wypowiedzi oraz reprezentacje niewerbalne (Hirstein 2005, s. 177). Zarówno jedne, jak i drugie można nazwać przekonaniem na jakiś temat. Przekonania mogą być komunikowane za pomocą języka naturalnego w postaci pełnych zdań i całych narracji językowych, jak też poprzez proste stwierdzenia „tak” lub „nie” oraz rysowanie obrazków czy wskazywanie na nie. Przekonanie może więc być także komunikowane niewerbalnie.

W ujęciu epistemicznym rezygnuje się z jeszcze jednego warunku, za pomocą którego charakteryzuje się często konfabulacje, to znaczy z warunku fałszywości wypowiedzi. W tym podejściu konfabulacją są zarówno wypowiedzi, jak i reprezentacje niejęzykowe błędnie ugruntowane, co nie znaczy, że fałszywe (Hirstein 2005, ss. 177-212). Konfabulacje mogą być zarówno fałszywe, jak i prawdziwe – kluczem do ich zrozumienia nie jest więc kryterium fałszywości, lecz błędnego ugruntowania. Błędne ugruntowanie przekonania dotyczy niesprawnych bądź chwilowo źle funkcjonujących procesów poznawczych związanych z danymi (percepcyjnymi, pamięciowymi) oraz procesów sprawdzających te dane. Poza tym to nowe kryterium nie tylko rozszerza termin „konfabulacja” na reprezentacje niejęzykowe, lecz także odróżnia konfabulację od intencjonalnego kłamstwa bądź oszustwa.

Hirstein proponuje, aby konfabulację charakteryzować jako skutek trzech współwystępujących czynników:

- deficytu systemu epistemicznego (np. deficyt systemu pamięciowego, percepcyjnego),
- deficytu systemu sprawdzającego (problem ze sprawdzaniem danych pamięciowych, percepcyjnych itd.),
- tendencji do wytwarzania błędnie ugruntowanych przekonań (Hirstein 2005, s. 179).

Współwystępowanie wszystkich trzech czynników jest niezwykle istotne przy powstawaniu konfabulacji. Z tego powodu nie wszyscy pacjenci z amnezją konfabulują. Dzieje się tak dlatego, że nie wystarczy deficyt systemu epistemicznego, np. systemu pamięci. Aby wystąpiła konfabulacja, potrzebny jest także deficyt systemu sprawdzającego informacje epistemiczne.

Hirstein zauważa, że nie wszyscy pacjenci z amnezją uzupełniają konfabulacjami luki w pamięci. Zwykle jest tak, że pacjenci amnestyczni raportują,

iz nie pamiętają przeszłych wydarzeń. Pacjenci z amnezją skroniową posiadają rozległe uszkodzenia struktur mózgu odpowiedzialnych za przechowywanie danych pamięciowych, w związku z tym zdarzenie, o które zostali zapytani, nie jest reprezentowane w ich mózgach. Niemniej ich procesy sprawdzające funkcjonują poprawnie – monitorują brak danych pamięciowych. W przypadku osób, które konfabulują na temat pamięci, jest inaczej. Pacjenci ci nie są zdolni do skorygowania własnych przekonań, gdyż ich procesy sprawdzające funkcjonują niepoprawnie<sup>8</sup>. Są więc przekonani, że opierają swoje sądy na rzetelnych danych pamięciowych (Hirstein 2005, s. 65).

Analogiczna sytuacja ma miejsce w przypadku osób z zespołem Antona. Zespół ten Hirstein za Emilem Redlichem i Giulio Bonvicinim nazywa percepcyjnym odpowiednikiem zespołu Korsakowa (Hirstein 2005, s. 146). Hirstein zwraca uwagę na to, że nie każdy pacjent z uszkodzoną korą wzrokową konfabuluje na temat bieżących percepcji. Pacjenci niewidomi zwykle raportują, iż nie widzą. Tylko niewielki odsetek osób z uszkodzoną korą wzrokową cierpi na zespół Antona. Zdaniem Hirsteina dzieje się tak dlatego, że pacjenci ci oprócz deficytu epistemicznego (tu: percepcyjnego) mają uszkodzone procesy sprawdzające dane percepcyjne<sup>9</sup>. Prowadzi to do sytuacji, w której pacjent nie zdaje sobie sprawy z tego, że nie widzi<sup>10</sup>.

Schemat owych trzech czynników powodujących konfabulacje pasuje, zdaniem Hirsteina, do wszystkich przypadków klinicznych konfabulacji. Za każdym razem mamy do czynienia z jakimś deficytem wiedzy (wiedzy pamięciowej, wiedzy wynikającej z percepcji wzrokowej bądź percepcji ciała, wiedzy dotyczącej emocji przypisywanej znajomej twarzy), trudnością ze sprawdzaniem danych (pamięciowych, percepcyjnych, emocjonalnych) oraz z tendencją do tworzenia przekonań nieopartych na rzetelnych danych i rzetelnych procesach sprawdzających. Dlatego też przekonania pacjentów są błędnie ugruntowane – stanowią efekt działania nieprawidłowych procesów poznawczych.

Epistemiczne ujęcie konfabulacji można scharakteryzować za pomocą kilku współwystępujących warunków. Są to: (1) warunki dla błędnie ugruntowanej wypowiedzi lub (2) warunki dla błędnie ugruntowanej reprezentacji niewerbalnej. W pierwszej wersji warunki konfabulacji wyglądają następująco:

---

<sup>8</sup> Wiele danych anatomicznych, fizjologicznych i pochodzących z badań nad lezjami mózgu wskazują, że procesy sprawdzające znajdują się w obszarach przebiegających przez korę oczodołowo-czołową (Hirstein 2005, s. 180).

<sup>9</sup> W niektórych przypadkach zespołu Antona, oprócz uszkodzenia kory wzrokowej, obserwuje się rozległe uszkodzenia kory oczodołowo-czołowej (Hirstein 2005, s. 146).

<sup>10</sup> W pewnym sensie odwróceniem zespołu Antona można nazwać ślepowidzenie. Pacjent cierpiący na ślepowidzenie „nie wie, że widzi” (Herzyk 2005, s. 46).

Jan konfabuluje werbalnie wtedy i tylko wtedy, gdy:

1. Jan twierdzi, że p.
2. Jan jest przekonany, że p.
3. Myśl Jana, że p, jest błędnie ugruntowana.
4. Jan nie wie, że jego myśl jest błędnie ugruntowana.
5. Jan powinien wiedzieć, że jego myśl jest błędnie ugruntowana.
6. Jan jest pewny, że p (Hirstein 2005, s. 187).

W wersji z błędnie ugruntowaną reprezentacją niewerbalną warunki te mają następującą postać:

Jan konfabuluje niewerbalnie „r” wtedy i tylko wtedy, gdy:

1. Jan ma reprezentację „r”.
2. Reprezentacja „r” jest prawidłowa dla Jana.
3. Reprezentacja „r” jest błędnie ugruntowana.
4. Jan nie wie, że reprezentacja „r” jest błędnie ugruntowana.
5. Jan powinien wiedzieć, że reprezentacja „r” jest błędnie ugruntowana.
6. Jan jest pewny, że reprezentacja „r” jest prawidłowa (Hirstein 2005, s. 198).

W przypadku konfabulacji klinicznych przekonania i reprezentacje ikonizacyjne pacjentów są skutkiem niepoprawnie funkcjonujących procesów formowania wiarygodnych przekonań i reprezentacji. Przekonania pacjentów są błędnie ugruntowane z uwagi na uszkodzenie systemu epistemicznego (np. pamięciowego lub percepcyjnego). Dodatkowo pacjenci mają deficyty systemu monitorującego reprezentacje werbalne i niewerbalne, o czym mówi warunek czwarty konfabulowania.

### 1.2.1. Dwie fazy konfabulowania według Hirsteina

Najważniejszym założeniem epistemicznej teorii konfabulacji Williama Hirsteina jest określenie dwóch etapów powstawania konfabulacji w różnych zespołach neuropsychologicznych. Po przeanalizowaniu wielu danych empirycznych filozof zauważył, że we wszystkich przypadkach konfabulacji klinicznych występują dwie fazy, które prowadzą do powstania konfabulacji:

- błąd w systemie epistemicznym – percepcyjnym bądź pamięciowym,
- dysfunkcja systemu sprawdzającego dane z danego systemu wiedzy.

System epistemiczny pozwala na wytwarzanie przekonań i jest częścią systemu poznawczego. Zdaniem Hirsteina pacjenci konfabulują, ponieważ z powodu deficytu wiedzy percepcyjnej bądź pamięciowej ich mózgi produkują twierdzenia o niższym prawdopodobieństwie niż wówczas, gdyby nie były uszkodzone. Usterka w systemie epistemicznym (czyli systemie tworzenia przekonań) prowadzi do mniejszych bądź większych błędów. Przy mniejszych deficytach tworzone są przekonania fałszywe, lecz prawdopodobne. Wraz ze

wzrostem uszkodzeń obszarów mózgu związanych z generowaniem przekonań powstają przekonania nieprawdopodobne. Zdaniem Hirsteina wspólną cechą konfabulacji klinicznych jest to, że pacjenci mają przede wszystkim deficyt w jakiejś dziedzinie wiedzy. Konfabulacje kliniczne dotyczą różnych poziomów wiedzy w zależności od tego, jakie obszary mózgu są u pacjenta uszkodzone. Konfabulacje mogą więc dotyczyć:

- wiedzy na temat przeszłych zdarzeń, w tym wiedzy na temat własnej przeszłości (zespół Korsakowa),
- wiedzy na temat percepcji wzrokowej (zespół Antona),
- wiedzy na temat własnego ciała oraz jego otoczenia (anozognozja przy hemiplegii, hemianopia, czyli utrata połowy pola widzenia),
- wiedzy na temat własnego umysłu (zespół postkomisurotomiczny, przypadku stymulacji mózgu),
- wiedza na temat innych ludzi (zespoły błędnej identyfikacji).

Wyróżnione przez Hirsteina rodzaje wiedzy można rozpatrywać jako różne rodzaje samowiedzy, co autor zdaje się pomijać w swojej książce. Zauważa jedynie, że każdy z owych rodzajów wiedzy jest podstawą tworzenia naszych przekonań i na co dzień nie zastanawiamy się nad tego rodzaju zdolnościami naszego systemu poznawczego. Stwierdzenia takie jak: „Potrafię poruszać ręką”, „Pamiętam, co wydarzyło mi się wczoraj”, „Rozpoznaję swoją rodzinę”, „Widzę, co jest przede mną”, „Wiem, czy zrobiłem coś celowo, czy też nie”, brzmią dziwnie z uwagi na ich potoczną oczywistość. Tymczasem, jak się okazuje, mózg wytworzył specjalne mechanizmy tworzenia ugruntowanych przekonań pamięciowych i percepcyjnych, które zwłaszcza w przypadkach uszkodzeń niektórych obszarów mózgu zawodzą.

Charakterystyczne dla rodzajów wiedzy wyróżnionych przez Hirsteina jest to, że większość z nich dotyczy wiedzy percepcyjnej (percepcja wzrokowa, percepcja ciała, percepcja własnego umysłu bądź umysłów innych), podczas gdy jeden typ wiedzy dotyczy wiedzy pamięciowej<sup>11</sup>. Obszary mózgu odpowiedzialne za gromadzenie i przetwarzanie tego rodzaju danych to płat poty-

<sup>11</sup> Hirstein zauważa jednak, że nie należy wprowadzać dychotomii między konfabulacjami percepcyjnymi a pamięciowymi, gdyż zarówno pamięć, jak i percepcja są dziedzinami wiedzy. Zabieg ten jest zrozumiały, kiedy zwrócimy uwagę na główny cel książki. Filozofowi zależy na tym, aby odnaleźć cechy wspólne wszystkich typów konfabulacji. Dokonanie podziału na konfabulacje pamięciowe i percepcyjne spowodowałoby doszukiwanie się różnych dla nich szczegółowych mózgowych mechanizmów konfabulowania i odejście od bardziej podstawowych wspólnych cech wszystkich typów konfabulacji. Hirstein widzi te różnice na poziomie deficytów poznawczych oraz uszkodzeń mózgu, które do nich prowadzą, o czym pisze w poszczególnych rozdziałach (II-VII), niemniej celem głównym książki jest ogólna epistemologiczna analiza wszystkich przypadków konfabulacji, która każe mu wskazać na znaczenie cech charakterystycznych konfabulowania w ogóle.

liczny i ciemieniowy w przypadku percepcji oraz płat skroniowy w przypadku pamięci.

Uszkodzenie obszarów systemu epistemicznego, związanego z generowaniem przekonań na podstawie niepełnej bądź brakującej wiedzy pamięciowej czy percepcyjnej, nie wystarcza zdaniem Hirsteina do powstania konfabulacji. Potrzebny jest jeszcze jeden czynnik, który powstaje na dalszych etapach formowania przekonań. Drugą istotną fazą konfabulowania jest uszkodzenie mechanizmów sprawdzających dane. Funkcje sprawdzania danych pełnią płaty czołowe. Czym są jednak owe mechanizmy sprawdzające dane?

Według Hirsteina mózg posiada wiele mechanizmów wykonawczych, których uszkodzenie prowadzi do niezdolności do autokorekty, autorefleksji oraz monitorowania własnych stanów umysłu. Należą do nich:

- sprawdzanie pamięci autobiograficznej,
- sprawdzanie wewnętrznych sprzeczności (przy użyciu pamięci semantycznej),
- sprawdzanie wiedzy faktualnej (poprzez pamięć semantyczną),
- sprawdzanie reprezentacji cielesnych,
- sprawdzanie poprzez obserwowanie słuchaczy,
- modelowanie wpływu bieżącej myśli na odbiorców,
- sprawdzanie prawdziwości bieżących obrazów wzrokowych,
- ocenianie wartości nagrody bieżącej myśli,
- dostrzeganie logicznych i przyczynowych konsekwencji bieżącej myśli oraz ocena wartości jej nagrody,
- wyznaczanie stopnia pewności dla bieżącej myśli,
- rozwijanie intencji do działania opartego na bieżących myślach (Hirstein 2005, s. 178).

Zdolności te rozwijają się albo w sposób naturalny – jak w przypadku monitorowania zawartości pamięci autobiograficznej, albo są jawnie wyuczone – jak w przypadku sprawdzania konsekwencji logicznej bieżących myśli. Hirstein zaznacza, że tym, co jest wspólne wszystkim rodzajom procesów sprawdzających, jest ich funkcja epistemiczna. Polega ona na tym, iż „myśli, które zostały wyselekcjonowane jako przekonania, mają najwyższy stopień prawdopodobieństwa prawdziwości, jakie system poznawczy danej osoby może osiągnąć” (Hirstein 2005, s. 178).

W jaki sposób dochodzi jednak do powstania konfabulacji jako błędnie ugruntowanego przekonania? Zdaniem Hirsteina pierwsza faza konfabulowania to wytworzenie błędnej reprezentacji przez procesy poznawcze skorelowane z któryś z obszarów wiedzy (błąd systemu wiedzy). Druga faza konfabulacji to niezdolność uchwycenia przez procesy sprawdzające nieprawności wytworzonej reprezentacji (błąd sprawdzania).

Zwrócenie uwagi na dwie fazy powstawania konfabulacji sprawiło, że koncepcja ta została nazwana dwufazową teorią konfabulacji (Hirstein i Ramachandran 2009, s. 120). Nacisk kładziony na procesy sprawdzające daje możliwość przyjrzenia się różnego rodzaju konfabulacjom pod kątem teorii wykonawczych. Ramachandran i Hirstein (2009, s. 121) przywołują w tym kontekście teorię wykonawczą odwołującą się do mechanizmów monitorowania rzeczywistości Marcii Johnson, która konfabulacje pamięciowe w zespole Korsakowa wyjaśnia na podstawie procesów monitorujących. W zdrowych i rozwiniętych systemach poznawczych monitorowanie rzeczywistych wspomnień w odróżnieniu od wyobrażeń nie stanowi problemu, jednak uszkodzenie obszarów w płatach czołowych prowadzi do niezdolności monitorowania statusu danej reprezentacji, w związku z czym pacjenci myślą wspomnienia z wyobrazeniami. Istotnym porównaniem jest system poznawczy małego dziecka, którego nierozwinięte w pełni płaty czołowe nie radzą sobie wystarczająco dobrze z odróżnianiem wspomnień od wyobrażeń. Zdolność rozpoznawania wspomnień właściwych rozwija się, zatem ten rodzaj sprawdzania danych jest czymś, co podlega wyuczeniu.

Co ciekawe, zdaniem Ramachandrana i Hirsteina procesy wykonawcze uczestniczą także w rozpoznawaniu obiektów w przestrzeni oraz ludzi. Dzięki nim jesteśmy w stanie odrzucić wstępne błędne identyfikacje przedmiotów oraz innych osób. Każdemu z nas czasem zdarza się rozpoznać w tłumie kogoś, kto ostatecznie okazuje się być tylko kimś podobnym do naszego znajomego. Wstępne rozpoznanie sprawia, że jesteśmy w stanie przyjąć, iż jest to ktoś, kogo znamy, jednak to dzięki procesom sprawdzającym orientujemy się, że osoba ta nie powinna się raczej znaleźć w tym tłumie ludzi, gdyż np. jesteśmy właśnie na koncercie jazzowym, a naszego znajomego nie interesuje ten rodzaj muzyki. Wstępna identyfikacja jest więc skorygowana przez wiedzę. Niestety w niektórych przypadkach zdolność sprawdzania zawodzi, co prowadzi do błędnej identyfikacji, tak jak w przypadku zespołu Capgrasa czy Fregoliego (Hirstein i Ramachandran 2009, s. 122).

Większość funkcji sprawdzających pełnią płaty czołowe, w szczególności grzbietowo-boczna kora czołowa, brzuszno-boczna kora czołowa i kora oczodołowo-czołowa (Hirstein i Ramachandran 2009). Podczas sprawdzania różnego rodzaju danych aktywna jest często także przednia część zakrętu obręczy. Obecnie uznaje się, że ten obszar mózgu aktywny jest przy rozwiązywaniu różnego rodzaju konfliktów pomiędzy działaniami rutynowymi, które nie są odpowiednie do aktualnego zadania oraz działaniami nowymi, które są istotne dla wykonywanego zadania (Hirstein i Ramachandran 2009).

### 1.3. Kategoryzacje konfabulacji klinicznych oraz normalnych

Hirstein wyznacza epistemiczne kryteria dla konfabulacji klinicznych w ramach swojej epistemicznej teorii konfabulacji. Kryteria te są następujące:

S konfabuluje (że p), wtedy i tylko wtedy, gdy:

1. S twierdzi, że p.
2. S jest przekonany, że p.
3. S-a myśl, że p jest błędnie ugruntowana.
4. S nie wie, że jego myśl jest błędnie ugruntowana.
5. S powinien wiedzieć, że jego myśl jest błędnie ugruntowana.
6. S jest pewien, że p (Hirstein i Ramachandran 2009, s. 111).

Wyróżnione kryteria odnoszą się do koniecznych warunków konfabulacji, a także odróżniają konfabulowanie od innych fenomenów epistemicznych. Kryterium pierwsze uwzględnia nie tylko werbalne akty mowy, ale też odpowiedzi na pytania w formie niewerbalnej, takie jak rysowanie obrazków czy wskazywanie na przedmioty oraz elementy obrazów. Kryterium drugie – bycie przekonanym, że p – odróżnia konfabulację od kłamstwa i intencjonalnego zwodzenia odbiorcy, a także od świadomego twórcy fikcji. Osoba konfabulująca ma przekonanie co do prawdziwości treści swojego twierdzenia i nie ma intencji oszukania. Trzeci warunek konfabulacji nawiązuje do tego, że któryś z systemów tworzenia wiedzy (percepcyjny bądź pamięciowy) funkcjonuje niepoprawnie, co skutkuje tym, iż myśl podmiotu konfabulującego jest błędnie ugruntowana. Następny warunek wskazuje na drugą fazę powstawania konfabulacji – niezdolność sprawdzenia, czy myśl jest poprawnie ugruntowana z uwagi na to, że procesy sprawdzające informacje są uszkodzone. Kolejne kryterium konfabulacji ma charakter normatywny – jeśli procesy sprawdzające byłyby nienaruszone, S byłby w stanie sprawdzić i odrzucić konfabulacyjne przekonanie, że p. Ostatnie kryterium nawiązuje do dużego stopnia pewności przypisywanego przekonaniom konfabulacyjnym przez pacjentów oraz niezdolności korygowania własnych przekonań nawet w przypadku dowodów im przeczących.

Wyróżnione przez Hirsteina kryteria są warunkami konfabulacji patologicznych. Czy takie same kryteria można wyróżnić w przypadku konfabulacji w normalnych systemach poznawczych?

Wydaje się, że dwa pierwsze kryteria w przypadku konfabulacji normalnych powinny być takie same. Kryteria te określają formę konfabulacji (werbalną bądź niewerbalną) oraz odróżniają fenomen konfabulowania od kłamstwa. Wątpliwości rodzą się w przypadku dwóch kolejnych kryteriów, nawiązujących do błędnego ugruntowania przekonania oraz nieświadomości, iż przekonanie jest błędnie ugruntowane.

Co to znaczy, że przekonanie jest błędnie ugruntowane przez system poznawczy, który funkcjonuje sprawnie? Jakie są wyznaczniki reliabilności i niereliabilności w zdrowym, normalnie funkcjonującym systemie poznawczym człowieka? Jeśli w przypadku uszkodzeń mózgu łatwo jest ekstermalnie stwierdzić, że dany system poznawczy jest niesprawny, jak można to sprawdzić w przypadku zdrowych mózgów, które powinny funkcjonować poprawnie? Czy generowanie błędów przez zdrowy system poznawczy jest cechą adaptacyjną, czy też nie? Ponadto rodzi się pytanie dotyczące niezdolności podmiotu do wglądu w ugruntowanie przekonania. Analogicznie do poprzedniego warunku pojawia się pytanie: Jak ustalić czy podmiot ma możliwość sprawdzenia, że jego przekonanie jest konfabulacyjne, skoro system sprawdzający informacje jest nieuszkodzony?

Kolejne pytanie rodzi się w związku z kryterium piątym, które jest warunkiem normatywnym konfabulacji – S powinien wiedzieć, że jego myśl jest błędnie ugruntowana. Według Hirsteina kryterium to oznacza, iż podmiot konfabulujący miałby wgląd w błędne ugruntowanie własnych przekonań i myśli, gdyby jego procesy sprawdzające informacje były sprawne. Jednak w przypadku konfabulacji w poprawnie funkcjonujących systemach poznawczych ustalenie sprawności procesów sprawdzających stanowi duży kłopot. W związku z tym normatywne kryterium piąte nie tylko rodzi wątpliwości w przypadku konfabulacji normalnych, ale także w przypadku samych konfabulacji klinicznych, dla których kryteria Hirsteina zostały opracowane. Co to bowiem znaczy, że podmiot konfabulujący wiedziałby, iż jego przekonania są błędnie ugruntowane, gdyby obszary mózgu odpowiedzialne za sprawdzanie informacji były nieuszkodzone? Powszechnie występowanie różnych rodzajów konfabulacji w poprawnie funkcjonujących systemach poznawczych pokazuje, że warunek normatywny („S powinien wiedzieć, że jego myśl jest źle ugruntowana”) traci swoją moc.

Oczywiście przypadki konfabulacji w poprawnie funkcjonujących systemach poznawczych mają inny charakter niż konfabulacje kliniczne. Niemniej warunek ten rodzi spore wątpliwości także z innych powodów. Jeśli bowiem byłoby tak, jak mówi Hirstein, to wśród zespołów neuropsychologicznych powinny być przypadki osób, które mają uszkodzone obszary mózgu odpowiedzialne za przetwarzanie informacji percepcyjnej bądź pamięciowej, a jednocześnie mają nieuszkodzone obszary sprawdzające. Takie osoby mogłyby okazjonalnie formułować twierdzenia konfabulacyjne, które podlegałyby jednak korekcie z uwagi na to, że system sprawdzający działałby sprawnie. Tymczasem większość opisów neuropsychologicznych mówi o nekorygowalnych konfabulacjach pamięciowych czy percepcyjnych albo o całkowitym braku wiedzy w danym zakresie (np. amnezja wsteczna, ślepotą). Dlatego druga faza powstawania konfabulacji rodzi pewne wątpliwości.

---

Ostatnie, szóste kryterium można zastosować do konfabulacji normalnych, jednak przy pewnych zastrzeżeniach. Przypisywanie dużego stopnia pewności przekonaniom konfabulacyjnym ma miejsce także w tym przypadku. Jednak stopień pewności przypisywany normalnym konfabulacjom nie musi być tak wysoki jak w przypadku konfabulacji klinicznych. Ponadto, w przeciwieństwie do konfabulacji klinicznych, normalne przypadki konfabulacji polegają na tym, że przekonania konfabulacyjne mogą zostać po jakimś czasie skorygowane, kiedy podmiot uwzględni dodatkową wiedzę lub dowody przeczące jego fałszywym przekonaniom.



# Konfabulacja: pojęcie, odmiany, sposoby wyjaśniania

## 1. Trudności definicyjne związane z terminem „konfabulacja”

Czym jest konfabulacja? Jakie są kryteria ją wyróżniające? Jaka definicję konfabulacji przyjąć? Co różni konfabulacje kliniczne od konfabulacji normalnych oprócz oczywistej różnicy, jaką są uszkodzenia mózgu?

Zajmuję się tu głównie konfabulacjami normalnymi. Zagadnienie to było do tej pory pomijane przez filozofów i przedstawicieli nauk kognitywnych. Zastanowię się nad tym, czym konfabulacje normalne różnią się od patologicznych, jakie są różnice, jeśli chodzi o mechanizmy ich powstawania, czym różnią się pod względem efektu, jaki przynoszą oraz funkcji, jakie pełnią.

Konfabulacja nie jest zjawiskiem wyłącznie patologicznym. Zakres tego terminu należałoby zatem określić szerzej, niż czyni to William Hirstein, przez uwzględnienie w nim przypadków błędnie ugruntowanych przekonań generowanych w poprawnie funkcjonującym systemie poznawczym człowieka. Przykładem konfabulacji jest zjawisko fałszywej pamięci, występujące dość powszechnie w ludzkiej populacji. Badania takich psychologów poznawczych, jak Elisabeth Loftus czy Daniel Schacter, wykazały, że dotyczy ono każdego człowieka. Schacter przedstawił kategoryzację błędów pamięci na błędy pominięcia: nietrwałość, blokowanie, roztargnienie oraz błędy przekształcania: błędna atrybucja źródła, tendencyjność, sugestywność i uporczywość pamięci (Schacter 2003). Drugą kategorię błędów można zaliczyć do konfabulacji pamięciowych.

Innym przykładem konfabulacji w poprawnie funkcjonującym systemie poznawczym są zmyślenia małych dzieci w zakresie pamięci. Dzieci często nie odróżniają danych pamięciowych od wyobrażeń, w związku z czym traktują swoje wyobrażenia na temat przeszłości jako przypomnienia. Co więcej, podatność na sugestie we wczesnym wieku sprawia, że dzieci są przekonane, iż

przypominają sobie rzeczy, których nie były wcześniej świadkami (Schacter 2003).

Listę konfabulacji w poprawnie funkcjonujących systemach poznawczych można rozszerzyć o grupę błędów poznawczych, do której należą: błędy w podejmowaniu decyzji, błędy w przekonaniach i ocenie prawdopodobieństwa oraz stereotypy społeczne. Konfabulacja w potocznym rozumieniu bywa utożsamiana z kłamstwem bądź celowym zmyśleniem. Pojęcia te nie są jednak synonimiczne, wręcz przeciwnie – tym, co odróżnia konfabulację od kłamania, jest to, iż w przeciwieństwie do kłamcy osoba konfabulująca nie ma intencji oszukiwania rozmówcy. Jest ona całkowicie bądź częściowo przekonana co do poprawności swojej wypowiedzi. Co ciekawe, towarzyszy temu zwykle ogromna pewność, że głoszone przekonania są prawdziwe. Pacjenci konfabulujący nie mają świadomości, że treść ich wypowiedzi nie jest oparta o wiarygodny proces poznawczy, w związku z czym nie odczuwają oni niepewności ani wątpliwości co do wygłaszanych przekonań. Hirstein nazywa to „patologiczną pewnością” (Hirstein 2005).

## 2. Rodzaje konfabulacji. Zarys dotychczasowych podziałów

Z uwagi na to, że pod koniec XIX i na początku XX wieku termin „konfabulacja” odnosił się wyłącznie do nieprawdziwych opowieści pacjentów na temat własnej przeszłości, pisano o różnych rodzajach konfabulacji pamięciowych. Jednym z pierwszych znanych podziałów pamięciowych konfabulacji klinicznych jest podział dokonany przez Emila Kraepelina (1904) na konfabulacje proste i fantastyczne (*simple confabulation* vs. *fantastic confabulation*). Konfabulacje proste są skutkiem błędów w czasowym porządkowaniu prawdziwych wspomnień, toteż ich treść jest związana z przeżyтыми przez pacjenta wydarzeniami, choć w nieadekwatnym porządku czasowym. Konfabulacje fantastyczne to z kolei nieprawdopodobne przekonania pamięciowe, które nie są skutkiem zgromadzonych wcześniej prawdziwych wspomnień pacjenta.

Michael D. Kopelman w 1980 r. odróżnił konfabulacje prowokowane od spontanicznych. Pierwsze z nich powstają na skutek prowokowanych pytań badacza, a drugie w sposób spontaniczny. Rozróżnienie to jest istotne, gdyż konfabulacje mogą powstawać w sposób, który nie jest zależny od badacza (Kopelman 1980).

Ze względu na rodzaj systemu poznawczego, jakiego konfabulacje dotyczą, Hirstein dokonał podziału na konfabulacje pamięciowe, percepcyjne (np. w zakresie percepcji wzrokowej oraz percepcji ciała), na temat umysłów innych, na temat własnych emocji, na temat własnych intencji i działań (Hirstein 2005;

Hirstein i Ramachandran 2009). Filozof ten pokazał na przykładach różnych zespołów neuropsychologicznych, że konfabulacje nie muszą dotyczyć wyłącznie pamięci (jak w zespole Korsakowa), ale też innych procesów poznawczych. Hirstein dokonał ponadto podziału na konfabulacje werbalne i niewerbalne (Hirstein 2005), przy czym te drugie to np. błędne wskazywanie na obrazki, fotografie, odrysowywanie z pamięci elementów obrazu (dodawanie czegoś do odpamiętywanych obrazków).

Z kolei Armin Schnider (2008) dokonał podziału konfabulacji na pamięciowe i niepamięciowe (*mnesitic vs. non-mnesitic confabulations*), a pamięciowe podzielił na:

- prowokowane w sposób prosty – powstają podczas zwykłego testowania pamięci,
- chwilowe – powstają przy zadawaniu pytań i w czasie konwersacji,
- fantastyczne – nie mają podstaw w rzeczywistości, lecz są wytworem wyobraźni,
- behawioralnie spontaniczne – są skutkiem niepowodzenia w wyhamowaniu wspomnień nieadekwatnych do kontekstu (Schnider 2008).

Inne kryterium podziału konfabulacji zaproponował Todd E. Feinberg. Uznał on, że konfabulacje można podzielić ze względu na wewnętrzną motywację podmiotu konfabulacyjnego na: konfabulacje neutralne, które nie są skutkiem wewnętrznej motywacji podmiotu, lecz wynikiem błędu poznawczego, oraz konfabulacje personalne, zwane inaczej urojeniowymi, które wymagają wewnętrznej motywacji i odnoszą się zazwyczaj do osobistych doświadczeń podmiotu (Feinberg 2009, s. 93). Często są one fantastyczne (tak jak w zespole Capgrasa).

Reasumując, najważniejsze podziały konfabulacji dokonane w XX i XXI wieku to:

- konfabulacje proste vs. fantastyczne (Kraepelin 1904),
- konfabulacje prowokowane vs. spontaniczne (Kopleman 1980),
- konfabulacje prowokowane w sposób prosty, chwilowe, fantastyczne i behawioralnie spontaniczne (Schnider 2008),
- konfabulacje pamięciowe vs. niepamięciowe (*mnesitic vs. non-mnesitic confabulations*) (Schnider 2008),
- konfabulacje werbalne vs. niewerbalne (Hirstein 2005),
- konfabulacje pamięciowe (np. zespół Korsakowa), percepcyjne, na temat umysłów innych, na temat własnych emocji, na temat własnych intencji i działań (Hirstein 2005; Hirstein i Ramachandran 2009),
- konfabulacje neutralne vs. personalne (Feinberg 2009).

### 3. Podziały konfabulacji naturalnych

Konfabulacje można podzielić ze względu na różne, starannie dobrane kryteria:

- pierwszy podział, dokonany według tego, jakiego procesu poznawczego konfabulacje dotyczą, pozwala wyróżnić konfabulacje pamięciowe, percepcyjne i predykcyjne. Osobna klasa konfabulacji dotyczy uzasadnień podjętych decyzji oraz emocji bieżących lub przeszłych;
- drugi podział, dokonany ze względu na czas, jakiego konfabulacje dotyczą, pozwala wyróżnić konfabulacje na temat przeszłości (retencyjne), na temat bieżących doświadczeń (aktualne) oraz na temat przyszłości (protencyjne);
- trzeci podział, dokonany ze względu na treść konfabulacji, pozwala wyróżnić konfabulacje na temat świata (naturalistyczne), na temat innych (socjocentryczne) oraz na własny temat (egocentryczne);
- czwarty podział, dokonany z uwagi na formę konfabulacji, pozwala wyróżnić formę językową oraz ikoniczną;
- piąty podział, utworzony ze względu na typ procesu konfabulowania, pozwala wyróżnić konfabulacje wypełniające, scalające i addytywne.

#### 3.1. Podział konfabulacji ze względu na rodzaj procesu poznawczego

Pięć rodzajów konfabulacji wyróżnionych z uwagi na rodzaj procesu poznawczego, który jest odpowiedzialny za ich powstawanie, przedstawia tabela 3.

Tabela 3. Podział konfabulacji ze względu na rodzaj procesu poznawczego

Zakres konfabulowania	Rodzaj konfabulacji normalnej
Percepcja (wzrokowa, słuchowa, percepcja ciała)	konfabulacje percepcyjne
Pamięć	konfabulacje pamięciowe
Oczekiwanie	<i>konfabulacje predykcyjne</i>
Podjęcie decyzji	konfabulacje dotyczące uzasadniania dokonanych wyborów i podjętych działań
Emocje	konfabulacje dotyczące uzasadniania emocji

Źródło: opracowanie własne na podstawie: Hirstein 2005; Hirstein i Ramachandran 2009.

Podział ten jedynie w przybliżeniu przypomina podział Hirsteina. Filozof ten bowiem wyróżnił konfabulacje percepcyjne i pamięciowe, w zakresie emocji oraz podejmowanych decyzji (Hirstein 2005; Hirstein i Ramachandran 2009), pominął natomiast konfabulacje w zakresie oczekiwań. Dlatego uzupełniłam

proponowany przez niego podział o konfabulacje predykcyjne (wyróżnione kursywą).

Konfabulować można zatem nie tylko w zakresie pamięci i percepcji, ale także w zakresie oczekiwań. Dzieje się tak wówczas, gdy podmiot dokonuje niezgodnych ze stanem faktycznym przewidywań, np. w zakresie własnych planów na przyszłość, bądź czyni to w odniesieniu do przyszłych stanów świata. Tego typu konfabulacje nazywam predykcyjnymi. Przykładem patologicznych konfabulacji predykcyjnych są opowieści na temat przyszłości pacjentów z zespołem Korsakowa. Osoby te nie tylko nie orientują się w aktualnej sytuacji oraz w czasie przeszłym, lecz także nie potrafią trafnie oceniać przyszłości. Pacjenci mogą snuć bardzo ambitne plany na dzień następny oraz na dalszą przyszłość, mimo że ich stan jest bardzo poważny i wymagają długiej hospitalizacji.

Oczywiście zespół Korsakowa to skrajny przykład konfabulacji predykcyjnych. Jednak tego typu konfabulacje mogą pojawiać się w poprawnie funkcjonujących systemach poznawczych, np. w sytuacjach błędnego szacowania przyszłych stanów świata.

Innym typem konfabulacji są konfabulacje dotyczące dokonanych wyborów i działań. Są to konfabulacje *ex post*, będące zazwyczaj racjonalizacjami podjętych działań. Na przykład osoba, która dokonuje zakupu produktu danej marki, jest skłonna potem twierdzić, że był to zakup przemyślany, a produkt jest najlepszym z tego rodzaju produktów, choć w trakcie zakupu nie przywiązywała zbyt dużej wagi do oferty sprzedaży. Innymi słowy, aby zmniejszyć dysonans poznawczy między ceną i oczekiwaniami a aktualną użytecznością, ludzie często przypisują dużą wartość zakupionym przez siebie produktom.

Osobna grupa konfabulacji obejmuje konfabulacje na temat emocji aktualnie przeżywanych bądź przeżytych w przeszłości. Racjonalizacja przeżywanych emocji występuje wówczas, gdy podmiot próbuje zmniejszyć powstający dysonans poznawczy. Tego rodzaju racjonalizacje to konfabulacje dotyczące uzasadniania emocji.

### 3.2. Podział konfabulacji ze względu na lokalizację czasową

Konfabulacje można podzielić ze względu na lokalizację czasową treści przekonania konfabulacyjnego na:

- retencyjne – na temat przeszłych zdarzeń i doświadczeń,
- aktualne – na temat aktualnych zdarzeń i doświadczeń,
- protencyjne – na temat przyszłych zdarzeń i doświadczeń.

Podział ten częściowo pokrywa się z poprzednim podziałem, ponieważ pamięć dotyczy przeszłości, percepcja – tego, co aktualne, a oczekiwania – tego, co przyszłe. Został on wprowadzony po to, by pokazać, że treść przekonania

konfabulacyjnego dotyczy zarówno tego, co terazniejsze, jak i tego, co przeszłe i przyszłe. W przypadku emocji konfabulacja jest retencyjna, gdy dotyczy przeszłych emocji przeżywanych przez podmiot, aktualna – gdy dotyczy aktualnie przeżywanych emocji, a protencyjna – gdy dotyczy emocji oczekiwanych w związku z danym zdarzeniem. Podobnie jest z podejmowaniem działań – konfabulacja jest retencyjna, gdy dotyczy podjętych w przeszłości działań, aktualna – gdy dotyczy podejmowanych aktualnie działań, a protencyjna – gdy dotyczy działań zamierzonych w przyszłości.

Celem przedstawionego podziału jest pokazanie, że konfabulacje mogą dotyczyć zarówno tego, co przeszłe, jak i tego, co terazniejsze i przyszłe. W tradycji neuropsychologicznej konfabulacjami określano zmyślenia na temat przeszłości głównie w zespole Korsakowa. Tymczasem pacjenci z zespołem Korsakowa konfabulują nie tylko na temat przeszłości, ale także na temat zdarzeń bieżących oraz przyszłości (Dalla Barba 2009, s. 225).

Konfabulacje retencyjne w zdrowych systemach poznawczych to fałszywe wspomnienia na temat zdarzeń z przeszłości, na temat innych bądź samego siebie. Konfabulacje aktualne to zaś konfabulacje na temat bieżących zdarzeń i doświadczeń. Należą do nich: konfabulacje percepcyjne, które dotyczą percepcji wzrokowej, percepcji ciała bądź percepcji własnego umysłu lub umysłów innych osób.

Konfabulować można także na temat przyszłych zdarzeń i doświadczeń, wówczas mamy do czynienia z konfabulacjami protencyjnymi. Przykładem może być sytuacja, w której osoba cierpiąca na lęk wysokości, stojąc na wysokiej wieży, oczekuje, że budynek się zawali, bądź osoba panicznie bojąca się latania samolotem oczekuje wypadku lotniczego (jest to też przykład konfabulacji naturalistycznych). Inny przykład to tak zwana iluzja wstrząsu, czyli przecenianie długości i intensywności swoich przyszłych stanów emocjonalnych.

Konfabulacje protencyjne to także nierealistyczne przekonania na temat przyszłości. Nie należy ich jednak utożsamiać z marzeniami. Ludzie wyobrażają sobie przyszłość w różny sposób i wyobrażenia te czasem rozmijają się z rzeczywistością. Niemniej konfabulacje protencyjne występują w sytuacji, gdy osoba jest w dużym stopniu przekonana, że jakieś zdarzenie będzie miało miejsce, mimo iż prawdopodobieństwo jego wystąpienia jest bardzo niskie.

W psychologicznych badaniach nad pamięcią prospektywną stosuje się wzorzec wnioskowań, według którego pamięć prospektywna to inaczej pamięć tego, co ma się wydarzyć: że jutro pójde do pracy, że wieczorem pójde do kina, w srode mam umowiona wizyte u lekarza, a w czwartek jestem umowiona ze znajomymi. Konfabulacje protencyjne można by uznać za związane z pamięcią prospektywną. Ludzie czasem konfabulują na temat własnych oczekiwań wobec przyszłości i na temat tego, co rzekomo zaplanowali.

### 3.3. Podział konfabulacji ze względu na treść

Konfabulacje z uwagi na treść konfabulowania dzielą się na:

- egocentryczne – w zakresie własnych doświadczeń,
- socjocentryczne – w zakresie umysłów innych osób,
- naturalistyczne – w zakresie zdarzeń w świecie.

Przykładem konfabulacji egocentrycznych jest przypisywanie dużej roli czynnikom wewnętrznym w sytuacji odniesionego sukcesu. W psychologii poznawczej nazywa się to podstawowym błędem atrybucji. Polega on na tym, że osoba tłumaczy własne sukcesy czynnikami osobowościowymi, natomiast porażki przypisuje czynnikom sytuacyjnym. Atrybucje wewnętrzne w sytuacji odniesionego sukcesu można nazwać konfabulacjami egocentrycznymi.

Konfabulacje egocentryczne to także przeszacowywanie własnych zdolności i umiejętności. Badania wykazały, że ludzie często uważają, iż ich zdolności w danej dziedzinie są większe niż przeciętne, a nie jest przecież tak, że wszyscy ludzie mają ponadprzeciętne zdolności. Tego rodzaju konfabulacje to inaczej pozytywne samozłudzenia. Czasem jednak się zdarza, że ludzie mają niskie poczucie własnej wartości i myślą, iż ich zdolności w danej dziedzinie są mniejsze niż w rzeczywistości. Tego rodzaju konfabulacje to inaczej negatywne samozłudzenia.

Konfabulacje egocentryczne to także racjonalizacje własnych emocji oraz podejmowanych działań, np. gdy ktoś kupuje jakiś produkt w sposób bezrefleksyjny, tylko dlatego, że widział wcześniej jego reklamę. Po dokonaniu zakupu racjonalizuje podjęte działanie, tłumacząc, że wybrał dany produkt, gdyż był on najlepszą z możliwych opcji. Tymczasem żadne inne opcje nie były wcześniej przez niego rozważane. Racjonalizacja polega tu na zmniejszeniu dysonansu poznawczego – inne opcje mogłyby okazać się korzystniejsze, ale ludzie nie biorą ich pod uwagę, gdy wybór już został dokonany. Podobnie dzieje się w przypadku racjonalizacji przeżytych emocji. Ludzie często tłumaczą własne reakcje emocjonalne w taki sposób, aby zmniejszyć dysonans poznawczy. Tymczasem przyczyną emocji mogło być zupełnie coś innego niż to, co jest treścią konfabulacyjnej racjonalizacji.

Przykładem konfabulacji socjocentrycznych jest tłumaczenie sukcesów innej osoby czynnikami zewnętrznymi. Jest to tzw. błąd asymetrii aktor – obserwator. Podczas gdy własne sukcesy tłumaczy się czynnikami wewnętrznymi, sukcesy innych przypisuje się czynnikom sytuacyjnym. Tego rodzaju atrybucje sytuacyjne względem innych osób można nazwać konfabulacjami socjocentrycznymi.

Konfabulacje socjocentryczne to także wszelkiego rodzaju stereotypy społeczne. Przykładem może być efekt aureoli, który polega na tym, że jedna

cecha osoby, np. uroda, wpływa na ocenę innych jej cech, np. kompetencji. Innym przykładem może być efekt horoskopowy, czyli ocenianie innych osób na podstawie rzekomo precyzyjnych opisów, które w rzeczywistości z uwagi na swoją ogólność mogą odnosić się do wielu osób.

Konfabulacje naturalistyczne dotyczą z kolei cech świata zewnętrznego. Mają one miejsce wówczas, gdy ktoś błędnie szacuje, że dane zjawisko wystąpi lub jaka będzie jego siła. Na przykład pasażer samolotu, który panicznie boi się latać, uważa, że samolot, którym leci, z pewnością będzie miał wypadek. Tymczasem prawdopodobieństwo wypadku lotniczego jest mniejsze niż wypadku samochodowego. Konfabulacja naturalistyczna dotyczy w tym przypadku błędnego oszacowania prawdopodobieństwa danego zdarzenia.

### 3.4. Podział konfabulacji ze względu na formę

Konfabulacje mogą mieć formę językową bądź ikoniczną. Z uwagi na to dzielią się na:

- językowe – mające formę raportów werbalnych,
- ikoniczne – mające formę obrazkową.

Konfabulacje językowe pojawiają się wówczas, gdy ludzie zdają relację na temat swoich przekonań, które są błędnie ugruntowane. Konfabulacje ikoniczne mają postać reprezentacji ikonicznych, a ich przykładem są zadania na odrysowywanie z pamięci obrazków. Polegają one na tym, że badani niekiedy dorysowują elementy, które nie były prezentowane na obrazkach do zapamiętania, uzupełniając obraz (Hirstein 2005, s. 18).

### 3.5. Podział konfabulacji ze względu na typ procesu konfabulowania

Ostatni podział dotyczy typu procesu konfabulacyjnego. W związku z tym konfabulacje można podzielić na:

- wypełniające – będące efektem procesu wypełniania luki,
- scalające – będące efektem błędnego scalania,
- addytywne – będące efektem procesu dodawania czegoś do istniejących danych lub uzmienniania danych.

Punktem wyjścia do tworzenia konfabulacji jest brak danych w systemie wiedzy, występowanie jedynie szczątkowych danych bądź prawidłowe dane w systemie wiedzy. W zależności od stanu wyjściowego mamy do czynienia z trzema trybami konfabulowania: wypełnianiem konfabulacyjnym, scalaniem konfabulacyjnym i dodawaniem konfabulacyjnym.

Tabela 4. Trzy tryby konfabulowania i trzy rodzaje konfabulacji

Stan wyjściowy	Proces	Rezultat
Luka w systemie poznawczym	wypełnianie konfabulacyjne	konfabulacja wypełniająca
Szczątkowe dane w systemie poznawczym	scalanie konfabulacyjne	konfabulacja scalająca
Prawidłowe dane w systemie poznawczym	dodawanie/uzmiennianie konfabulacyjne	konfabulacja addytywna

Źródło: opracowanie własne.

Pozostaje jeszcze odpowiedź na pytanie, jakie konfabulacje, w zespołach neuropsychologicznych wymienionych powyżej, to konfabulacje wypełniające, scalające i addytywne. W przypadku zespołu Korsakowa neuropsychologowie zgadzają się, że zaburzenia polegają przede wszystkim na utracie części wspomnień. Powodem tworzenia konfabulacji może więc być brak danych w systemie wiedzy pamięciowej, a tryb konfabulowania to wypełnianie konfabulacyjne, którego efektem jest konfabulacja wypełniająca.

Podobnie jest w przypadku zespołu Antona. Pacjenci z tym zespołem zdają się nie wiedzieć, że nie widzą. Według Hirsteina wypełniają oni luki w wiedzy percepcyjnej poprzez nieświadome zdawanie relacji z rzekomego wyglądu sceny wzrokowej. Ten rodzaj konfabulacji również ma charakter wypełniający, a jej powodem jest luka w systemie wiedzy, zaś trybem konfabulowania – wypełnianie luki.

Innym przykładem są konfabulacje u osób po przecięciu spoidła wielkiego. Przy krótkiej ekspozycji bodźców percepcyjnych po lewej stronie widzenia osoby te mają dostęp do informacji. Informacja z lewej strony pola widzenia trafia do prawej półkuli, która przetwarza tę informację w sposób nieświadomy. Z uwagi na to, że prawa półkula nie może zinterpretować językowo danych percepcyjnych, lewa półkula tworzy konfabulacje na temat podjętych przez pacjenta decyzji. Pacjent nie wie, że jego prawa półkula przetwarza informację wzrokową, jednak zachowuje się, jakby informacja była przetwarzana. Jest to przykład konfabulacji scalającej, której podstawą są szczerkowe dane, a trybem konfabulowania – scalanie informacji.

Najbardziej kontrowersyjnym przykładem są konfabulacje addytywne, których punktem wyjścia są prawidłowe dane w systemie poznawczym, a trybem konfabulowania – dodawanie bądź uzmiennianie konfabulacyjne. Wydaje się, że sposób konfabulowania przysługuje głównie poprawnie funkcjonującym systemom poznawczym w sytuacji, gdy dostęp do informacji, np. pamięciowej

lub percepcyjnej, jest wystarczający, natomiast dodatkowe informacje wpływają na zabarwienie danych.

Zarówno konfabulacja wypełniająca, scalająca, jak i addytywna występują w zdrowych systemach poznawczych w różnych domenach wiedzy.

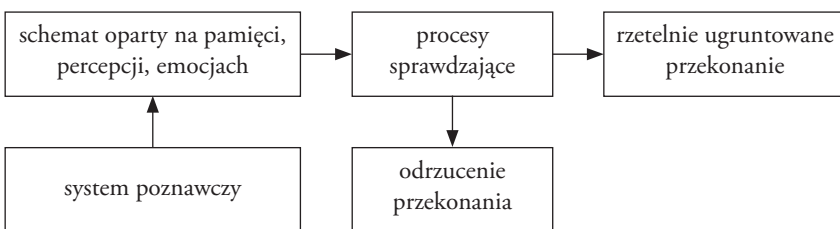
## 4. Mechanizmy konfabulowania i procesy konfabulacyjne

Złożoność badanego fenomenu konfabulacji wymaga precyzyjnego rozróżnienia na mechanizmy konfabulowania oraz procesy konfabulacyjne, jakie są następstwem działania owych mechanizmów.

### 4.1. Mechanizmy konfabulowania

Zgodnie z koncepcją Hirsteina konfabulacje powstają w wyniku dwóch błędów w systemie poznawczym: błędu systemu epistemicznego (tj. systemu pamięciowego lub percepcyjnego) oraz błędu systemu sprawdzającego. Powstające konfabulacje mogą więc mieć charakter patologiczny lub normalny.

Hirstein (2005) uważa, że rzetelnie ugruntowane przekonania powstają wówczas, gdy procesy sprawdzające wychwycą nieprawidłowości i pozwolą na powstanie przekonania opartych jedynie na rzetelnych procesach pamięciowych bądź percepcyjnych. System epistemiczny działa w tym przypadku sprawnie, a powstające przypadkowo nieprawidłowości zostają natychmiast wychwycone. Powstawanie rzetelnie ugruntowanych przekonań przedstawia schemat 1.

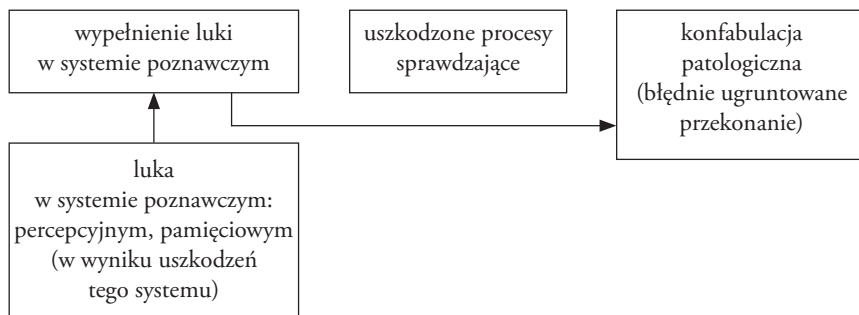


Schemat 1. Przebieg procesu poznawczego prowadzącego do rzetelnie ugruntowanego przekonania

Źródło: opracowanie własne na podstawie: Hirstein 2005.

Konfabulacje patologiczne powstają z kolei w wyniku uszkodzenia dwóch rodzajów procesów: epistemicznych i sprawdzających. W systemie wiedzy powstaje zatem luka, która jest uzupełniana. Następnie dane te przechodzą przez system sprawdzający, który nie wychwytyje jednak nieprawidłowości, gdyż

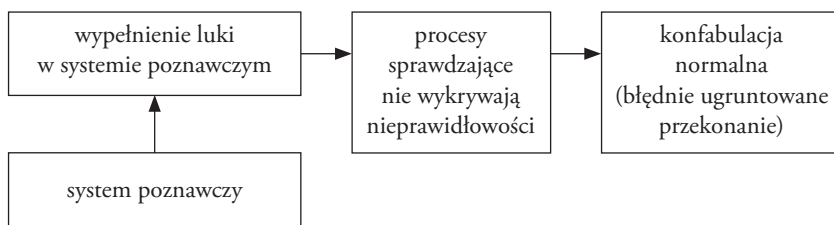
jest uszkodzony. Skutkiem takiego uszkodzenia jest konfabulacja patologiczna. Sposób jej powstawania przedstawia schemat 2.



Schemat 2. Etapy powstawania konfabulacji patologicznych

Źródło: opracowanie własne na podstawie: Hirstein 2005.

Jak jednak powstają konfabulacje normalne w poprawnie funkcjonujących systemach poznawczych? Hirstein nie precyzuje zagadnienia konfabulacji normalnych, skupiając się wyłącznie na konfabulacjach patologicznych. Jednak z nielicznych fragmentów jego książki można wnioskować, że jego zdaniem konfabulacje normalne powstają w wyniku wypełniania luk w systemie poznawczym (Hirstein 2005, ss. 13-14). System epistemiczny i sprawdzający nie zawsze są sprawne na tyle, aby tworzyć rzetelnie ugruntowane przekonania. Niekiedy brakuje danych w systemie epistemicznym, a luka ta jest uzupełniana informacjami, które nie są rzetelnie ugruntowane. Z kolei system sprawdzający nie zawsze wychwytuje nieprawidłowości. W ten sposób powstaje konfabulacja normalna w zdrowym systemie poznawczym. Sposób jej powstawania przedstawia schemat 3. Należy jednak pamiętać, że jest to tylko jeden z trzech możliwych rodzajów konfabulacji: wypełniającej, scalającej i addytywnej.



Schemat 3. Etapy powstawania konfabulacji normalnych

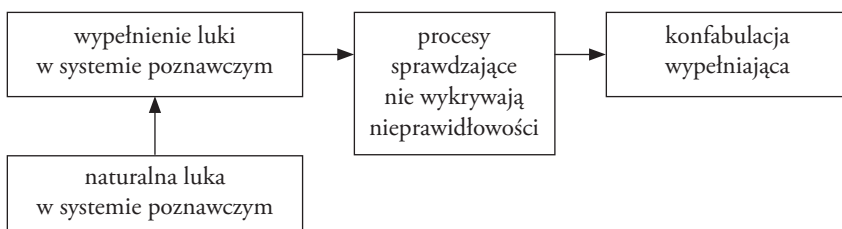
Źródło: opracowanie własne na podstawie: Hirstein 2005.

## 4.2. Charakterystyka procesu konfabulowania

### 4.2.1. Modele procesów konfabulowania

Modele procesów konfabulowania odwołują się do podziału konfabulacji ze względu na typ procesu konfabulowania, którego dotyczą. Pokazują one, w jaki sposób powstają wyróżnione konfabulacje: wypełniająca, scalająca i addytywna.

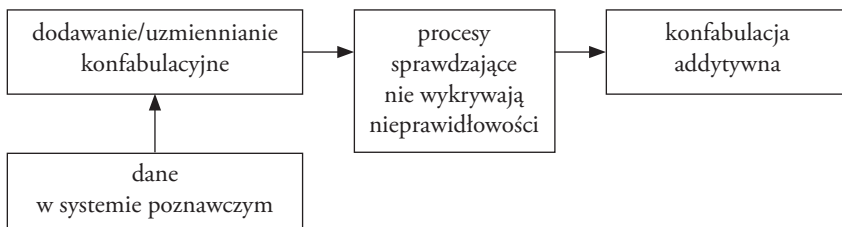
Wyróżniam trzy modele procesów konfabulowania. Pierwszy dotyczy konfabulacji wypełniającej. W systemie wiedzy pojawia się luka, która wymaga uzupełnienia. System poznawczy nie wychwytyje nieprawidłowości, co skutkuje konfabulacją wypełniającą. Wypełnianie konfabulacyjne przedstawia schemat 4.



Schemat 4. Schemat procesu wypełniania konfabulacyjnego

Źródło: opracowanie własne.

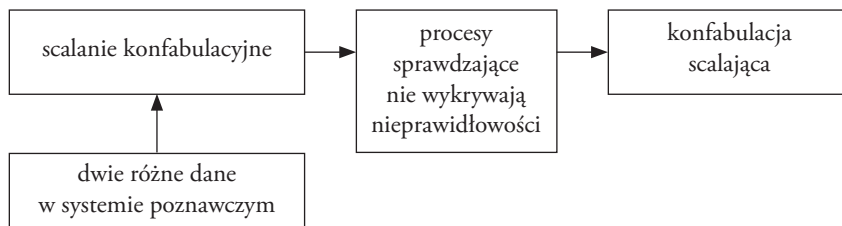
Drugim procesem konfabulacyjnym jest dodawanie bądź uzmiennianie konfabulacyjne. Ma ono miejsce wówczas, gdy dane w systemie poznawczym są w miarę pełne. Do istniejących danych dodawane są nowe elementy bądź dane są uzmienniane. Procesy sprawdzające nie wychwytyją nieprawidłowości, co skutkuje konfabulacją addytywną. Powstawanie konfabulacji addytywnej przedstawia schemat 5.



Schemat 5. Schemat procesu dodawania konfabulacyjnego

Źródło: opracowanie własne.

Trzecim procesem konfabulacyjnym jest scalanie konfabulacyjne. Następuje ono wówczas, gdy dwie różne dane w systemie poznawczym są scalane w jedną. Procesy sprawdzające, podobnie jak w poprzednich przypadkach, nie wychwytyują nieprawidłowości, co prowadzi do konfabulacji scalającej. Powstawanie konfabulacji scalającej przedstawia schemat 6.



Schemat 6. Schemat procesu scalania konfabulacyjnego

Źródło: opracowanie własne na podstawie: Hirstein 2005.

## 5. Konfabulacja a pokrewne fenomeny epistemiczne, czyli co jest, a co nie jest konfabulacją

### 5.1. Mówienie prawdy – konfabulacja a kłamstwo i fikcyjotwórstwo

W przypadku konfabulacji istotne jest przekonanie podmiotu konfabulującego o jego prawdomówności. Dlatego przedmiotem dalszych rozważań będą przekonania o szczerości i prawdomówności osoby konfabulacji, a nie definicja i problematyka prawdy jako takiej. Oznacza to, że konfabulacja, jako błędnie ugruntowane przekonanie, nie musi być utożsamiana z fałszem. Mogą bowiem istnieć błędnie ugruntowane przekonania, które są prawdziwe, np. w sytuacjach przypadkowych, i są one również konfabulacjami.

Kwestia kłamstwa i kłamania poruszana jest zwykle w dyskursie filozoficznym, głównie przez logików i semiotyków. Dla logików istotne są kwestie logiczne, w tym relacja między kłamstwem a fałszem, a także zagadnienie logiczno-semantycznych konsekwencji tego, że podmiot kłamie, orzekając o stanach rzeczy. Semiotycy z kolei zwracają uwagę na charakter znakowy kłamania. Jak zauważa Jolanta Antas (2008), proces kłamania to nie tylko logicznie bardzo złożona, ale i bardzo inwencyjna strategia, która sprawia, że skuteczny kłamca jest jednocześnie językowo sprawnym aktorem i fikcyjotwórcą. W związku z tym zagadnieniem kłamania zajmowali się także teoretycy literatury, których celem było „odgraniczenie fikcjonalnej wypowiedzi artystycznej od każdego innego

zdarzenia językowego, a w rezultacie także od wszelkich aktów życiowego zmyślenia, takich jak konfabulacja, dziecięce wymysły, językowe przesady i mistyfikacje, retoryczne demagogie czy wreszcie językowe kłamstwa i oszustwa” (Antas 2008, s. 6). Nowsza teoria literatury nie stawia już tych granic tak ostro, ponieważ narzędziem kłamstwa, fikcyjotwórstwa, zmyślenia i konfabulacji jest w tym przypadku język i jego właściwości referencjalne. Fikcja literacka ma swoje źródło w zwykłej językowej przesadzie, która prowadzi do zmyślenia, opowiadania historii, anegdot, często fantastycznych, oraz w aktach udawania przez mówiącego kogoś innego (Antas 2008).

Mimo że język jest narzędziem kłamania, językowe uchwycenie kłamstwa jest dość problematyczne. Kłamstwo nie daje się opisać za pomocą żadnej gramatycznej zasady ani reguły. Głównym powodem trudności w definiowaniu kłamstwa na poziomie języka jest to, iż skuteczne kłamstwo przybiera pozory prawdy lub prawdopodobieństwa, a dostęp do intencji kłamcy ma jedynie sam kłamca (Antas 2008).

Kwestia kłamania jest w pewnym stopniu podobna do zagadnienia konfabulacji. Chociaż narzędziem konfabulacji werbalnych jest język, to konfabulowania nie da się opisać za pomocą reguł gramatycznych, zwłaszcza w przypadku spójnych konfabulacji, które pod względem gramatycznym przypominają inne spójne wypowiedzi. Ponadto niektóre typy konfabulacji mają pozory prawdy. Jednak, w przeciwieństwie do kłamstwa, podmiot konfabulujący nie ma bezpośredniego wglądu w konfabulację. Konfabulacja nie jest bowiem świadomym oszukiwaniem odbiorcy, lecz nieświadomym uzupełnianiem luk w systemie poznawczym.

Często kłamstwo w potocznym rozumieniu niesłusznie kojarzone jest z wypowiadaniem fałszywych twierdzeń (Antas 2008). Fałszywe twierdzenia to tylko błędne poznawcze sądy lub logiczne przeczenia, ale nie kłamstwa. Kłamstwo można odkryć dopiero wówczas, gdy zna się rzeczywiste przeświadczenie mówiącego co do prawdziwości wygłaszanego sądu. Ponadto kłamca wcale nie musi uciekać się do wypowiadania twierdzeń ontologicznie fałszywych, aby kogoś wprowadzić w błąd. Strategią kłamcy jest posługiwanie się sądami prawdziwymi, aby kogoś oszukać (Antas 2008).

Jak widać, prawda z punktu widzenia pragmatyki językowej nie jest tożsama z jej mówieniem. Konfabulacja, podobnie jak kłamstwo, nie musi polegać na wypowiadaniu zdań fałszywych. Może się zdarzyć, że osoba konfabulująca powie prawdę, ale nie znaczy to, że nie konfabuluje. Na przykład kiedy pacjent z zespołem Antona stwierdzi, że widzi przed sobą stół, może się zdarzyć, iż stół stoi przed nim, co nie znaczy, że pacjent go widzi. Podobnie w przypadku pacjentów z zespołem Korsakowa – może się zdarzyć, że pacjent stwierdzi, iż przedwczoraj był czwartek, a przedwczoraj rzeczywiście był czwartek, ale

oznacza to raczej, że pacjent konfabuluje, niż zdaje relację z własnej pamięci. Konfabulacja jest więc – w przeciwieństwie do kłamstwa – wsparta przekonaniem o mówieniu prawdy.

Konfabulowanie różni się jednak od świadomego tworzenia fikcji. Opowiadacz fikcji zmyśla celowo, choć jego celem nie jest wprowadzenie w błąd odbiorcy. Zakłada on, że odbiorca ma dystans do komunikatu językowego. Z kolei osoba, która konfabuluje, sama nie mając dystansu do swojego komunikatu, nie zakłada także takiego dystansu u odbiorcy.

Zdaniem Antas kłamstwo jest nadrzędne wobec fikcji z dwóch zasadniczych powodów: po pierwsze, nie musi mieć wymiaru falsyfikacyjnego, gdyż może bazować na semantycznej prawdzie; po drugie, może korzystać z mechanizmów tworzenia fikcji, ale w przeciwieństwie do fikcji jest wypowiedzią fałszywą, która rości sobie pretensje do prawdziwości (Antas 2008, s. 24).

Konfabulacja to niejako tworzenie fikcji bez świadomości jej tworzenia. Mogą to być np. fikcje na temat przeszłości bądź na temat bieżących percepcji bez świadomości ich stosowania. Konfabulacja, podobnie jak kłamstwo, może być oparta na prawdzie semantycznej. Podobnie też jak kłamstwo, może opierać się na fikcji, która rości sobie pretensje do prawdziwości. Jedyna różnica jest taka, że konfabulant nie zdaje sobie sprawy z tego, iż jego sądy są fikcjonalne.

Fikcjotwórca jest w pewnym sensie oszustem, podobnie jak kłamca, jednak nie zakłada on, tak jak kłamca, zaangażowania odbiorcy w prawdziwość komunikatu. Konfabulant z kolei angażuje się w prawdziwość wypowiedzianego przez siebie zdania, czyli stwierdza je z mocną asercją. Zaangażowanie w prawdziwość wypowiedzianego zdania jest – zdaniem filozofa Johna Searle'a – podstawowym semantycznym i pragmatycznym warunkiem aktu illokucyjnego, jakim jest asercja.

Zdaniem Searle'a są dwa warunki aktu asercji:

- mówiący musi być przekonany o oczywistości lub być w stanie dostarczyć racji na rzecz prawdziwości wypowiedzianego zdania,
- wyrażone zdanie nie może być w sposób zbyt oczywisty prawdziwe ani dla nadawcy, ani dla odbiorcy (za: Antas 2008, s. 26).

Według Antas kłamca i fikcjotwórca zawieszają warunek pierwszy, ale w zupełnie inny sposób: fikcjotwórca nie zamierza nikogo oszukiwać ani zwodzić; sam niezaangażowany prawdziwościowo zakłada ten sam rodzaj niezaangażowania u odbiorcy; kłamca zaś, sam niezaangażowany prawdziwościowo, angażuje w to odbiorcę (Antas 2008). Inna różnica jest taka, że fikcjotwórca używa jedynie formy asertywnej, zaś kłamca dokonuje aktu asercji. Jest to różnica zasadnicza z punktu widzenia pragmatyki językowej, gdyż moc illokucyjna aktu zostaje zawieszona w przypadku opowiadania fikcji. Jest to akt jedynie pozorny i nikt nie zamierza fikcji realizować (Antas 2008).

W moim przekonaniu konfabulant sam jest zaangażowany prawdziwościami i tym różni się zarówno od kłamcy jak i opowiadacza fikcji. Konfabulant, podobnie jak kłamca, angażuje prawdziwościami odbiorcę komunikatu, ale z innej przyczyny – nie dlatego, żeby go oszukać, lecz dlatego, że sam jest zaangażowany w prawdziwość komunikatu.

Konfabulacja występuje tylko w przypadkach, gdy osoba jest zaangażowana prawdziwościami. Przypadek świadomego opowiadacza fikcji (literackiej, teatralnej, filmowej) bądź też świadome marzenia, wyobrażanie sobie czegoś lub fantazjowanie nie są przypadkami konfabulacji.

Te trzy fenomeny epistemiczne zostały ujęte w tabeli 5, która przedstawia zaangażowanie w prawdziwość wypowiedzi, angażowanie odbiorcy oraz intencje kłamcy, ficjotwórcy i konfabulanta.

Tabela 5. Zaangażowanie w prawdziwość wypowiedzi oraz intencje kłamcy, opowiadacza fikcji oraz konfabulanta

Rodzaj fenomenu epistemicznego	Asercja	Zaangażowanie prawdziwościami	Angażowanie odbiorcy komunikatu	Intencje
Kłamca	akt asercji	niezaangażowany prawdziwościami	angażuje odbiorcę prawdziwościami	ma intencję oszukania odbiorcy
Ficjotwórca	używa formy asertywnej	niezaangażowany prawdziwościami	nie angażuje odbiorcy prawdziwościami	nie ma intencji oszukania odbiorcy
Konfabulant	akt asercji	zaangażowany prawdziwościami	angażuje odbiorcę prawdziwościami	nie ma intencji oszukania odbiorcy

Źródło: opracowanie własne.

## 5.2. Pewność przypisywana przekonaniom.

### Konfabulacja a samooszustwo – podobieństwa i różnice

Innym ważnym kryterium konfabulowania jest pewność przypisywana przekonaniom konfabulacyjnym. Osoba konfabulująca przypisuje swoim przekonaniom duży stopień pewności. Ten warunek odróżnia konfabulację od samooszukiwania się, w którym stopień pewności przypisywany przekonaniom może być niski. Samooszustwo występuje w przypadku, gdy podmiot podtrzymuje przekonania, które są sprzeczne z posiadanymi dowodami. Samooszustwo i konfabulacja to fenomeny epistemiczne zbliżone do siebie, ale nie tożsame.

Czym jest samooszukiwanie się? Przykłady z życia łatwo jest podać: narkoman twierdzi, że panuje nad swoim nałogiem, student, który nie przygotował się do egzaminu, twierdzi, że uda mu się go zdać. Przykłady te dotyczą pozytywnego samooszukiwania się, czyli podtrzymywania przekonań korzystnych dla podmiotu wbrew dowodom im przeczącym. Istnieją jednak także negatywne samooszustwa, kiedy podmioty podtrzymują niekorzystną dla nich wersję rzeczywistości, np. hipochondryk może twierdzić, że jest nieuleczalnie chory, student, który przygotował się do egzaminu, twierdzi, że nic nie umie.

Zagadnieniu samooszukiwania się poświęcono wiele uwagi w refleksji filozoficznej. Definicje samooszustwa różnią się w zależności od tego, czy przyjmuje się stanowisko intencjonalne, czy nieintencjonalne (Piłat 2009). Intencjonalne definicje samooszustwa kładą nacisk na to, że to podmiot intencjonalnie wprowadza się w stan samooszustwa, a nieintencjonalne podkreślają, iż samooszustwo nie polega na intencjonalnym oszukiwaniu samego siebie.

Tradycyjny model samooszukiwania się wzorowany jest na modelach oszukiwania interpersonalnego, gdzie istnieje oszust, który jest przekonany, że nie-p, i oszukiwany, który jest przekonany, że p. W przypadku samooszukiwania się oszust i oszukiwany to ta sama osoba: A jest przekonany, że nie-p i oszukuje sam siebie, że p. Zwolennikiem tego podejścia jest m.in. Donald Davidson (1988). Według niego osoba A oszukuje się względem przekonania p wtedy, gdy:

1. A ocenia, że całość dostępnych jej danych, wspiera przekonanie, że nie-p.
2. A uważa, że nie-p.
3. Przekonanie, że nie-p wzbudza w A dyskomfort.
4. Z powodu 3. osoba A posiada motyw do poszukiwania argumentów przemawiających za p.
5. Osoba A celowo działa tak, by wzmocnić pożądane przekonanie p.
6. Na skutek 5. osoba A jest przekonana, że p.
7. W celu zdobycia i utrzymania przekonania p osoba A musi pogwałcić warunek kładący w nabywaniu przekonań odwoływać się do całości dostępnych danych.
8. Spełnienie warunków 1-7 czyni osobę A winną samoindukowanego słabego uzasadnienia jej własnego przekonania (za: Piłat 2009, s. 33).

Jak zauważa Robert Piłat, Davidson próbował połączyć w swojej definicji dwie strony zjawiska: motywację i rację.

Zarysował drogę od racji przekonania nie-p, przez motywację do przekonania p (opartą na niechęci do stanu rzeczy [nie-p] który jest przyczyną dyskomfortu) do ostatecznej racji przekonania p [...]. Przekonania te są logicznie sprzeczne, lecz podmiot ich tak nie traktuje, dlatego Davidson uważa, że osoba spełniająca warunki 1-8 jest irracjonalna, jej stany przekonaniowe są wprawdzie psychologicznie umotywowane, lecz nie są oparte na dobrych racjach (Piłat 2009, s. 33).

W tradycyjnym modelu samooszukiwania podmiot podtrzymuje dwa sprzeczne przekonania. Ponadto intencjonalnie prowadzi sam siebie do podtrzymywania przekonania, co do którego jest przekonany, że jest fałszywe. Podejścia intencjonalne powodują więc powstanie dwóch paradoksów:

– paradoksu statycznego, który koncentruje się na problemie stanu umysłu osoby samooszukującej się i polega na tym, że podmiot w tym samym czasie musi podtrzymywać dwa sprzeczne ze sobą przekonania;

– paradoksu dynamicznego, zwanego inaczej strategicznym, który koncentruje się na dynamicznym procesie samooszukiwania się i polega na tym, że podmiot jako oszust musi być świadomy, że rozwija strategię oszukiwania, lecz żeby zostać skutecznie oszukanym, powinien być nieświadomy rozwijanej przez siebie strategii.

Podejścia intencjonalne uznają, że tradycyjny model jest trafny, gdyż pomaga to wyjaśnić oczywistą odpowiedzialność osoby samooszukującej się za oszukiwanie samej siebie oraz odróżnić samooszukiwanie od innych rodzajów motywowanych przekonań, takich jak myślenie życzeniowe.

Podejścia nieintencjonalne uznają natomiast problematyczność paradoksów statycznego oraz dynamicznego, które wypływają z tradycyjnego modelu wzorowanego na intencjonalnym oszustwie interpersonalnym. W różny sposób starają się one rozwiązać owe paradoksy, odwołując się do nieintencjonalnych modeli.

Według Herberta Fingaretta samooszukiwanie się widoczne jest w działaniu i nie wymaga deklarowania przekonań (Fingarette 1969). Zdaniem Briana P. McLaughlina i Alfreda Mele samooszukiwanie zbliżone jest do myślenia życzeniowego (McLaughlin 1988; Mele 2001). Paul Thagard i Baljinder Sahdra uznają, iż samooszukiwanie nie jest intencjonalne, gdyż wiąże się z nieświadomymi emocjami (Sahdra i Thagard 2003). Natomiast zdaniem Daniela Golemana samooszustwo jest wynikiem działania nieświadomych schematów poznawczych (Goleman 1999).

Z kolei w ujęciu Williama Hirsteina samooszustwo to fenomen zbliżony do konfabulacji. Zjawisko konfabulacji i samooszustwa to pewne kontinuum z uwagi na stopień posiadanej pewności (Hirstein 2005). Uważa on, że zarówno w przypadku samooszustwa, jak i konfabulacji podstawa określonego przekonania jest słaba:

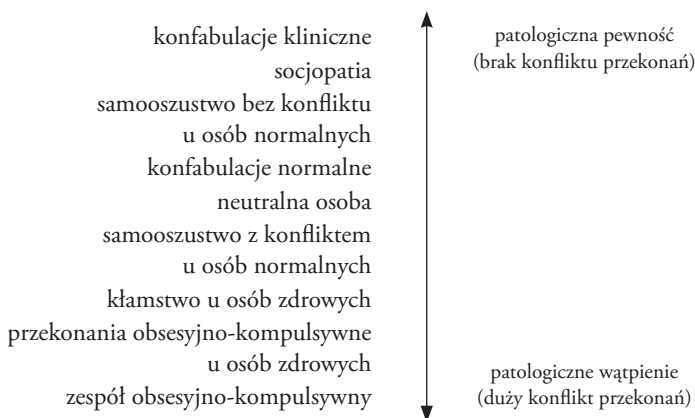
Przekonanie osoby samooszukującej się nie jest oparte na najbardziej rzetelnym procesie formowania przekonań, jaki jego mózg mógłby dostarczyć, gdyby nie był w stanie samooszukiwania się (Hirstein 2005, s. 213).

Różnica między samooszustwem a konfabulacją jest jednak taka, że osoba, która oszukuje sama siebie, ma większy dostęp do informacji lub procesów, które

potrzebne są do uznania, iż przekonanie jest błędnie ugruntowane. W przypadku konfabulacji klinicznej nie ma możliwości dostępu do informacji.

Dokonany przez Hirsteina podział przekonań w zależności od stopnia subiektywnej pewności im przypisywanym odwołuje się do kontinuum kilku pokrewnych dysfunkcji systemu poznawczego: konfabulacji patologicznej, socjopatii, samooszustwa patologicznego, zespołu obsesyjno-kompulsywnego oraz kontinuum fenomenów epistemicznych w zdrowym systemie poznawczym: konfabulacji normalnej, samooszustwa normalnego, przekonań poprawnie ugruntowanych, kłamstwa u osób zdrowych, zachowań obsesyjno-kompulsywnych u osób zdrowych (schemat 7).

Im bardziej schodzimy w dół piramidy fenomenów epistemicznych, tym większe pojawia się napięcie lub konflikt pomiędzy przekonaniami. Według Hirsteina problematyczna jest tu definicja „napięcia”. Dlatego kontinuum fenomenów epistemicznych można opisać za pomocą pojęć „ekstremalnej pewności” na górze oraz „ekstremalnego wątpienia” na dole. Na samej górze znajdują się konfabulacje patologiczne, w których pewność przypisywana błędnym przekonaniom jest bardzo wysoka. Kliniczne przypadki konfabulacji to takie, w których konfabulant nie doświadcza konfliktu przekonań. Pacjent z zespołem Korsakowa twierdzi, iż pamięta, co wydarzyło się w ciągu ostatniego tygodnia, a pacjent z zespołem Antona twierdzi, iż widzi. Hirstein uważa, że pacjenci konfabulujący, zwłaszcza ci odrzucający istnienie własnej choroby (tj. pacjenci z anozognozą przy hemiplegii czy pacjenci z zespołem Antona), tak skutecznie się samooszukują, że ich samooszustwo jest pozbawione napięcia.



Schemat 7. Kontinuum fenomenów epistemicznych w zdrowych i zaburzonych systemach poznawczych pod względem stopnia pewności przypisywanego przekonaniom

Źródło: opracowanie własne na podstawie: Hirstein 2005.

Po konfabulacjach klinicznych pojawia się socjopatia. Socjopaci, jeśli w ogóle posiadają procesy sprawdzające, to „ich procesy wydają się być pozbawione emocjonalnej siły potrzebnej do wyhamowania niemądrych lub bolesnych działań” (Hirstein 2005, s. 215). Socjopatia umiejscowiona jest między konfabulacją a samooszustwem bez konfliktu u osób zdrowych. Dalej są konfabulacje w zdrowych systemach poznawczych, samooszustwo z konfliktem, a następnie kłamstwo u zdrowych ludzi, zachowania obsesyjno-kompulsywne zdrowych osób i na samym końcu kliniczne zachowania obsesyjno-kompulsywne. Im niższy stopień piramidy fenomenów epistemicznych, tym większe jest napięcie bądź konflikt przekonań.

Warto jeszcze rozważyć wyróżnione przez Hirsteina konfabulacje patologiczne i normalne oraz samooszustwo z konfliktem i bez konfliktu (tab. 6). Hirstein wyróżnia dwa rodzaje samooszukiwania się: bez konfliktu oraz z konfliktem. Pierwsze z nich występuje wówczas, gdy nie ma konfliktu przekonań. Pojawia się ono nie tylko w przypadkach klinicznych u osób zaprzeczających własnej chorobie, ale także w zdrowych systemach poznawczych. Samooszustwo bez konfliktu zbliżone jest do konfabulacji.

Tabela 6. Stopnie pewności przypisywanej przekonaniom w wybranych fenomenach epistemicznych

Samooszustwo z konfliktem	Konfabulacja normalna	Samooszustwo bez konfliktu	Konfabulacja kliniczna
napięcie między przekonaniem	przypisywanie pewności przekonaniu (może być skorygowane)	duży stopień pewności przypisywanej przekonaniu	b. duży stopień pewności przypisywanej przekonaniom (patologiczna pewność)

Źródło: opracowanie własne.

Czym jednak jest samooszustwo bez konfliktu? Filozof zauważa, że nieobecność konfliktu przekonań może prowadzić do stwierdzenia, iż mamy do czynienia z konfabulacją, a nie samooszustwem. Można jednak podać przykłady samooszustwa bez konfliktu. Jednym z nich są iluzje pozytywne. Osoby, które samooszukują się w tym, iż ich zdolności są ponadprzeciętne, mogą nie doświadczać żadnego napięcia, gdyż nie mają wglądu w swoją prezentację lub w wymagane kompetencje. Ten brak napięcia zbliża je do osób, które konfabulują (Hirstein 2005, s. 214).

Samooszustwo z napięciem występuje zaś wtedy, gdy samooszukujący się nie jest całkiem pewny wygłaszanych przez siebie sądów, gdyż posiada dowody im przeczące.

Samooszukująca się normalna osoba, która doświadcza napięcia, nie wygłasza swoich sądów ze spokojną pewnością osoby konfabulującej (Hirstein 2005, s. 214).

Zatem o ile samooszustwo bez konfliktu zbliżone jest do konfabulacji, o tyle samooszustwo z konfliktem odróżnia się od niej. Związek tych dwóch fenomenów – samooszustwa i konfabulacji – można zdaniem Hirsteina opisać na poziomie neurologicznym. Filozof przywołuje w tym kontekście dwie fazy konfabulowania: wytworzenie błędnie ugruntowanego przekonania oraz niezdolność sprawdzenia, że myśl jest błędnie ugruntowana. W przypadku konfabulacji klinicznych procesy sprawdzające są uszkodzone, w związku z czym nie ma możliwości sprawdzenia bieżących myśli. W przypadku samooszustwa jakieś inne procesy w mózgu trzymają procesy sprawdzające z daleka od myśli, która ma być sprawdzona.

Hirstein uważa, że, podobnie jak w przypadku konfabulacji, istnieją dwie te same fazy samooszustwa: najpierw wytwarzane jest błędnie ugruntowane przekonanie, a potem przekonanie to jest chronione i podtrzymywane.

Metafora ochrony może być stosowana na różne sposoby. W niektórych przypadkach osoba może intencjonalnie nie pozwalać sobie myśleć, przypominać lub percypować pewnych faktów. Inny rodzaj występuje wówczas, gdy osoba samooszukująca się intencjonalnie nie używa pewnych procedur sprawdzających (Hirstein 2005, s. 216).

Wyróżnienie przez filozofa samooszustwa bez konfliktu jest moim zdaniem mylące, z uwagi na to, że trudno je odróżnić od konfabulacji. Sam Hirstein przyznaje, że osoba, która oszukuje sama siebie i nie odczuwa konfliktu, w rzeczywistości konfabuluje (Hirstein 2005, s. 226). Dlatego też uważam, że wyróżnienie osobnego fenomenu epistemicznego, jakim jest samooszustwo bez konfliktu, jest problematyczne. Tym bardziej że filozof ten podaje takie same warunki samooszukiwania się dla konfabulacji i dla samooszustwa bez konfliktu.

Warunki samooszustwa z konfliktem są zbliżone według Hirsteina do warunków konfabulowania patologicznego, ale z jednym wyjątkiem. Warto przyrzeć się warunkom konfabulacji wymienionym przez Hirsteina i sprawdzić, jak odnoszą się one do samooszustwa u normalnych osób.

Jan oszukuje sam siebie wtedy, gdy:

1. Jan twierdzi, że p.
2. Jan jest przekonany, że p.
3. Myśl Jana, że p jest błędnie ugruntowana.
4. Jan nie wie, że jego myśl jest błędnie ugruntowana.
5. Jan powinien wiedzieć, że jego myśl jest błędnie ugruntowana.
6. Jan jest pewien, że p (Hirstein 2005, s. 225).

Pierwszy warunek mówi o tym, że, podobnie jak konfabulant, osoba, która oszukuje sama siebie, wygłasza sąd, że p. W przypadku samooszustwa z konfliktem osoba, wygłaszając sąd, może przekonywać sama siebie i słuchaczy. Hirstein podkreśla, że może być tak, iż osoba, która oszukuje sama siebie, może obawiać się wygłoszenia sądu *explicite*, aby słuchacz nie zdemaskował jej samooszustwa. Dzieje się tak tylko w przypadku samooszustwa z konfliktem. Ktoś, kto oszukuje siebie i nie doświadcza konfliktu, jest podobny w tym względzie do konfabulanta.

Drugi warunek mówi o tym, że podmiot jest przekonany, iż p. Zdaniem Hirsteina warunek ten jest identyczny z warunkiem konfabulowania, gdyż samooszukujący się podmiot podtrzymuje przekonania, na podstawie których działa. Co ciekawe, zarówno konfabulant, jak i osoba oszukująca sama siebie „może mieć wyspy błędnie ugruntowanych przekonań wśród sieci przekonań dobrze ugruntowanych” (Hirstein 2005, s. 225).

Warunek trzeci mówi o tym, że myśl podmiotu samooszukującego się jest błędnie ugruntowana. Hirstein podkreśla, iż „być przekonanym, że p, jedynie z obawy, że nie-p, nie jest dobrą epistemiczną strategią, mimo że mogą istnieć nieepistemiczne powody dla tego przekonania, takie jak idea, iż odrzucenie go chroni nas od popadnięcia w depresję” (Hirstein 2005, s. 225).

Czwarty warunek mówi o tym, że podmiot nie wie, iż jego myśl jest błędnie ugruntowana. Uważam, że jest to warunek wysoce problematyczny, ponieważ w przypadku samooszustwa z konfliktem osoba może do pewnego stopnia zdawać sobie sprawę z tego, iż istnieją dowody przeczące jej przekonaniom. Według Hirsteina warunek ten z pewnością dotyczy samooszustwa bez konfliktu.

U osoby samooszukującej się, która nie wie, że jej przekonanie jest błędnie ugruntowane, rozgrywa się bitwa pomiędzy procesami chroniącymi przekonanie a procesami wypierającymi przekonanie. Osoba nie zdaje sobie sprawy lub nie wie, że p jest błędnie ugruntowane. Taka bitwa nie może mieć miejsca w umyśle konfabulanta, ponieważ procesy wypierania przekonania są niesprawne (Hirstein 2005, s. 225).

Piąty warunek mówi o tym, że osoba powinna wiedzieć, iż jej myśl jest błędnie ugruntowana. Ten warunek dotyczy zdaniem Hirsteina zarówno konfabulanta, jak i osoby samooszukującej się, tyle że w inny sposób. W przypadku konfabulacji patologicznych mózg jest uszkodzony, a w przypadku samooszustwa nie jest w najbardziej optymalnym stanie z epistemicznego punktu widzenia, aby wychwycić samooszustwo.

Ostatni warunek konfabulacji – duży stopień pewności przypisywany przekonaniom – jest w przypadku samooszustwa z konfliktem nieadekwatny. Osoba taka nie przypisuje bowiem dużego stopnia pewności swoim przekonaniom, gdyż doświadcza napięcia. Warunek ten, jak podkreśla Hirstein, można zastosować wyłącznie do samooszustwa bez konfliktu.

Hirstein mówi jednak wyłącznie o konfabulacjach klinicznych, zaś nie porusza wcale zagadnienia konfabulacji w zdrowych systemach poznawczych, choć wyróżnia takie konfabulacje, umiejscawiając je między samooszustwem bez konfliktu u osób zdrowych a neutralną osobą z dobrze ugruntowanymi przekonaniem. Warunki konfabulacji normalnych są więc jego zdaniem zbliżone do warunków samooszustw bez konfliktu, a co za tym idzie – do warunków konfabulacji patologicznych.

Samooszustwo i konfabulacje mają zdaniem Hirsteina takie samo podłoże neuronalne. Zwraca on uwagę zwłaszcza na jądro przyśrodkowo-grzbietowe wzgórza, które często jest uszkodzone u osób cierpiących na konfabulacje kliniczne. Ten obszar mózgu jest odpowiedzialny za koncentrację uwagi i podtrzymywanie rozpoczętej czynności. U osób konfabulujących mamy do czynienia z niekontrolowanym przedłużeniem natręctw myśli i mowy. Sytuacja odwrotna jest w przypadku osób samooszukujących się. Występuje u nich brak koncentracji na niechcianych myślach i emocjach. Samooszustwo polega więc w tym podejściu na hamowaniu niechcianych treści (Hirstein 2005, s. 228).

Innym ośrodkiem mózgu, na który wskazuje Hirstein, jest kora oczodołowo-czołowa. Uszkodzenie jej prowadzi do niemożności generowania wyhamowanych emocji. Dlatego też pacjenci konfabulujący nie wyhamowują swoich werbalnych zachowań. Tymczasem ten obszar mózgu odgrywa również dużą rolę w przypadku samooszustwa.

Emocje mogą także odgrywać rolę w tym, by osoba oszukująca sama siebie trwale porzuciła samooszukujące przekonania (Hirstein 2005, s. 228).

Według Hirsteina samooszustwo i konfabulacja to zatem fenomeny zbliżone do siebie, choć nie tożsame. Jego głównym dokonaniem było wyróżnienie dwóch rodzajów samooszukiwania się: samooszustwa z konfliktem i bez konfliktu. Ten drugi rodzaj samooszustwa jest prawie nieodróżnialny od konfabulacji, z uwagi na to, że nie występuje w nim napięcie pomiędzy przekonaniem. Uważam jednak, że używanie w tym przypadku określenia „samooszukiwanie się” jest mylące. Należy bowiem oddzielić fenomen konfabulacji od samooszustwa. Konfabulowanie występuje wówczas, gdy nie ma napięcia pomiędzy dwoma sprzecznymi przekonaniem. Jeśli napięcie występuje, mamy do czynienia z samooszukiwaniem się. W terminologii Hirsteina jest to samooszustwo z konfliktem.



CZĘŚĆ II

## ZAGADNIENIE KONFABULACJI – ZASTOSOWANIA



# Konfabulacje pamięciowe – badanie fałszywych wspomnień

## 1. Mechanizmy powstawania fałszywych wspomnień

Konfabulacje wyróżnione ze względu na rodzaj procesu poznawczego, którego są skutkiem, to konfabulacje: percepcyjne, pamięciowe, predykcyjne, dotyczące uzasadniania podjętych wyborów i działań, a także uzasadniania emocji. Szczególne znaczenie wśród nich mają konfabulacje pamięciowe. Etymologia terminu „konfabulacja” wskazuje na to, że pierwotnie stosowany był on wyłącznie na określenie błędnych przekonań pamięciowych. Ponadto konfabulacje pamięciowe to szeroka klasa konfabulacji, którymi warto zająć się z uwagi na możliwość szerszego i bardziej wnikliwego opracowania zagadnienia konfabulacji normalnych. Uwagę poświęcam tu konfabulacjom pamięciowym w poprawnie funkcjonujących systemach poznawczych, odwołując się do paradygmatu badań nad wspomnieniami fałszywymi.

Termin „konfabulacja” został po raz pierwszy użyty w psychiatrii na określenie zmyśleń pacjentów w zakresie pamięci. Przeanalizuję więc zagadnienie konfabulacji pamięciowych z uwagi na genezę samego pojęcia, jak i z uwagi na dyskusje toczone wokół konfabulacji pamięciowych. Jednym z podziałów konfabulacji jest podział dokonany przez Armina Schnidera, który wyróżnił konfabulacje pamięciowe i niepamięciowe (Schnider 2008). Podział ten wskazuje na to, jak szeroką klasą są konfabulacje odnoszące się do ludzkiej pamięci.

Analizując konfabulacje pamięciowe w poprawnie funkcjonujących systemach poznawczych, na podstawie wyników badań eksperymentalnych pokazuję, że ludzie w codziennym życiu podtrzymują błędne przekonania pamięciowe oparte na fałszywych wspomnieniach, które nazywam konfabulacjami pamięciowymi. Oznacza to, że ludzie nie tylko „przypominają” sobie zdarzenia, które w rzeczywistości nie miały miejsca, ale także ze szczegółami relacjonują fragmenty tych zdarzeń. Fałszywe wspomnienia to bowiem zarówno pamięć zdarzeń, które nigdy nie miały miejsca, jak i pamięć zdarzeń, które wprawdzie

wystąpiły, lecz w innej formie, niż przedstawiają to podmioty (zniekształcenia pamięci).

W artykule *False memories: A kind of confabulation in non-clinical subjects* (2009) Lauren French, Maryanne Garry i Elisabeth Loftus relacjonują obecny stan badań nad zagadnieniem fałszywych wspomnień, które nazywają konfabulacjami. Według nich konfabulacje pamięciowe są dość powszechne w codziennym życiu zdrowych ludzi i stanowią skutek uboczny normalnie funkcjonujących mechanizmów pamięciowych. Konfabulacje pamięciowe można podzielić na całkowicie fałszywe wspomnienia oraz częściowe zniekształcenia pamięci. Najczęściej występującymi są częściowe zniekształcenia pamięci, takie jak błędne pamiętanie czasu, miejsca lub treści danego zdarzenia.

French, Garry i Loftus wskazują, że nawet wspomnienia zaskakujące i społecznie znaczące – nazywane przez psychologów wspomnieniami błyskowymi (*flashbulb memories*) – podlegają zniekształceniom. Termin „wspomnienia błyskowe” zaproponowali psychologowie Roger Brown i James Kulik w 1977 r. do opisywania wspomnień zdarzeń zaskakujących i znaczących społecznie. Badacze ci użyli metafory lampy błyskowej w aparacie fotograficznym, która pozwala na uchwycenie większej liczby detali, aby pokazać, że również w mózgu coś „rozświeć się” podczas doświadczania znaczącego społecznie zdarzenia, a także pozwala uchwycić i zachować więcej szczegółów niż zwykle. Według French, Garry i Loftus wspomnienia błyskowe są intensywne, żywe i szczegółowe oraz nie podlegają dużym zniekształceniom.

French, Garry i Loftus omawiają badania, które wykazały, że nawet tego rodzaju wspomnienia są podatne na zafałszowania. W 1986 r. Ulric Neisser i Nicole Harsh (1992) w ciągu 24 godzin od eksplozji Challengeera przeprowadzili na dużej grupie osób badanie, które miało wykazać, czy wspomnienia zdarzeń społecznie znaczących podlegają zniekształceniom. Zadali oni badanym pytania dotyczące tego wydarzenia: jak i kiedy dowiedzieli się o wypadku, gdzie i z kim wówczas byli, co robili, jak czuli się, gdy się dowiedzieli o eksplozji oraz co robili potem. Dwa i pół roku później tym samym osobom zadano te same pytania. Okazało się, że średnio mniej niż połowa szczegółów z późniejszego kwestionariusza pasowała do tych w oryginalnym kwestionariuszu. Ponadto kwestionariusze żadnej z osób nie były w pełni spójne (Neisser i Harsh 1992). Według Browna i Kulika wspomnienia błyskowe są intensywne, żywe i szczegółowe oraz nie podlegają dużym zniekształceniom.

French, Garry i Loftus podają, że inne badania nad wspomnieniami błyskowymi są zgodne z wynikami badań, jakie uzyskali Neisser i Harsh. Dlaczego więc nawet wspomnienia błyskowe podlegają tak wielu zniekształceniom? Powołując się na Neissera (1982), badaczki podkreślają, że wspomnienia błyskowe mają duże znaczenie kulturowe, to znaczy opierają się na zdarzeniach, które często

są przywoływane przez inne źródła, takie jak media czy inni ludzie. Zniekształcenia pamięci powstają właśnie dlatego, że po zdarzeniu nastawieni jesteśmy na informacje z innych źródeł, które ostatecznie nakładają się na pierwotnie zapamiętane fakty. Dowiadujemy się o jakimś zdarzeniu z gazet czy telewizji bądź rozmawiając z innymi ludźmi, czytając listy lub e-maile. Czasami zapamiętujemy coś, wyobrażając sobie, co mogło się zdarzyć. Tego rodzaju źródła dostarczają nowych informacji, które włączane są do posiadanej już wiedzy, co powoduje uaktualnienie i rewizję wspomnień. Podczas takich rewizji dochodzi właśnie do zniekształceń pamięci.

Jak jednak badać zafałszowania pamięci? Psychologowie poznawczy stworzyli wzorzec badań nad tego rodzaju zafałszowaniem pamięci, który nazwali paradygmatem efektu dezinformacji (French i in. 2009, s. 39). W eksperymentach pokazuje się badanym jakieś zdarzenie, np. wypadek samochodowy bądź przestępstwo. Następnie podaje się im mylną informację o zdarzeniu. Informacja ta może być zawarta w pytaniu lub w formie pisemnej bądź ustnej narracji czy też w rozmowie z inną, podstawioną osobą. Okazuje się, że badani często twierdzą, iż pamiętają rzeczy, o których tylko usłyszeli lub przeczytali po obejrzeniu zdarzenia. Nie tylko więc nie pamiętają oni szczegółów zdarzenia, ale także często dodają nowe szczegóły do swoich raportów werbalnych.

Efekt dezinformacji może również polegać na dodawaniu detali do zapamiętanego zdarzenia, zmianie detali zdarzenia czy ich ewaluacji (French i in. 2009, ss. 39-40). Dodawanie detali dotyczy dodawania obiektów, zachowań bądź innych elementów. W jednym z badań pokazano badanym sekwencję slajdów, na których przedstawiony był wypadek motocyklowy. Następnie podzielono badanych na dwie grupy. Osobom z pierwszej grupy zadano pytanie, czy widziały „ten zбитy reflektor” (*the broken headlight*), osobom z drugiej – czy widziały „jakiś zбитy reflektor” (*a broken headlight*). Okazało się, że osoby z grupy, którą zapytano o „ten zбитy reflektor”, częściej twierdziły, iż przypominają sobie zбитy reflektor, choć w rzeczywistości go nie było (French i in. 2009, s. 39). Dezinformacja zawarta w pytaniu doprowadziła zatem do tego, że badani pamiętali coś, czego w rzeczywistości nie mogli zobaczyć. Tego rodzaju nieistniejące szczegóły mogą być zawarte w występującym po zdarzeniu pytaniu, narracji, dyskusji bądź wspólnym przypominaniu.

Badani mogą także zmieniać szczegóły konkretnych obiektów lub zdarzeń, a także błędnie oceniać szczegóły zdarzenia. W kolejnym badaniu poproszono osoby badane, by obejrzały nagranie wypadku samochodowego (Loftus i Palmer 1974). Następnie podzielono je na grupy. Członkom pierwszej grupy zadano pytanie: „Jaka była prędkość w momencie roztrzaskania się (*smash*) samochodów?”, a członków drugiej grupy zapytano: „Jaka była prędkość w momencie zderzenia się (*hit*) samochodów?”. Grupie kontrolnej nie zadawano pytania

o prędkość samochodów. Odpowiedzi badanych różniły się. Pierwsza grupa oszacowała prędkość na 66 km/godz., druga – na 55 km/godz. Po tygodniu badanym zadano pytanie: „Czy na miejscu wypadku widzieli jakieś rozbite szkło?”. W grupie pierwszej (której wcześniej zadano pytanie o roztrzaskanie się samochodów) odpowiedzi twierdzących było 32%, podczas gdy w grupie drugiej – 14%. W grupie kontrolnej odpowiedzi twierdzących było zaś 12%. Różnica między pierwszą a drugą grupą okazała się znacząca. Wniosek jest taki, że informacja zawarta w pytaniu może przyczynić się do szacowania szczegółów zdarzenia (takich jak prędkość samochodu), a także wpływać na dodawanie bądź zmienianie szczegółów zdarzenia.

Podatność na dezinformację zależy od dwóch czynników: jakości pamięci oraz przekąźnika informacji. Na jakość pamięci wpływają głównie okoliczności zapamiętywania informacji, takie jak: uwaga poświęcona wydarzeniu w trakcie zapamiętywania, poziom stresu osoby zapamiętującej, to, jak długo trwało wydarzenie, jakość zapamiętywanych informacji (np. jakie było oświetlenie), ale także okoliczności przypominania sobie zdarzeń, takie jak to, po jakim czasie zdarzenie zostało przypomniane. Im lepsza jakość pamięci, tym mniej osoby badane podatne są na dezinformację.

Drugim czynnikiem jest to, kto przekazuje błędną informację. Osobę eksperymentatora uważa się zwykle za wiarygodne źródło informacji, dlatego efekt dezinformacji w sytuacjach eksperymentalnych jest tak silny. Badania wykazały, że im wyższa wiarygodność i autorytet źródła, tym silniejszy jest efekt dezinformacji (French i in. 2009, ss. 41-43). W codziennych sytuacjach autorytetem może być np. policjant, który stawiając świadkowi pytania zawierające sugestie, może doprowadzić do zniekształceń pamięci. Wynika z tego, że zniekształcenia pamięci stanowią duże wyzwanie dla psychologów zajmujących się zeznaniami świadków.

Co ciekawe, jak zauważają French, Garry i Loftus, błędna informacja może być zawarta nie tylko w pytaniu, narracji czy wspólnym przypominaniu, ale także gdy ktoś w określony sposób opowiada przeżyte wydarzenia. Często wówczas wyolbrzymia lub minimalizuje, zmyśla i omija detale wydarzeń, o których opowiada innym ludziom.

Badania wykazały, że tego rodzaju świadome zmyślanie może prowadzić do fałszywych wspomnień. W jednym z takich badań Elizabeth Marsh i Barbara Tversky poprosiły osoby badane o ocenę tego, czy ich opowieści na temat przeszłości są zniekształcone (Marsh i Tversky 2004). Uczestnicy eksperymentu określili 61% swoich opowieści jako zniekształcone, czyli zawierające wyolbrzymienia, minimalizacje, dodatki i ominięcia. Jednak gdy zapytano ich, ile z ich opowieści jest niedokładnych, okazało się, że ocena była niższa – 42% opowieści zostało uznanych za nieściśle. Zdaniem badaczy świadczy to o tym,

że ludzie nie stawiają znaku równości między zniekształceniem a nieściłością. Opowiadana historia zmienia się w zależności od tego, jaki jest cel opowiadania (dokładność czy rozrywka) oraz komu opowiadana jest historia (rówieśnikowi czy eksperymentatorowi) (Marsh i Tversky 2004).

Jak wynika z badań, częściowe zniekształcenia pamięci pojawiają się dość często. Co jednak z całkowicie fałszywymi wspomnieniami?

Elizabeth Loftus badała formowanie się fałszywych wspomnień autobiograficznych. Postawiła sobie pytanie, czy można zaszczerpić osobom badanym treść fałszywych wspomnień. W jednym z badań prowadzonych wraz z Jacqueline E. Pickrell (1995) poprosiła starszych krewnych o opisanie zdarzeń z dzieciństwa osób badanych. Na tej podstawie przygotowała broszurki, w których zostały opisane trzy prawdziwe zdarzenia z życia osoby badanej i jedno fałszywe, które dotyczyło zagubienia się w centrum handlowym. Następnie badani zostali poproszeni o przypomnienie sobie szczegółów każdego ze zdarzeń. Uczestnicy badania przypominali sobie średnio 68% zdarzeń prawdziwych oraz 25% zdarzeń fałszywych.

Jeszcze inne badania potwierdziły, że ludzie przypominają sobie różnego rodzaju zdarzenia, które nie zdarzyły się w ich życiu, takie jak lot balonem, zrobienie psikusa nauczycielowi, bycie uratowanym przez ratownika, bycie zaatakowanym przez zwierzę (French i in. 2009). Oznacza to, że ludzie mogą fałszywie przypominać sobie rzeczy nie tylko typowe, z życia codziennego, ale także specyficzne i dziwne.

Całkowicie fałszywe wspomnienia mogą zatem dotyczyć zarówno zdarzeń typowych, codziennych, jak i zdarzeń fantastycznych, nieprawdopodobnych, które nie zdarzyły się w życiu danego człowieka. Do zdarzeń nieprawdopodobnych należą np. wspomnienia porwań przez kosmitów. French, Garry i Loftus powołują się na badanie George'a H. Gallupa Jr. i Franka Newporta (1991), które wykazało, że 27% amerykańskich respondentów wierzy, iż kosmici odwiedzili Ziemię, a 14% twierdziło, iż widzieli UFO. W innym badaniu 54% respondentów twierdziło, że wierzy, iż kosmici porywają ludzi, a 1% respondentów twierdziło, że byli porwani przez kosmitów (Chequers, Joseph i Diduca 1997, za: French i in. 2009). Co ciekawe, osoby przypominające sobie bycie porwanym przez UFO z detalami relacjonowały sytuacje z tego porwania (French i in. 2009, s. 34).

Jednym z najważniejszych czynników wpływających na wiarygodność wspomnień jest pamięć źródła informacji. Rozróżnienie na pamięć faktów i pamięć źródła zostało wprowadzone w 2002 r. przez Patricka Davidsona i Elizabeth Glisky (2002; Nęcka, Orzechowski i Szymura 2008). Pamięć faktów przechowywana jest w pamięci semantycznej bądź epizodycznej, wyróżnionych przez Endela Tulvinga w 1972 r. Pamięć źródła informacji to z kolei kontekst

nabywania nowej informacji, który jest o wiele ważniejszy w przypadku zapamiętywania epizodów z życia niż wiedzy ogólnej.

Błąd monitorowania źródła informacji to niezdolność przypisywania zdarzeniu odpowiedniego kontekstu, przede wszystkim czasu lub miejsca. Dzieje się tak dlatego, że pamięć źródła informacji kodowana jest inaczej niż pamięć faktów. Na przykład kiedy czterdziestoletni Jan próbuje sobie przypomnieć, na które urodziny w dzieciństwie dostał pierwszy aparat fotograficzny, musi sięgnąć pamięcią daleko wstecz, aby to ustalić. Oczywiście może wspomóc się dodatkową wiedzą, zaglądając do pierwszych albumów ze zdjęciami, i wywnioskować, iż pierwsze zdjęcia są z okresu, kiedy miał 9 lat. Jednak może się okazać, że zdjęcia z tego albumu robione były innym aparatem fotograficznym, a Jan dostał swój pierwszy aparat, kiedy miał 11 lat.

Przypominając sobie czas i miejsce zdarzeń, często posługujemy się dodatkową wiedzą, na podstawie której wnioskujemy o kontekście zdarzenia. Błędne monitorowanie źródła może jednak dotyczyć nie tylko zewnętrznych elementów zdarzenia, ale także mylenia wewnętrznych przemyśleń lub wyobrażeń z czymś, co mogłoby się wydarzyć. Przypuśćmy, że Jan całe dzieciństwo marzył o tym, żeby mieć piłkę nożną i spędzał wiele czasu, wyobrażając sobie, że gra własną piłką z kolegami z klasy. Przypuśćmy też, że Jan rzeczywiście grał często w piłkę nożną z kolegami z klasy, lecz piłka była własnością kogoś z kolegów. Po 30 latach Jan może przypisywać te wyobrażenia nie tyle wyobraźni, ile pamięci prawdziwych zdarzeń, twierdząc, iż pamięta, że miał własną piłkę nożną.

Błędy pamięci, takie jak opisane wyżej, nie mają być może dużego znaczenia społecznego. Błąd monitorowania źródła pamięci stanowi jednak poważny społeczny problem w przypadku zeznań świadków. Wymiar sprawiedliwości w dużej mierze opiera się na zeznaniach świadków – pamięć świadka może dostarczyć ważnych informacji na temat zdarzeń. Tymczasem badania z zakresu psychologii poznawczej dowodzą, że ludzka pamięć jest zawodna. Ludzie myślą twarze przestępców z twarzami widzianymi wcześniej w innym kontekście oraz ubrania, jakie przestępca miał na sobie, ponieważ wcześniej widzieli w telewizji lub gazecie kogoś ubranego w ten sposób (French i in. 2009; Loftus 1979; Schacter 2003).

Istotnym czynnikiem tworzenia się zniekształceń pamięci są błędy sklejanie śladów pamięciowych, czyli mylne wiązanie różnych elementów zdarzeń w jedną całość. Poszczególne elementy zdarzenia mogą być dobrze pamiętane, jednak zawieść może ich łączenie. Daniel Schacter w książce *Siedem grzechów pamięci* podaje wiele przykładów tego rodzaju błędnego sklejanie. Oto jeden z nich:

Poznawszy pana Wilsona i pana Alberta na spotkaniu roboczym, śmiało odpowiesz na pytanie współnika o wiceprezesa ich firmy: „pan Wilbert”. Zapamiętałeś

prawidłowo części składowe obydwu nazwisk, ale pomyłkowo „skleiłeś” je, tworząc nowe (Schacter 2003, s. 148).

Błędy wiązania mogą dotyczyć błędnego sklejanego nazwisk, nazw, zdań, obrazków czy cech twarzy. W jednym z eksperymentów badano ten efekt na przykładzie rozpoznawania widzianych wcześniej twarzy. Osoby badane rozpoznawały twarz jako wcześniej widzianą, pomimo iż nie była ona dotąd pokazywana, lecz stanowiła jedynie sklejenie cech twarzy prezentowanych wcześniej (rys. 1) (Schacter 2003, ss. 148-149).



Rysunek 1. Rodzaj błędnej atrybucji zwany błędem sklejanego śladów pamięciowych

Źródło: Schacter 2003, s. 149.

W procesach wiązania śladów pamięciowych ważną rolę odgrywa część mózgu zwana hipokampem. Okazuje się, że pacjenci z uszkodzeniami tego obszaru mózgu częściej niż osoby z grupy kontrolnej popełniają błędy sklejanego poznanych ostatnio słów i twarzy (Schacter 2003, ss. 148-149).

Pacjenci ci postrzegają słowa lub twarze jako całości, ale kiedy są testowani kilka sekund lub minut później, często przestawiają sylaby między słowami lub błędnie łączą poszczególne cechy twarzy. Zniszczony hipokamp przestał dostarczać mnemonicznego kleju, potrzebnego do utrzymania w pamięci wszystkich części słowa lub twarzy jako całości (Schacter 2003, s. 149).

Innym obszarem mózgu związanym z błędnym sklejeniem śladów pamięciowych są płaty czołowe. Odpowiadają one za wydobycie informacji z pamięci trwałej. Płaty czołowe monitorują efekty działania pamięci. W sytuacji gdy jakaś twarz wydaje się znajoma, płaty czołowe monitorują tę informację, próbując sprawdzić, dlaczego tak jest (Schacter 2003, s. 149). Uszkodzenia tego obszaru mózgu prowadzą do niezdolności monitorowania, czy dana informacja pochodzi z pamięci trwałej, czy też nie. Osoby z uszkodzeniami płatów czołowych popełniają zatem więcej błędów monitorowania źródła informacji (Schacter 2003, s. 150).

## 2. Jak odróżnić fałszywe wspomnienia od prawdziwych?

Czy można dokonać tego rozróżnienia w toku analizy samych wypowiedzi podmiotu? Czy istnieją czynniki, które pozwalają na odróżnienie jednych wspomnień od drugich? Psychologowie od dawna zadają sobie te pytania, lecz jak dotąd ich odpowiedź jest negatywna. Zdaniem French, Garry i Loftus (2009) nie ma możliwości jednoznacznego odróżnienia fałszywych wspomnień od prawdziwych.

Psychologiczne próby odróżnienia fałszywych wspomnień od prawdziwych można zaliczyć do trzech klas: badań ilościowych pamięci, badań jakości pamięci, badań fizjologicznych. Badania ilościowe polegają na sprawdzeniu długości wspomnień oraz detali zawartych we wspomnieniach. Loftus i Pickrell (1995) w badaniu dotyczącym zagubienia w centrum handlowym zauważyli, że opisy fałszywych wspomnień są znacznie krótsze niż opisy wspomnień prawdziwych. Okazało się jednak, iż opisy wspomnień prawdziwych były w niektórych przypadkach krótkie, dlatego uznano, że długość raportu werbalnego nie może być wyznacznikiem fałszywości wspomnienia. Inną próbą odróżnienia wspomnień fałszywych od prawdziwych jest badanie liczby detali w raportach werbalnych. Badania tego rodzaju dowodzą jednak, że nie ma znaczących różnic w liczbie szczegółów w obu rodzajach raportów werbalnych (za: French i in. 2009, s. 50).

Do jakościowych prób odróżnienia konfabulacji pamięciowych od prawdziwych wspomnień należą badania: subiektywnej jasności wspomnień, subiektywnego przeżycia na nowo, zawartości emocjonalnej wspomnień oraz pewności przypisywanej wspomnieniom.

Loftus i Pickrell (1995) badały subiektywną ocenę jasności wspomnień i stwierdziły, że ludzie oceniają wyrazistość wspomnień prawdziwych jako wyższą niż wspomnień fałszywych. W innych badaniach wykazano, iż ludzie oceniają wyżej jakość prawdziwego wspomnienia opartą na subiektywnym ponownym przeżyciu wspomnienia (French i in. 2009, s. 50). Subiektywna ocena wyrazistości wspomnienia nie może być jednak czynnikiem odróżniającym wspomnienia fałszywe od prawdziwych, gdyż w niektórych przypadkach wspomnienia fałszywe również są pod tymi względami oceniane wysoko. Innym jakościowym czynnikiem odróżniającym może być emocjonalna zawartość pamięci. Badania potwierdzają, że ludzie mają skłonność do oceniania własnych prawdziwych wspomnień jako bardziej emocjonalnych. Niemniej French, Garry i Loftus zauważają, iż same emocje nie mogą być wskaźnikiem autentyczności. Ponadto niektóre fałszywe wspomnienia również oceniane są jako posiadające dużą zawartość emocjonalną (French i in. 2009, s. 50).

Podobna sytuacja dotyczy rozważania pewności przypisywanej wspomnieniom. Niektóre badania pokazują, że ludzie mają tendencję do przypisywania wyższego stopnia pewności wspomnieniom prawdziwym niż fałszywym (French i in. 2009, s. 50). Inne badania wykazały, iż ludzie w takim samym stopniu przypisują duży stopień pewności wspomnieniom prawdziwym i fałszywym, dlatego pewność przypisywana przekonaniom nie może być czynnikiem odróżniającym prawdziwe raporty werbalne od fałszywych (Hyman i Billings 1998; Hyman i Pentland 1996). Te ostatnie badania dowodzą, że stopień pewności przypisywany przekonaniom konfabulacyjnym może być wysoki.

Czy istnieją inne sposoby odróżnienia fałszywych wspomnień od prawdziwych? Psychologowie postanowili sprawdzić, czy badania fizjologiczne pozwolą odróżnić wspomnienia fałszywe od prawdziwych. W tym celu zbadali fizjologiczne reakcje osób, które twierdziły, że zostały porwane przez kosmitów (McNally i in. 2004; French i in. 2009). Osoby te zostały poproszone o opisanie: (1) wspomnienia z porwania, (2) doświadczenia stresującego w ich życiu, (3) pozytywnego doświadczenia z życia, (4) neutralnego doświadczenia z życia. Wszystkie wypowiedzi były nagrywane. Następnie nagrania odtwarzano osobom badanym w celu zbadania ich reakcji fizjologicznych (m.in. przewodnictwa skóry, częstości tętna) w czasie słuchania opisów wspomnień. Na dalszym etapie eksperymentu badani zostali poproszeni o ocenę własnych reakcji emocjonalnych. Istotnym wynikiem badania było to, iż reakcje fizjologiczne podczas słuchania nagrań opisujących porwanie przez kosmitów były takie same jak reakcje pacjentów cierpiących na zespół stresu pourazowego. Wniosek jest więc taki, że nawet metody fizjologiczne nie pozwalają skutecznie odróżnić wspomnień fałszywych od prawdziwych.

Dlaczego fałszywe wspomnienia nie są odróżnialne od wspomnień prawdziwych? Według French, Garry i Loftus przyczyną może być to, że prawdziwe wspomnienia są tworzone w sposób podobny do wspomnień fałszywych (French i in. 2009, s. 51). Duże znaczenie w procesie powstawania wspomnień ma monitorowanie źródła informacji. Innym czynnikiem jest prawdopodobieństwo powstawania fałszywych wspomnień.

Badaniom dotyczącym fałszywych wspomnień można zarzucić to, że często próbuje się przedstawiać badanym zdarzenia codzienne, przyziemne, które z dużym prawdopodobieństwem mogły się zdarzyć w życiu każdego człowieka. Co jednak ze zdarzeniami mało prawdopodobnymi? W literaturze przedmiotu znane są przypadki nieprawdopodobnych fałszywych wspomnień, takich jak wspomniane porwania przez kosmitów. Osoby, które posiadają takie wspomnienia, muszą wierzyć, iż zdarzenia te są prawdopodobne, choć większość ludzi twierdzi, że tak nie jest. Istnieją trzy warunki, które muszą być spełnione, aby ktoś mógł wytworzyć fałszywe wspomnienie: (1) musi on być przekonany, że

dane zdarzenie z dużym prawdopodobieństwem mogło mieć miejsce; (2) musi zbudować w sobie przeświadczenie, że wydarzyło się ono jemu; (3) musi popełnić błąd monitorowania źródła, czyli niewłaściwie twierdzić, iż szczegóły zdarzenia pochodzą z prawdziwego doświadczenia (Mazzoni, Loftus i Kirsch 2001; French i in. 2009).

### 3. Podziały błędów pamięci a trzy tryby konfabulowania

Wśród wielu podziałów błędów ludzkiej pamięci wartą uwagi jest klasyfikacja Daniela Schactera wyróżniająca siedem błędów pamięci. Schacter uznał błędy pamięci nie tyle za wady, ile raczej za zalety pamięci bądź skutki uboczne poprawnych i użytecznych mechanizmów poznawczych. Błędy pamięci podzielił na błędy pomijania i błędy przekształcania. Do błędów pomijania należą: nietrwałość pamięci, roztargnienie i blokowanie, a do błędów przekształcania: błędna atrybucja, sugestywność pamięci, tendencyjność pamięci oraz uporczywość pamięci (Schacter 2003). Druga grupa błędów będzie dalej rozpatrywana jako przykłady konfabulacji pamięciowych w zdrowych systemach poznawczych.

Podział błędów pamięci dokonany przez Schactera można odnieść do przedstawionego wcześniej podziału konfabulacji ze względu na typ procesu konfabulowania, w tym konfabulację wypełniającą, scalającą oraz addytywną. Rozróżnienie to zostało wprowadzone z uwagi na trzy różne stany wyjściowe: brak danych w systemie wiedzy, szczątkowe dane w systemie wiedzy i w miarę pełne dane w systemie wiedzy. Trzy tryby konfabulowania odpowiadające rodzajom konfabulacji to: wypełnianie konfabulacyjne, scalanie konfabulacyjne oraz dodawanie bądź uzmiennianie konfabulacyjne. Zarówno konfabulacja wypełniająca, jak i scalająca czy addytywna mogą występować w poprawnie funkcjonujących systemach poznawczych w różnych domenach wiedzy.

W jaki sposób ten podział konfabulacji odnosi się do błędów pamięci Schactera? W moim przekonaniu konfabulacją wypełniającą jest całkowita fałszywa pamięć w sytuacji wypełniania luk w systemie poznawczym przy braku dostępu do informacji. Stanem wyjściowym tej konfabulacji jest całkowity brak informacji w pamięci, a trybem konfabulowania – wypełnianie konfabulacyjne luki w pamięci. Powodem takiego wypełniania konfabulacyjnego może być efekt dezinformacji (French i in. 2009; Schacter 2003).

Konfabulacja scalająca opiera się z kolei na błędzie scalania śladów pamięciowych. Występuje on wówczas, gdy np. dwa różne ślady pamięciowe scalane są w jeden. Punktem wyjścia scalającej konfabulacji pamięciowej są szczątkowe dane pamięciowe, a trybem konfabulowania – scalanie konfabulacyjne. Może to dotyczyć sklejanie słów, zdań, elementów zdarzeń, obrazów czy też twarzy

Tabela 7. Siedem błędów pamięci w ujęciu Daniela Schactera

Cecha pamięci	Rodzaj błędu	Efekt
Nietrwałość Roztargnienie Blokowanie	błędy pomijania	informacja nie zostaje zakodowana bądź wydobyta z pamięci
Błędna atrybucja Sugestywność Tendencyjność	błędy przekształcania	nieprawidłowa forma pamięci; zafałszowanie pamięci
Uporczywość		wyolbrzymienie i mimowolne narzucanie się niechcianych wspomnień

Źródło: opracowanie własne na podstawie: Schacter 2003.

(Schacter 2003, s. 148). Konfabulacje scalające to także błędne atrybucje źródła informacji, np. przypisanie znajomej twarzy do nieprawidłowego kontekstu. Cechą charakterystyczną tego błędu jest to, że istotna treść pamięci jest prawidłowa, natomiast błędny jest kontekst, w którym tę treść sytuujemy.

Konfabulacja addytywna to uporczywość pamięci, w której pamięć przeszłych zdarzeń może być prawidłowa, natomiast mogą pojawiać się naddatki w postaci wyolbrzymionego emocjonalnego zabarwienia przeszłych zdarzeń. Uporczywość pamięci to niechciane pojawianie się wspomnień o negatywnym bądź pozytywnym zabarwieniu emocjonalnym. Takie wspomnienia często są zniekształcone przez wyolbrzymienie cech przeszłych zdarzeń. Tego rodzaju zjawisko jest domeną poprawnie funkcjonujących systemów poznawczych, choć w skrajnych przypadkach prowadzi do zaburzeń w postaci nerwicy pourazowej i depresji (Schacter 2003, ss. 266-267). Tabela 8 przedstawia zestawienie omówionych konfabulacji pamięciowych.

Tabela 8. Trzy rodzaje konfabulacji pamięciowych w zdrowych systemach poznawczych

Konfabulacja wypełniająca	Konfabulacja scalająca	Konfabulacja addytywna
– fałszywa treść pamięci – sugestywność pamięci	– błąd sklejanego śladów pamięciowych – błędna atrybucja źródła	– uporczywość pamięci

Źródło: opracowanie własne.

Podane przykłady trzech trybów konfabulowania w odniesieniu do konfabulacji pamięciowych w zdrowych systemach poznawczych są tylko wstępem do szerszego, epistemicznego ujęcia zagadnienia konfabulacji. Konfabulacje

pamięciowe nie wyczerpują bowiem wszystkich przykładów konfabulowania. Aby epistemiczna teoria konfabulacji miała szersze odniesienia, potrzebna będzie analiza innych przykładów konfabulacji, jakie zachodzą w poprawnie funkcjonujących systemach poznawczych, np. konfabulacja na temat percepcji ciała, percepcji stanów umysłów innych osób czy też uzasadniania własnych przekonań i działań.

## Konfabulacje socjocentryczne a teoria Mind-Reading

### 1. Klasyczna teoria Mind-Reading

William Hirstein zastosował teorię Mind-Reading do analizy konfabulacji pacjentów w różnych zespołach neuropsychologicznych. Zauważył on, że teoria ta może być pomocna w wyjaśnieniu dużego stopnia pewności przypisywanego przez pacjentów przekonaniom konfabulacyjnym. Pomimo wyraźnych dowodów przeczących ich przekonaniom oraz oczywistego niedowierzania, wskazówek czy korekt ze strony słuchaczy osoby z zespołami, w których występują konfabulacje, nie odczuwają dysonansu poznawczego, w związku z czym nie korygują swoich przekonań i nie uzgadniają ich z istniejącymi dowodami czy przekonaniem innych. Jest to cecha charakterystyczna wszystkich rodzajów konfabulacji klinicznych. Stopień subiektywnej pewności przypisywany przekonaniom jest zazwyczaj bardzo wysoki (Hirstein 2005, s. 101).

Jednym z pierwszych teoretyków teorii Mind-Reading, czyli „czytania umysłów”, był psycholog Simon Baron-Cohen, który w książce *Mindblindness. An Essay on Autism and Theory of Mind* (1995) zaproponował rozbudowaną koncepcję mechanizmu pozwalającego na identyfikację różnego typu podmiotów w otoczeniu. Dyspozycja do detekcji innych istot żywych w otoczeniu jest – zdaniem tego badacza i innych zorientowanych ewolucyjnie filozofów, psychologów i kognitywistów – jedną z podstawowych zdolności przystosowania się organizmów do otoczenia. Ma ona kluczowe znaczenie w rozpoznawaniu przez organizm tego, czy ma on do czynienia z obiektem nieożywionym, czy z istotą żywą posiadającą umysł, oraz w ustaleniu, czy rozpoznana istota żywa jest pomocnym przyjacielem, atrakcyjnym partnerem, niebezpiecznym przeciwnikiem, czy ofiarą. Z uwagi na to, że otoczenie społeczne jest częścią otoczenia człowieka, a także innych naczelnych oraz innych zwierząt, dyspozycja ta musiała pojawić się w toku ewolucji jako jeden z ważniejszych wymogów adaptacyjnych.

Teoria „czytania umysłów” Barona-Cohena tłumaczy, jak w toku rozwoju ontogenetycznego człowieka pojawiają się kolejne zdolności rozpoznawania stanów mentalnych istot żywych. Na całkowitą zdolność rozpoznawania stanów mentalnych innych osób składają się cztery moduły: detektor intencjonalności (ID – *Intentionality Detector*), detektor kierunku patrzenia (EED – *The Eye-Direction Detector*), detektor współlniania pola uwagi (SAM – *The Shared Attention Mechanism*) oraz mechanizm teorii umysłu (ToMM – *The Theory of Mind Mechanism*). Pozwalają one na interpretowanie zachowań innych osób jako celowych działań dzięki wcześniejszemu rozpoznaniu ich intencji. Dwa pierwsze detektory występują już u zwierząt, jednak osiągnięciem gatunku ludzkiego, które w pełni pozwala na rozpoznawanie innych jako istot posiadających własny punkt widzenia oraz własne przekonania na temat świata, jest zdolność współlniania uwagi oraz mechanizm teorii umysłu.

Każdy z detektorów odzwierciedla inną umiejętność w ramach zdolności „czytania umysłów”. Detektor intencjonalności (ID) służy do rozpoznawania celów działań innych osób lub zwierząt. Jest to mechanizm percepcyjny dokonujący interpretacji bodźców ruchowych w kategoriach prymitywnych wolicjonalnych stanów umysłu. Nadaje on sens zachowaniom ruchowym wszystkich zwierząt, takim jak: zbliżanie się, unikanie, chwytanie. Pozwala na przypisywanie innym umysłom celów i pragnień. ID działa nie tylko za pośrednictwem wzroku, ale także dotyku i słuchu (są to według Barona-Cohena modalności podstawowe). Za każdym razem gdy pojawiają się dane percepcyjne wskazujące, że organizm ma do czynienia z podmiotem działającym (agentem), detektor intencjonalności jest aktywowany.

Detektor kierunku patrzenia (EDD) działa wyłącznie za pomocą zmysłu wzroku i jest częścią układu wzrokowego. EDD pełni trzy funkcje: pozwala na detekcję oczu lub bodźców podobnych do oczu, sprawdza, na jaki obiekt skierowane są oczy, oraz wnioskuje na podstawie własnego doświadczenia, że jeśli oczy innego organizmu skierowane są na coś, to organizm ten widzi dany obiekt.

ID oraz EDD potrafią dokonywać wyłącznie reprezentacji diadycznych. ID reprezentuje takie stany rzeczy, jak: [Agent-chce-pożywienia], [Agent-ma cel-otworzyć drzwi], natomiast EDD takie jak: [Agent-widzi-mnie], [Agent-patrzy-na mnie] (Baron-Cohen 2009, s. 157). Reprezentacje te zwane są diadycznymi, gdyż reprezentują relację umysłową pomiędzy dwoma obiektami, takimi jak „agent i obiekt” lub „agent i »ja«”. Aby móc komunikować się z innymi na temat rzeczywistości, potrzebne są jednak reprezentacje triadyczne, czyli takie, które pozwalają na reprezentowanie tego, iż „ja” i agent patrzymy na tę samą rzecz. Na tego rodzaju zabieg pozwala mechanizm współlniania uwagi (SAM). Reprezentuje on takie stany rzeczy, jak: [Agent-widzi-(ja-widzę)-autobus].

SAM buduje reprezentacje triadyczne dopiero wówczas, gdy otrzyma informacje na temat stanu percepcyjnego innego agenta. Mechanizm uwspólniania uwagi działa w ten sposób, że porównuje stan percepcyjny agenta z bieżącym stanem percepcyjnym „ja”.

Działa jak komparator, który łączy w reprezentację tridyczną dwie reprezentacje diadyczne – na temat stanu percepcyjnego innej osoby oraz stanu percepcyjnego Ja. Sprawia to, że SAM oblicza, że oboje, ty i ja, widzimy to samo, czujemy ten sam zapach, dotykamy tej samej rzeczy, smakujemy to samo lub słyszymy to samo (Baron-Cohen 1995/2009, s. 159).

SAM może korzystać z dowolnej modalności zmysłowej, jednak uprzywilejowanym związkiem jest relacja między EDD i SAM. To bowiem w zakresie percepcji wzrokowej najczęściej zachodzi uwspólnianie uwagi.

Mechanizm teorii umysłu (ToMM) pozwala z kolei na reprezentowanie stanów epistemicznych drugiej osoby, takich jak: myślenie, udawanie, wiedza, przekonania, wyobrażanie sobie czy oszukiwanie. Przetwarza on reprezentacje postaw propozycyjalnych, takich jak: [Agent-postawa propozycyjalna-„twierdzenie”], np. [Jan-sądzi-„pada śnieg”], [Maria-jest przekonana-„Pluton jest planetą”]. ToMM pozwala na rozumienie, że inni ludzie mogą mieć inne przekonania niż my.

## 2. Teoria Mind-Reading a konfabulacje

Hirstein uważa, że niezdolność pacjentów konfabulujących do korekty przekonań wynika z zaburzeń w „systemie czytania umysłów innych osób” (Hirstein 2005, s. 102). Patologiczne zachowania konfabulacyjne są rezultatem niezdolności hamowania i korygowania przekonań tworzonych w wyniku niesprawnych procesów poznawczych. Do powstania konfabulacji patologicznych nie wystarczy zatem dysfunkcja bądź chwilowa niesprawność systemu poznawczego, która prowadzi do powstania błędnie ugruntowanego przekonania. W sytuacjach normalnych taka nieprawidłowość zostałaby wyhamowana bądź po wyartykułowaniu treści przekonania wychwycona i poddana korekcie dzięki informacji zwrotnej. Co istotne, osoby cierpiące na konfabulacje kliniczne upierają się przy swoim stanowisku, bez poczucia zażenowania czy zdenerwowania tym, że inni próbują dokonać korekty ich przekonań.

Punktem wyjścia dla Hirsteina jest przekonanie, że pacjenci z zespołami neuropsychologicznymi nie tylko nie zdają sobie sprawy, iż ich twierdzenia są błędnie ugruntowane, ale też nie zdają sobie sprawy, że rozmówcy uznają ich twierdzenia za błędnie ugruntowane. Pacjenci konfabulujący mają kłopoty ze społecznymi interakcjami z innymi. Warto przypomnieć, że konfabulant to

nie kłamca, gdyż skuteczne kłamanie wymaga wiedzy na temat stanów umysłu osoby okłamywanej. Hirstein zauważa, że są dwa powody tego, iż konfabulanci nie są kłamcami: (1) mają oni przekonania wyrażane przez ich konfabulacje i działają często na ich podstawie, (2) są nieczuli na niedowierzanie ze strony słuchaczy (Hirstein 2005, s. 103).

Hirstein uważa, że ludzki system poznawczy posiada wiele modułów rozumienia innych osób. Po pierwsze, potrafi zrozumieć percepcję innych osób, ich działanie w świecie z innymi obiektami i ludźmi. W zakres tego rozumienia wchodzi uwspólnianie uwagi z innymi, czyli zdolność kierowania uwagi na obiekty, na które patrzą inni. Po drugie, ludzki system poznawczy posiada zdolność rozumienia działań innych. Po trzecie, posiada zdolność rozumienia przekonań i rozpoznawania emocji innych osób.

Hirstein powołuje się na Leslie Brothers (1995), która wyróżnia zimne i ciepłe podejście w ramach teorii Mind-Reading. Zimne podejście koncentruje się na tym, w jaki sposób poznajemy percepcje oraz prawdziwe i fałszywe przekonania innych osób. Ciepłe podejście dotyczy tego, w jaki sposób poznajemy emocje innych. Hirstein zastanawia się, w jaki sposób działa emocjonalny Mind-Reading. Jego zdaniem jedną z ról kory oczodołowo-czołowej może być wytwarzanie społecznych i symulowanych emocji (Hirstein 2005, s. 113). Emocje społeczne to emocje wyższego rzędu, takie jak: poczucie winy, zakłopotanie, wstyd, współczucie (Damasio 2005). Emocje symulowane reprezentują natomiast emocje innych osób, z którymi wchodzimy w interakcje (Hirstein 2005).

Warto zwrócić uwagę na dwie kwestie dotyczące teorii Mind-Reading w omawianym kontekście: po pierwsze, na kwestię związku tej teorii z konfabulacjami dotyczącymi stanów umysłów innych osób; po drugie, na kwestię przypisywania dużej pewności własnym przekonaniom konfabulacyjnym (bez względu na ich rodzaj) przy równoczesnej niezdolności bądź nieprawidłowości „czytania umysłów” innych. W pierwszym przypadku niezdolność „czytania umysłów” ma istotne znaczenie przy tworzeniu przekonania konfabulacyjnego na temat umysłów innych. W drugim przypadku niezdolność „czytania umysłów” innych ma znaczenie przy testowaniu i korekcie przekonań własnych.

### 3. Zespoły błędnej identyfikacji a teoria Mind-Reading według Williama Hirsteina

William Hirstein (2005) dokonuje analizy syndromów błędnej identyfikacji w kontekście teorii Mind-Reading. Ujęcie to ma go doprowadzić do bardziej ogólnego stwierdzenia, że pacjenci ze wszystkimi zespołami konfabulacyjnymi mają deficyt „czytania umysłów” innych.

Zespoły błędnej identyfikacji to np. zespół Capgrasa, zespół Fregoliego czy zespół Cotarda (Hirstein 2005, s. 116). Błędna identyfikacja w tych zespołach dotyczy często umysłów innych osób, dlatego można je nazwać klinicznymi konfabulacjami socjocentrycznymi. Niemniej zdarzają się przypadki błędnej identyfikacji samego siebie i wówczas mamy do czynienia z klinicznymi konfabulacjami egocentrycznymi.

Hirstein uważa, że w każdym z zespołów błędnej identyfikacji można dostrzec cechy niezdolności „czytania umysłów” innych. Na przykład pacjenci z zespołem Capgrasa twierdzą, że osoby bliskie są w rzeczywistości obce bądź są podstawionymi dublerami. Mimo że wyglądają tak samo jak ojciec czy brat, to wydają się być kimś obcym, co prowadzi pacjenta do przekonania, że pod jego krewnego podszywa się oszust. Dotychczas uważano, iż przyczyną tego zaburzenia jest dysocjacja między rozpoznaniem kognitywnym a afektywnym<sup>1</sup>.

W jaki sposób ludzki mózg rozpoznaje inne osoby w otoczeniu, a w szczególności osoby bliskie? Aby mogło dojść do rozpoznania osoby bliskiej, potrzebne są dwa rodzaje reprezentacji: zewnętrzna i wewnętrzna. Reprezentacja zewnętrzna to wygląd osoby, jej sposób poruszania się, głos, śmiech czy zapach. Reprezentacja wewnętrzna to przypisanie umysłu drugiej osobie, czyli charakterystycznych dla niej emocji, przekonań, nastrojów. Rozpoznawanie bliskiej osoby różni się od patrzenia na osobę nieznaną tym, że w pierwszym przypadku mamy dwa rodzaje reprezentacji – zewnętrzną i wewnętrzną, natomiast w drugim – tylko reprezentację zewnętrzną (Hirstein 2005, s. 123).

Relacja między obiema reprezentacjami jest złożona i może szwankować na wiele sposobów. Według Hirsteina osoby z zespołem Capgrasa posiadają reprezentację zewnętrzną osoby bliskiej, podczas gdy ich reprezentacja wewnętrzna jest zmieniona, dlatego też pacjent przypisuje swojemu krewnemu umysł innej, obcej osoby. Nie jest to więc ta sama reprezentacja krewnego, jakiej zwykł używać przed chorobą. Filozof analizuje także inne przypadki błędnej identyfikacji, takie jak zespół Fregoliego<sup>2</sup>. Polega on na tym, że pacjent twierdzi, iż różni nieznanemu mu ludzie są tą samą, znaną mu osobą. Mimo że osoby rozpoznane wyglądają inaczej niż osoba bliska, pacjent upiera się, iż jest to ta sama osoba. Mamy tu do czynienia z sytuacją, w której ta sama wewnętrzna reprezentacja aktywowana jest przy reprezentacjach zewnętrznych innych, często obcych osób (Hirstein 2005, s. 123).

---

<sup>1</sup> Twarze osób bliskich są nadal rozpoznawane, ale nie ma reakcji emocjonalnych, np. widocznych zmian oporności skóry (Ramachandran i Blakeslee 1998, s. 167). Jedną z hipotez głosi, że uszkodzone zostały obszary odpowiedzialne za emocje towarzyszące widzianej twarzy. Podejście to jest krytykowane przez Ramachandrana i Hirsteina (2009).

<sup>2</sup> Nazwa zespołu Fregoliego pochodzi od nazwiska włoskiego aktora Leopolda Fregoliego, który słynął z tego, że potrafił wcielić się w różnych ludzi.

Innym przykładem jest zespół Cotarda, który polega na tym, że pacjent twierdzi, iż on sam bądź inni nie żyją lub są w środku puści. W tej sytuacji reprezentacja zewnętrzna (zewnętrzny wygląd osoby) jest zachowana, natomiast reprezentacja wewnętrzna przestaje działać. To dlatego pacjent twierdzi, że widzi pustkę w drugim człowieku bądź w sobie samym (Hirstein 2005, s. 127).

Uszkodzenie systemu „czytania umysłów” innych zdaniem Hirsteina polega w szczególności na niezdolności tworzenia symulowanych emocji:

Może być tak, że zarówno emocjonalne reakcje pacjenta, jak i emocje, które przypisuje innym, zmieniły się. Jeśli to prawda, że używamy własnego systemu emocjonalnego w trybie symulacji, aby zrozumieć emocje innych, symultaniczna zmiana w obu tych sferach jest mniej zaskakująca. Jeden z argumentów za podejściem teorii umysłu do zespołów błędnej identyfikacji jest taki, że uszkodzenie często ma duży wpływ na pacjenta teorię umysłu (Hirstein 2005, s. 126).

Jakie obszary mózgu są istotne dla systemu „czytania umysłów” innych osób? Hirstein powołuje się w tej materii na Barona-Cohena, wymieniając korę oczodołowo-czołową, górną bruzdę skroniową i ciało migdałowe (Hirstein 2005, s. 106). Ponadto przywołuje badania, które wskazują, iż pacjenci z uszkodzeniami tych obszarów mają problemy z „czytaniem umysłów” innych osób.

W jednym z badań okazało się, iż osoby z uszkodzeniami kory oczodołowo-czołowej zachowywały się podobnie do pacjentów z zespołem Aspergera w testach na rozumienie społecznych relacji, takich jak rozpoznanie popełnionego *faux pas* (Stone i in. 1998). Inne badanie wykazało, że z obustronnym uszkodzeniem środkowej kory oczodołowej mają osłabione zdolności rozpoznawania przekonań u innych, np. nie rozpoznają, że druga osoba chce ich oszukać (Struss i in. 2001). Pacjenci z uszkodzoną brzuszną korą czołową nie potrafili zidentyfikować emocji na twarzach i w głosach innych osób (Horniak i in. 1996). Przy uszkodzeniach brzuszno-przyśrodkowej kory przedczołowej pacjenci nie przechodzili testu nazywanego The Reading the Mind in the Eyes Test, podczas którego pacjenci oglądają fotografie twarzy innych i mają ocenić np. czy osoba na fotografii patrzy im w oczy, czy też nie (Gregory i in. 2002).

Górna bruzda skroniowa to obszar odpowiadający za analizę ruchów mięśni twarzy, w szczególności tych, które dotyczą ekspresji emocji (Haxby, Hoffman i Gobbini 2000). Niektórzy uważają, że przez górną bruzdę skroniową biegnie szlak odpowiedzialny za ukryte rozpoznanie twarzy (rozpoznanie emocji, kierunku patrzenia). Ciało migdałowe ma neurony wrażliwe na kierunek patrzenia, a jego dysfunkcja może być jedną z przyczyn autyzmu (Baron-Cohen i in. 1997). Stymulacja ciała migdałowego powoduje przywoływanie wspomnień związanych z emocjami, poczuciem znajomości (Gloor 1986).

Po bliższym przyjrzeniu się propozycji Hirsteina pojawia się zasadnicze pytanie, czy trudności z „czytaniem umysłów” innych osób dotyczą wyłącznie niektórych ludzi, takich jak krewni. Jak wiadomo, pacjenci z zespołem Capgrasa nie rozpoznają umysłu wyłącznie bliskiej osoby. Jak zatem wyjaśnić ten syndrom w ramach klasycznie rozumianej teorii Mind-Reading? W koncepcji Barona-Cohena uszkodzenie poszczególnych detektorów systemu „czytania umysłów” prowadzi do autyzmu. Osoby z autyzmem nie potrafią „czytać umysłów” innych, jednak nie jest to ułomność wybiórcza, lecz globalna. Nie potrafią często w ogóle zidentyfikować umysłów w otoczeniu. W przypadku zespołów błędnej identyfikacji pacjenci potrafią przypisywać umysły innym ludziom, jednak myślą się co do tego, czyj to jest umysł. Mają zatem problem z reprezentacją wewnętrzną (Hirstein 2005).

Zaproponowany przez Hirsteina podział na reprezentację zewnętrzną i wewnętrzną także prowadzi do trudności. Zasadne jest pytanie, czy rzeczywiście jest tak, że kiedy widzimy twarz człowieka po raz pierwszy, posiadamy tylko jego reprezentację zewnętrzną. W świetle teorii Barona-Cohena wydaje się to wątpliwe. Ludzki umysł posiada cztery detektory, z czego dwa pierwsze – detektor intencjonalności i detektor kierunku patrzenia – służą do identyfikacji w otoczeniu systemów umysłopodobnych. Jest więc raczej tak, że przyglądając się nieznanej nam osobie, nie tylko zapoznajemy się z jej wyglądem, lecz także tworzymy reprezentację wewnętrzną jej umysłu.

Koncepcja reprezentacji pojawia się w artykule Williama Hirsteina i Vilyanura S. Ramachandrana (2009). Badacze najpierw krytykują dwa inne podejścia do zespołów błędnej identyfikacji: podejście odwołujące się do emocji oraz podejście odwołujące się do personalnego pokrewieństwa. Następnie wprowadzają nowe reprezentacjonistyczne podejście nawiązujące do teorii Mind-Reading.

W podejściu odwołującym się do emocji zespół Capgrasa wyjaśnia się za pomocą twierdzenia o braku emocjonalnej reakcji pacjenta na widok znajomej osoby, co prowadzi go do przekonania, że ma do czynienia z oszustem. Podejście to nawiązuje do stanowiska samego Capgrasa, od którego pochodzi nazwa tego zespołu neuropsychologicznego. Uważał on, że urojenie to ma niewiele wspólnego z procesami sensorycznymi, a więcej z procesami emocjonalnymi (za: Hirstein i Ramachandran 2009, s. 114). Wielu współczesnych badaczy odwołuje się do tego podejścia (Ellis i Young 1990; Hirstein i Ramachandran 2009). Uważają oni, że w zespole Capgrasa dochodzi do dyskoneksji połączeń obszarów emocjonalnych z obszarami rozpoznawania twarzy, w związku z czym pacjenci nie czują żadnych emocji wobec bliskich im osób i konfabulują, aby ten brak emocji uzasadnić. Podejście to jednak, jak zauważają Hirstein i Ramachandran (2009), prowadzi do dwóch ważnych pytań: Dlaczego zespół

Capgrasa ogranicza się do konfabulacji na temat krewnych? Dlaczego pacjenci nie twierdzą po prostu, że widzą swoich krewnych, tylko nie czują do nich emocji, zamiast głosić, że bliscy są oszustami?

Jeśli chodzi o pierwsze pytanie, to można na nie odpowiedzieć w ten sposób, że emocjonalne reakcje na twarze bliskich krewnych są większe niż reakcje na twarze osób nieznanymi. Nieobecność takiej emocjonalnej reakcji prowadzi pacjenta do konfabulacji, iż krewny jest w rzeczywistości oszustem (Hirstein i Ramachandran 1997).

Odpowiedź na drugie pytanie wydaje się jednak podważać zasadność podejścia emocjonalnego. Dlaczego bowiem pacjent nie twierdzi: „Wiem, że wygląda jak mój ojciec, lecz nic do niego nie czuję”, a zamiast tego tworzy bardzo rozbudowane i niewiarygodne konfabulacje o oszuście, który podszywa się pod krewnego? Jednym z wyjaśnień tego zjawiska, zdaniem Hirsteina i Ramachandrana, może być to, iż jakieś dodatkowe uszkodzenia w prawym płacie czołowym powodują wytworzenie tak ekstremalnych konfabulacji. Przy okazji wyjaśniania asomatognozji Ramachandran (1995) przedstawił teorię, że lewa półkula mózgu tworzy spójne obrazy świata, podczas gdy prawa półkula odpowiedzialna jest za sprawdzanie owej spójności. Zdaniem badacza teoria ta wyjaśnia, dlaczego pacjenci z uszkodzeniami prawej półkuli tak często tworzą mało prawdopodobne konfabulacje. To bowiem lewa półkula tworzy nieprawdopodobne historie, a prawa, która zwykle jest cenzorem, w wyniku uszkodzenia nie wychwytuje nieprawidłowości.

Podejście odwołujące się do pokrewieństwa personalnego, autorstwa Todda E. Feinberga i Davida M. Roane'a (1997), głosi, że zespół Capgrasa jest podobny do asomatognozji pod względem utraty poczucia pokrewieństwa personalnego – w pierwszym przypadku z krewnym, w drugim z częściami własnego ciała. Feinberg i Roane zestawiają także oba zespoły z zespołem Fregoliego, który ich zdaniem różni się od poprzednich tym, że pacjenci widzą pokrewieństwo personalne tam, gdzie go nie ma. Zdaniem Feinberga wyjaśnienie w kategoriach pokrewieństwa personalnego może być dobrym wyznacznikiem tworzenia taksonomii zespołów błędnej identyfikacji.

Zdaniem Hirsteina i Ramachandrana (2009) żadne z wyżej wymienionych podejść nie wyjaśnia dobrze zespołów błędnej identyfikacji. Badacze ci proponują więc podejście konkurencyjne zwane reprezentacjonistycznym. Twierdzą, iż bardzo wysoki procent reprezentacji przechowywanych przez nasz mózg dotyczy rzeczy, które są powiązane z nami personalnie. Nasze mózgi reprezentują innych ludzi zarówno allocentrycznie, jak i egocentrycznie. Allocentryczne reprezentacje innych ludzi to „niezależne od punktu widzenia reprezentacje ich zewnętrznych cech cielesnych” (Hirstein i Ramachandran 2009, s. 116), czego przykładem może być reprezentowanie ludzkich twarzy. Z kolei egocentryczne

reprezentacje innych „koduują pozycję rzeczy i przestrzeni relatywnie do centralnego »ego«” (Hirstein i Ramachandran 2009, s. 116). Zawierają więc one wewnętrzny punkt widzenia.

Egocentryczny system reprezentacyjny posiada dwa tryby: tryb-ja (*self-mode*) oraz tryb-innych (*other-mode*):

W trybie-ja ten system reprezentuje mój umysł usytuowany w moim ciele, które jest usytuowane w moim środowisku. Kiedy system funkcjonuje w trybie-innych, reprezentuje innych jako umysły usytuowane w ciałach, usytuowanych w środowisku. [...] Kiedy reprezentuję ciebie z mojej perspektywy, reprezentuję ciebie allocentrycznie. Kiedy reprezentuję ciebie z twojej perspektywy, reprezentuję ciebie egocentrycznie. [...] W trybie-ja to ja jestem ego w centrum tego systemu, w trybie-innych system egocentryczny reprezentuje lub symuluje egocentryczny system innej osoby (Hirstein i Ramachandran 2009, ss. 116-117).

Badacze sugerują, że egocentryczny system „czytania umysłów” innych jest jednym z wielu mózgowych procesów Mind-Reading.

Zdaniem Hirsteina i Ramachandrana percypowanie osobowości innej osoby wiąże się z reprezentowaniem jej emocji, nastrojów, charakteru<sup>3</sup>. Kiedy dobrze znamy jakąś osobę, przypisujemy jej zindywidualizowaną egocentryczną reprezentację jej umysłu. Kiedy nie znamy drugiej osoby i dopiero ją poznajemy, przypisujemy jej zgeneralizowaną egocentryczną reprezentację umysłu. Taka zgeneralizowana reprezentacja jest swego rodzaju stereotypem charakteru, jaki przypisujemy nieznanemu dobrze osobie. Te dwa rodzaje reprezentacji egocentrycznych wskazują na to, w jaki sposób rozpoznajemy znanych nam ludzi w odróżnieniu od nieznanymi.

Co dzieje się zatem w zespołach błędnej identyfikacji z wyróżnionymi przez badaczy reprezentacjami? Według nich zespoły błędnej identyfikacji powstają wówczas, gdy któreś z egocentrycznych reprezentacji umysłu przestają działać tak jak powinny (Hirstein i Ramachandran 2009). Na przykład zespół Capgrasa powstaje, gdy egocentryczne reprezentacje jakiejś osoby (krewnego) są uszkodzone bądź niedostępne, a w zamian są podstawione inne niewłaściwe reprezentacje, najprawdopodobniej reprezentacje zgeneralizowane. Pacjent odbiera swojego krewnego jako kogoś, kto wygląda tak samo, lecz ma inny umysł – inne emocje, nastroje, charakter. To prowadzi go do przekonania, że jest on oszustem.

Według Ramachandrana i Hirsteina tylko ujęcie reprezentacjonistyczne jest w stanie uchwycić specyfikę zespołów błędnej identyfikacji, gdyż traktuje

<sup>3</sup> Hirstein i Ramachandran odwołują się do symulacyjnej koncepcji „czytania umysłów” innych Daniela Goldmana, która głosi, że poznając umysły innych osób, symulujemy we własnych umysłach stany umysłów innych osób.

poważnie wypowiedzi samych pacjentów. Pacjent z zespołem Capgrasa twierdzi, że jego ojciec jest oszustem, dlatego iż odbiera umysł krewnego jako umysł obcej osoby. Jego przekonanie, że krewny jest oszustem, należy potraktować literalnie. Pacjent nie żywi tego przekonania dlatego, że stracił emocjonalną relację z krewnym (podejście emocjonalne) czy też stopień personalnego pokrewieństwa z nim się zmniejszył (podejście odwołujące się do personalnego pokrewieństwa). Twierdzenie, że krewny jest kimś obcym, to zatem twierdzenie dotyczące identyfikacji osoby, jej personalnej idyntityczności, a nie emocjonalnego czy personalnego pokrewieństwa. W tym przypadku indywidualistyczna reprezentacja egocentryczna umysłu krewnego zmieniła się w reprezentację zgeneralizowaną. W przypadku zespołu Fregoliego mamy do czynienia z sytuacją odwrotną – reprezentacje zgeneralizowane zmieniają się w zindywidualizowane reprezentacje egocentryczne bliskiej osoby. Dlatego pacjenci z tym zespołem twierdzą, że nieznanymi im ludziami są w rzeczywistości kimś bliskim.

Hirstein i Ramachandran argumentują, że idyntityczność osoby jest niezależna od stopnia personalnego pokrewieństwa. Jednak zmiana w percypowaniu idyntityczności osoby bliskiej pociąga za sobą zmianę w personalnym pokrewieństwie. Nie jest jednak na odwrót – zmiana w personalnym pokrewieństwie niekoniecznie pociąga za sobą zmianę idyntityczności osoby. Można stracić bliską relację z drugą osobą, lecz nie oznacza to, że będziemy tej osobie przypisywać inny umysł. Wówczas mówimy zwykle, że bliska osoba stała się dla nas kimś obcym. Osoba z zespołem Capgrasa żywi prawie takie samo przekonanie, tyle że czyni to literalnie, a nie metaforycznie.

Reasumując, wedle propozycji Hirsteina i Ramachandrana poznając innych ludzi, tworzymy dwa rodzaje reprezentacji: allocentryczne i egocentryczne. Allocentryczne reprezentacje to wygląd drugiej osoby, natomiast egocentryczne reprezentacje dzielą się na zindywidualizowane egocentryczne reprezentacje innych oraz zgeneralizowane egocentryczne reprezentacje innych. Zgeneralizowane reprezentacje innych posiadamy, gdy dopiero poznajemy daną osobę. Tworzymy wówczas stereotypową reprezentację drugiej osoby. Zindywidualizowane reprezentacje innych posiadamy zaś wówczas, gdy dobrze poznamy inną osobę i potrafimy reprezentować jej charakter, nastroje i emocje, a także przekonania, jakie ona posiada.

Być może więc w przypadkach naturalnych konfabulacji socjocentrycznych mamy do czynienia z podobną sytuacją mieszania reprezentacji zgeneralizowanych i zindywidualizowanych. Na przykład w przypadku stereotypów społecznych przypisujemy innym ludziom cechy na podstawie schematów poznawczych, jakie posiadamy. Tego rodzaju stereotypy społeczne, opierające się na schematach poznawczych, to nic innego jak zgeneralizowane reprezentacje innych osób.

## Podmiotowe i świadomościowe determinanty konfabulacji

### 1. Konfabulacje a podmiotowość

Czym jest ludzkie „ja”? Czy jest ono czymś realnym, czy tylko teoretyczną fikcją? Czy jest czymś jednolitym, czy też różnorodnym, tak iż istnieją różne rodzaje podmiotowości? Czy kiedy prezentujemy nasze własne „ja” innym osobom, dbamy o to, aby przedstawiać je rzetelnie, czy też jest ono raczej produktem konfabulacji? A jeśli nasze „ja” jest lub bywa konfabulacyjne, to w jaki sposób? Przykładem może być sytuacja, gdy opowiadając o sobie samych, wyolbrzymiamy pewne cechy, a inne pomijamy lub minimalizujemy. Ukazując siebie innym i sobie samemu, chcemy pokazać się z jak najlepszej strony. Ulegamy wówczas iluzjom i koloryzujemy cechy naszego charakteru. W ten sposób tworzymy czasami konfabulacyjne „ja”.

Czy nasze „ja” jest czymś jednorodnym? Czy ma jednolitą strukturę? A może raczej ma kilka poziomów lub warstw? Pierwszy podział na różne rodzaje „ja” pojawił się w koncepcji Williama Jamesa, który wyróżnił „ja” podmiotowe i „ja” przedmiotowe, a następnie podzielił „ja” przedmiotowe na: „ja” fizyczne, „ja” mentalne oraz „ja” duchowe (za: Northoff i in. 2006). Podobne rozróżnienia pojawiają się we współczesnych badaniach filozoficznych i kognitywistycznych dotyczących tożsamości osobowej. Przykładem może być koncepcja znanego neurobiologa Antonio Damasio, który wprowadził podział na „proto-ja”, „ja” rdzenne i „ja” auto-biograficzne, wpisujący się w koncepcję Jamesa (Northoff i in. 2006). „Proto-ja” to inaczej „ja” fizyczne u Jamesa, „ja” rdzenne przypomina „ja” mentalne, natomiast „ja” autobiograficzne jest odzwierciedleniem „ja” duchowego. W podobny sposób można zestawić rodzaje „ja” wyróżnione przez Shauna Gallaghera: „ja” minimalne oraz „ja” narracyjne z tymi, które wyróżnił James (Gallagher 2000). „Ja” minimalne koresponduje, zdaniem Northoffa, z „ja” mentalnym Jamesa, a „ja” narracyjne – z „ja” duchowym (Northoff i in. 2006).

## 1.1. Koncepcja pięciu rodzajów „ja” według Ulrica Neissera

Ulric Neisser w artykule pt. *Five kinds of self-knowledge* (1988) wyróżnia pięć rodzajów „ja”. Zauważa on, że kiedy rozpatrujemy podmiotowość ludzką jako obiekt jednorodny, staje się ona pełna sprzeczności. Z jednej strony jest czymś fizycznym, uniwersalnym dla całego gatunku, z drugiej – czymś mentalnym i charakterystycznym dla danej kultury. Czasem rozpatrujemy ją jako coś, co można percypować bezpośrednio oraz jako fenomen publiczny, a czasem jako fenomen prywatny, który jest konstruowany przez naszą wyobraźnię. Sprzeczności te są tylko pozorne, bo gdy przyjrzymy się podmiotowości z różnych stron, okaże się, że fenomen ten ma wiele aspektów, które można pogodzić.

Neisser wyróżnia pięć rodzajów „ja”: ekologiczne, interpersonalne, conceptualne, rozszerzone i prywatne. Kryterium tego podziału jest ontogeneza człowieka, czyli pojawianie się różnego rodzaju podmiotowości na różnych etapach ludzkiego życia. Badacz analizuje różne informacje, z jakimi ma do czynienia człowiek, co prowadzi go do wyszczególnienia pięciu rodzajów „ja”, które nie są odczuwane jako odrębne, lecz zintegrowane ze sobą.

Neisser nie traktuje pięciu rodzajów „ja” jako poszczególnych części ludzkiej osoby. Nie uważa też, by „ja” dało się zlokalizować w mózgu. Odrzuca także hipotezę homunkulusa<sup>1</sup> (Neisser 1993, s. 3). Pod tym względem, jak sam zauważa, jego koncepcja zgodna jest z ustaleniami niektórych współczesnych filozofów, zwłaszcza Daniela C. Dennetta oraz neuronaukowców, takich jak Patricia Churchland i Terrence J. Sejnowski (Neisser 1993, s. 3).

Koncepcja ta pokazuje, że podmiot rozpatrywany jest z kilku różnych perspektyw. Na przykład „ja” ekologiczne to podmiot widziany jako działający w otoczeniu. „Ja” interpersonalne to ten sam podmiot oglądany z perspektywy jego bezpośrednich kontaktów z innymi ludźmi. „Ja” rozszerzone wskazuje na to, że podmiot wykracza poza teraźniejszość poprzez zdolność pamiętania. „Ja” conceptualne pozwala podmiotowi stosować pojęcia na określenie samego siebie. „Ja” prywatne umożliwia zaś wyodrębnienie własnych stanów mentalnych.

### „Ja” ekologiczne

„Ja” ekologiczne pojawia się w momencie narodzin człowieka i dotyczy poznawania środowiska fizycznego przez noworodka. Badacz zauważa, że poznawanie

---

<sup>1</sup> Pojęcie homunkulusa funkcjonuje w filozofii umysłu na określenie obszaru mózgu, który zbiera informacje z innych obszarów i steruje nimi. Jest to też miejsce, w którym dochodzi do uświadomienia informacji. Podejście takie krytykowane jest przez wielu kognitywistów, w tym Dennetta i Neissera.

to odbywa się poprzez percepcję bezpośrednią, podczas działania człowieka w świecie. Nawiązuje on do percepcji bezpośredniej Jamesa Gibsona (1976), twórcy psychologii ekologicznej, wedle której podmiot poznaje świat poprzez działanie. Percepcja i działanie mogą być odróżniane tylko w ramach teoretycznych analiz, w rzeczywistości zaś są ze sobą powiązane.

To, co percypujemy, to my sami, jako *zakorzeni* w środowisku, i działający względem niego. [...] Oprócz specjalnych okazji, nie jest tak, że najpierw percypujemy, a potem kontynuujemy ruch. Percypujemy *jak* działamy i *że* działamy [...] (Neisser 1988, s. 40).

Cechą „ja” ekologicznego jest to, że podobnie jak środowisko istnieje ono obiektywnie, a na jego charakterystykę wpływają obiektywnie istniejące informacje, które nie tylko pozwalają na percypowanie lokalizacji tego „ja”, ale także na jego interakcję ze środowiskiem (Neisser 1988, s. 41). Większość tych informacji ma charakter kinetyczny. „Ja” ekologiczne jest percypowane od wczesnego niemowlęstwa w sposób bezpośredni i nie polega na budowaniu reprezentacji. Dlatego też nie jest uświadamiane w takim sensie, w jakim zwykle mówimy o samoświadomości. Niemowlęta nie budują jeszcze reprezentacji „ja”, a mimo to posiadają „ja” ekologiczne. Z uwagi na to, że „ja” ekologiczne postrzegane jest bezpośrednio, nie podlega ono iluzjom z wyjątkiem rzadkich stanów patologicznych, takich jak kończyny fantomowe u osób po amputacjach.

Neisser utrzymuje, że jego koncepcja „ja” ekologicznego i twierdzenie, że możemy percypować samych siebie, jest istotna tylko wówczas, gdy przyjmujemy teorię percepcji Gibsona (Neisser 1993, s. 6). Teoria ta głosi, że system percepcyjny pobiera informację ze środowiska bezpośrednio, a informacja ta niezmiennie wyszczególnia obiektywnie istniejące stany rzeczy. Percepcja, w przeciwieństwie do wiedzy, nie jest konstruktem, lecz pozwala na widzenie środowiska i własnych działań w świecie takimi, jakimi są one w rzeczywistości. Studia nad percepcją mają na celu zidentyfikowanie obiektywnie istniejącej informacji oraz zbadanie, w jaki sposób jest ona pozyskiwana. Podejście to zakłada, że organizmy wyposażone w system percepcyjny osiągają ewolucyjne korzyści z pozyskiwania takiej informacji.

Cechami charakterystycznymi „ja” ekologicznego są zdolność rozumienia własnej lokalizacji oraz zdolność do ruchu w przestrzeni. Istotną rolę w odbiorze lokalizacji i ruchu odgrywa percepcja wzrokowa. Widzenie dostarcza bogatej informacji na temat środowiska oraz nas samych. Zaproponowane przez Gibsona pojęcie optycznego umieszczenia (*optic array*) pozwala określić przestrzeń w środowisku, w której się poruszamy, zawierającą nieskończoną liczbę potencjalnych punktów obserwacji, z których każdy otoczony jest „skorupą” optycznej struktury (Neisser 1993, s. 6). Ruch obserwatora powoduje optyczny przepływ (*optic flow*), który pozwala na percypowanie naszych ruchów w przestrzeni.

Innym ważnym elementem „ja” ekologicznego jest działanie oraz poczucie sprawstwa, które Gibson określa mianem afordancji. Przedmioty w otoczeniu stanowią swego rodzaju oferty, dzięki którym działamy, np. pióro jest ofertą pisania, drzwi ofertą przechodzenia, podłoga ofertą chodzenia. Każda aktualna sytuacja oferuje nieskończoną liczbę potencjalnych działań. Jednak w danym momencie postrzegamy niewielką liczbę ofert przedmiotów, a jeszcze mniej realizujemy w działaniu.

### „Ja” interpersonalne

„Ja” interpersonalne powstaje od wczesnego dzieciństwa i dotyczy interakcji społecznych. Podobnie jak „ja” ekologiczne, nie jest ono zapośredniczone przez reprezentacje, lecz cechuje je percepcja bezpośrednia, dokonywana na podstawie obiektywnie istniejącej informacji. Ta informacja powstaje wówczas, gdy dochodzi do interakcji pomiędzy dwiema lub więcej osobami. Interakcje pomiędzy podmiotami są percypowane przez nie same, ale czasem także przez zewnętrznych obserwatorów. To, co charakterystyczne dla „ja” interpersonalnego, to odpowiedzi na zachowania innej osoby, które tworzą intersubiektywność.

„Ja” interpersonalne jest specyficzne dla danego gatunku. Genetyczne wyposażenie człowieka pozwala na wytwarzanie pewnego rodzaju ekspresji, gestów i wokalizacji oraz reagowanie na tego rodzaju zachowania innych ludzi. Inne gatunki zwierząt również posiadają wyposażenie pozwalające na komunikowanie się z pozostałymi, jednak pewne ludzkie zdolności komunikowania się są specyficzne tylko dla nas.

Neisser podkreśla, że „ja” interpersonalne opiera się raczej na percepcji bezpośredniej niż na inferencji, co potwierdzają badania nad interakcją niemowląt z opiekunami. W wieku dwóch miesięcy niemowlęta posiadają już wiele zdolności społecznych. Od szóstego tygodnia życia poszukują wzroku opiekuna, zaczynają się uśmiechać, a w wieku dwóch miesięcy stają się szczególnie wrażliwe na głos matki i zaczynają reagować oraz niewerbalnie odpowiadać na wokalizacje (Trevvarthen 1983).

### „Ja” konceptualne

„Ja” konceptualne to inaczej pojęcia, jakie tworzymy na określenie nas samych w środowisku społecznym. Różnią się one w zależności od społeczności czy kultury. „Ja” konceptualne stanowią przede wszystkim określenia naszych ról społecznych, takich jak bycie matką czy ojcem, żoną czy mężem, córką, synem, pracownikiem, Polakiem. Co więcej, na „ja” konceptualne składają się też określenia nas samych, naszych zdolności, umiejętności, zainteresowań, cech charakteru.

Są to wszystkie pojęcia, jakie przypisujemy sobie samym. „Ja” konceptualne nie bazuje na percepcji bezpośredniej, lecz na reprezentacjach.

„Ja” konceptualne powstaje pod koniec pierwszego roku życia, kiedy niemowlę zaczyna traktować siebie jako obiekt myślenia. Według Neissera umiejętność ta jest związana ze zdolnością uwspólniania uwagi, która zaczyna się nie prędzej niż w dziesiątym miesiącu życia. Uwspólnianie uwagi to triadyczna relacja między dzieckiem, opiekunem oraz obiektem, która pozwala na zrozumienie, iż dziecko i matka patrzą na ten sam obiekt. Od tego momentu dziecko zaczyna rozumieć, na co inni ludzie zwracają uwagę i o czym mówią, zwłaszcza gdy samo staje się obiektem zainteresowania innych osób. Dopiero wówczas, gdy dziecko zrozumie, iż samo jest przedmiotem zainteresowania innych, może powstać pojęcie dotyczące „ja” (Neisser 1993, s. 5).

Zdaniem Neissera pojęcia odnoszące się do „ja” mogą być rozpatrywane jako pojedynczy model kognitywny:

Moje pojęcie tego, kim jestem, tak jak twoje pojęcie tego, kim jesteś, odzwierciedla poznawczy model zakorzeniony w teoretycznej sieci [...]. Tak jak inne teorie, niekoniecznie ta jest poprawna; każdy z nas zna ludzi, których teorie na temat „ja” wydają się być błędne pod pewnymi względami [...] (Neisser 1988, s. 53).

Poznawczy model „ja” konceptualnego składa się z kilku subteorii. Należą do nich: teorie roli (*role theories*), modele wewnętrzne (*internal models*) oraz atrybucje cech (*trait attributions*). Teorie roli dotyczą naszych pojęć odnoszących się do nas samych w ramach pełnionych ról społecznych. Pojęcia te określają, w jaki sposób jesteśmy dopasowani do społeczeństwa, jak realizujemy swoje obowiązki i egzekwujemy prawa. Modele wewnętrzne to z kolei pojęcia na temat nas samych, naszych ciał oraz umysłów. Pojęcia odnoszące się do ciała pochodzą w większości z medycyny i biologii, a pojęcia dotyczące umysłu mają oparcie w psychologii, filozofii czy religii. Atrybucje cech stanowią natomiast pomost między teoriami ról oraz modelami wewnętrznymi. Oznaczają przypisywanie sobie różnego rodzaju cech wyglądu i charakteru.

„Ja” konceptualne różni się od innych rodzajów „ja” tym, że jego podstawą są pojęcia ustalane w dyskursie społecznym, które mogą być komunikowane werbalnie.

Jednakże, kiedy rozwijamy nasze własne konceptualne „ja”, często próbujemy wziąć pod uwagę inne rodzaje doświadczenia. Zatem nasze pojęcia dotyczące „ja” zawierają idee na temat naszych fizycznych ciał, komunikacji interpersonalnej, na temat tego, jakie rzeczy robiliśmy w przeszłości i najprawdopodobniej zrobimy w przyszłości, a w szczególności na temat znaczenia naszych własnych odczuć i myśli. W rezultacie każdy z czterech innych rodzajów samowiedzy jest także reprezentowany w „ja” konceptualnym (Neisser 1988, s. 54).

## „Ja” rozszerzone

Kolejnym rodzajem „ja”, które kształtuje się w ciągu naszego życia, jest „ja” rozszerzone, zwane też przez Neissera „ja” pamięciowym. „Ja” ekologiczne i „ja” interpersonalne bazowały na percepcji bezpośredniej i dotyczyły tego, co dzieje się w teraźniejszości. „Ja” rozszerzone opiera się natomiast na pamięci zdarzeń przeszłych i antycypacjach przyszłości (Neisser 1988, s. 46). Wykorzystuje ono reprezentacje z naszej przeszłości. Korzysta zarówno z pamięci epizodycznej, bezpośrednio dotyczącej poszczególnych zdarzeń z naszego życia, jak i z pamięci schematów poznawczych, czyli ogólnej wiedzy na temat tego, jak przebiegają dane zdarzenia (Neisser 1988, s. 47).

Z uwagi na to, że „ja” rozszerzone opiera się na reprezentacjach pamięci autobiograficznej, nie każdy rodzaj pamięci stosuje się do tego rodzaju podmiotowości. W psychologii poznawczej wyróżniono pamięć deklaratywną i niedeklaratywną (Squire 1986). Podział ten nawiązuje do wprowadzonego przez Gilberta Ryle’a podziału na wiedzę proceduralną („wiedza jak”) i deklaratywną („wiedza że”) (Ryle 1949). Larry Squire wprowadził zaś ogólniejszy termin „pamięć niedeklaratywna”, gdy okazało się, że istnieją zjawiska podobne do pamięci proceduralnej. Jego podział opiera się na dwóch kryteriach: sposobie przechowywania informacji w pamięci oraz mechanizmach wykorzystywanych do wydobycia informacji z pamięci (za: Maruszewski 2001, s. 161).

Informacje zawarte w pamięci deklaratywnej przechowywane są w postaci abstrakcyjnych lub konkretnych reprezentacji angażujących język i właśnie dzięki wykorzystaniu języka mogą być łatwo przywoływane w postaci słownej. Dostęp do tych informacji jest względnie niezależny od kontekstu, to jest możemy je odtwarzać w innych sytuacjach aniżeli te, w których owe informacje zostały przyswojone [...]. Natomiast informacje [...] w pamięci niedeklaratywnej przechowywane są w postaci związków S-R, czyli – jak powiedzieliby behawioryści – nawyków. Nawyki te mogą być bardzo proste, jak na przykład nawyk cofania palców przed bardzo gorącymi przedmiotami, ale mogą być też bardzo złożone, jak na przykład nawyk pisania tekstu za pomocą komputera (Maruszewski 2001, s. 161).

Pamięć deklaratywna dzieli się dalej na pamięć epizodyczną, tj. dotyczącą epizodów z życia podmiotu, i semantyczną, tj. dotyczącą faktów (Tulving 1972).

Pamięć proceduralna, czyli pamięć umiejętności, jest niezależna od pamięci deklaratywnej, gdyż nie opiera się na reprezentacjach mentalnych. Osoby cierpiące na amnezję są niezdolne do deklaratywnego przypominania sobie zdarzeń z przeszłości, ale badania wykazały, że ich pamięć proceduralna jest nienaruszona – mogą uczyć się nowych umiejętności, mimo że nie pamiętają, iż się ich uczą. Dlatego też amnezja prowadzi do uszkodzenia „ja” rozszerzonego (Neisser 1988, s. 46).

„Ja” rozszerzone pojawia się około trzeciego roku życia, kiedy dzieci zaczynają tworzyć pamięć autobiograficzną. Jest to związane z amnezją dziecięcą, doświadczaną w wieku dorosłym, czyli niezdolnością przypomnienia sobie, co wydarzyło się przed trzecim rokiem życia. Dzieci przed trzecim rokiem życia zaangażowane są w mówienie na temat teraźniejszości i rzadko zainteresowane przeszłością, chociaż potrafią przypominać sobie różne rzeczy. Zainteresowanie przeszłością wzrasta, gdy dzieci po trzecim roku życia zaczynają zdawać sobie sprawę ze społecznego znaczenia pamięci autobiograficznej, która pozwala na rozszerzenie relacji z innymi ludźmi poza teraźniejszość (Neisser 1988, s. 48).

Pierwszym kryterium pojawienia się pamięci autobiograficznej jest – zdaniem Tomasza Maruszewskiego – zdolność relacjonowania własnej przeszłości, która pojawia się między trzecim a czwartym rokiem życia (Maruszewski 2005, s. 129). Choć wcześniej dziecko uczy się mówić, że jakiś fakt miał miejsce, to pamięć autobiograficzna polega na zdolności opowiadania prostych historii na temat własnej przeszłości. Drugim kryterium pojawienia się pamięci autobiograficznej jest zdolność odróżniania przez dziecko tego, co przydarzyło się jemu, od tego, co przydarzyło się innym. Około osiemnastego miesiąca życia dziecko przechodzi test lustra, czyli zaczyna rozpoznawać siebie. Wraz z tym zyskuje zdolność przypisywania wspomnień samemu sobie (Maruszewski 2005, s. 129).

Zdaniem Neissera zdolność opowiadania historii na temat własnej przeszłości to umiejętność, którą nabywamy powoli w toku ontogenezy (Neisser 1991, s. 206).

Pamięć, w przeciwieństwie do percepcji, podlega wielu zniekształceniom. Opowiadane przez ludzi historie na temat własnej przeszłości mogą znacznie się różnić w zależności od kontekstu, w jakim są wypowiedziane:

Doskonale wrażliwe na wymogi sytuacji, w której są wypowiedziane, historie te mogą różnić się wyraźnie od okazji do okazji. Ponadto nie bazują one na przechowywanej „pamięci” samej w sobie; polegamy na konwencjach narracyjnych, teoriach *implicite*, a nawet na przypomnieniach innych osób, gdy konstruujemy dzisiejszą relację wczorajszego dnia (Neisser 1991, s. 206).

W artykule *Two Perceptually Given Aspects of the Self and Their Development* Neisser wyróżnia dwa aspekty „ja” rozszerzonego: standardowo przechowywane „ja” (*a standard stored self*) oraz obecnie przypominane „ja” (*a presently remembered self*) (Neisser 1991, ss. 206-207). Standardowo przechowywane „ja” to stabilna kognitywna struktura dotycząca doświadczeń z przeszłości, którą posiada każdy dorosły człowiek. Jest to coś, co posiadamy, czyli pewnego rodzaju reprezentacja mentalna. W przeciwieństwie do tego obecnie przypominane „ja” to coś, czego dokonujemy – jest to opowiadanie historii na temat naszej przeszłości, które różni się w zależności od okoliczności. Stanowisko Neissera

jest więc takie, iż pamięć jest konstruktem, a nie dokładnym odtwarzaniem minionych zdarzeń:

Przypominanie jest zawsze konstruktem czegoś nowego, odpowiedniego względem okazji: *obecnie przypominające ja* (Neisser 1991, s. 207).

Jaka jest relacja między „ja” rozszerzonym a innymi rodzajami „ja”? Po pierwsze, jest ono niezależne od „ja” ekologicznego w tym sensie, że przypominamy sobie coś niezależnie od okoliczności czasu teraźniejszego. Nie doświadczamy jednak „ja” ekologicznego i „ja” pamięciowego jako czegoś odrębnego; przypominamy sobie coś, percypując jednocześnie chwilę obecną, i w tym sensie te dwa rodzaje „ja” są ze sobą powiązane. Odrębność „ja” pamięciowego widoczna jest jednak w kulturowych i indywidualnych różnicach, które przejawiają się tym, iż niektórzy ludzie bardziej skupiają się na wspomnianiu przeszłości niż inni.

Z kolei „ja” rozszerzone powiązane jest z „ja” konceptualnym w tym sensie, że nabyta wiedza na temat nas samych i świata wpływa na to, jak sobie przypominamy przeszłość (co więcej, wiedza płynąca z „ja” konceptualnego może wpływać na „ja” ekologiczne oraz „ja” interpersonalne, czyli na to, jak postrzegamy samych siebie w świecie fizycznym i nasze relacje z innymi). Pamięć jest konstruktem, który podlega zniekształceniom, często w zależności od tego, jakie posiadamy teorie na temat nas samych, które tworzą nasze „ja” konceptualne.

### „Ja” prywatne

Ostatni rodzaj podmiotowości to „ja” prywatne, które powstaje wówczas, gdy dziecko zaczyna zdawać sobie sprawę, że posiada indywidualne doświadczenia, niezależne od doświadczeń innych osób (Neisser 1988, s. 50). Na „ja” prywatne składają się indywidualne przeżycia odnoszące się zarówno do percepcji i działania, jak również pamięci, wyobraźni i snów. Podczas gdy dziecko od początku życia posiada doświadczenia percepcyjne pozwalające na działanie w świecie fizycznym i społecznym (posiada „ja” ekologiczne i „ja” interpersonalne), świadomość tego, iż jego doświadczenia są prywatne, pojawia się później w rozwoju ontogenetycznym.

Wiele badań wskazuje na to, że ta świadomość prywatności doznań przychodzi, zanim dziecko skończy piąty rok życia. Neisser powołuje się na badania, z których wynika, że dopiero czterolatki rozumieją pojęcie „sekrety” (Mosslera i in. 1976, za: Neisser 1988, s. 50). W sytuacji gdy czteroletnie dziecko i badacz, porozumiewając się za pomocą gestów, wiedzą coś, czego nie wie osoba trzecia, która nie patrzy na nich, czterolatek rozumie, że nie wie ona tego samego, co on i badacz.

To, jak bardzo nasze „ja” prywatne dochodzi do głosu, zależy głównie od cech indywidualnych. Niektórzy z nas częściej niż inni odwołują się do doświadczenia wewnętrznego. Neisser powołuje się także na rozróżnienie Carla Gustava Junga na ekstrawertyków i introwertyków, aby dowiedzieć, że introwertycy częściej niż ekstrawertycy opierają się na własnych wewnętrznych stanach mentalnych (Neisser 1988, s. 51).

Jakie są powiązania między „ja” prywatnym a innymi rodzajami „ja”? „Ja” prywatne nie dotyczy tylko uświadamiania sobie, jakie stany mentalne aktualnie posiadam, ale również odnosi się do przypominania i wyobrażania sobie, w tym sensie związane jest z „ja” rozszerzonym. „Ja” prywatne jest też powiązane z „ja” konceptualnym, gdyż korzysta z pojęć odnoszących się do siebie samego, nabytych w toku doświadczenia. Jest z kolei stosunkowo niezależne od „ja” ekologicznego i „ja” interpersonalnego. Można sobie bowiem wyobrazić sytuację, w której percypując jakiś stan rzeczy, wyobrażam sobie coś innego, albo rozmawiając z jakąś osobą, myślę o innej osobie. Nie należy jednak z owej niezależności „ja” prywatnego od „ja” ekologicznego wnioskować, że „ja” prywatne jest również niezależne od materialnego mózgu (Neisser 1988, s. 51).

Czy „ja” prywatne poddaje się złudzeniom? W filozofii uważano dotąd, że „ja” prywatne jest jedynym wartym rozważenia właśnie dlatego, iż nie poddaje się złudzeniom, nie jest narażone na błędy. Reprezentantem takiego stanowiska był Kartezjusz, który twierdził, że jedyne, czego możemy być pewni, to własna świadomość istnienia. Neisser w przeciwieństwie do tego argumentuje, że jedyne „ja” niepodlegające złudzeniom to „ja” ekologiczne i „ja” interpersonalne. Zarówno „ja” ekologiczne, jak i „ja” interpersonalne bazują na percepcji bezpośredniej, podczas gdy trzy kolejne rodzaje podmiotowości są związane z reprezentacjami. „Ja” konceptualne, a w szczególności „ja” rozszerzone, narażone jest na zniekształcenia, gdyż jego domeną są reprezentacje przechowywane przez ludzki umysł (Neisser 1988).

### 1.1.1. Koncepcja Neissera a konfabulacje

Według Neissera zniekształceniom podlegają te rodzaje „ja”, które opierają się na reprezentacjach mentalnych. Są to: „ja” konceptualne, „ja” rozszerzone oraz „ja” prywatne. „Ja” ekologiczne i „ja” interpersonalne, oprócz kilku patologicznych przypadków, w zasadzie nie poddają się błędom ani iluzjom, gdyż do ich powstania niezbędna jest percepcja bezpośrednia.

Koncepcję Neissera można odnieść do zagadnienia konfabulacji, ponieważ wyróżnione przez niego rodzaje „ja”, które podlegają zniekształceniom, odpowiadają pod względem sposobu rozumienia ich funkcji konfabulacjom wyróżnionym przeze mnie (por. rozdział III). Konfabulacje pamięciowe i pre-

dykcyjne to zniekształcenia „ja” rozszerzonego. Konfabulacje egocentryczne to zniekształcenia „ja” konceptualnego oraz „ja” prywatnego. Pozostają jednak jeszcze inne, już wyróżnione. Powstaje więc pytanie: Jaki jest ich charakter w kontekście rozumienia fenomenu konfabulacji? Jeśli „ja” ekologiczne nie podlega zniekształceniom, to jak możliwe są konfabulacje percepcyjne oraz konfabulacje naturalistyczne? Jeśli „ja” interpersonalne nie poddaje się iluzjom i jest niepodatne na błędy, to jak możliwe są konfabulacje socjocentryczne?

Według mnie konfabulacje percepcyjne opierają się na reprezentacjach mentalnych i odnoszą do „ja” konceptualnego, bazującego na „ja” ekologicznym i „ja” interpersonalnym. Tak powstają konfabulacje percepcyjne interpersonalne oraz konfabulacje percepcyjne naturalistyczne.

W jaki sposób „ja” rozszerzone poddaje się konfabulacjom? Opiera się ono na narracjach branych z naszego życia, ale „rozciągniętych” w czasie. Ten rodzaj „ja” przez niektórych filozofów i kognitywistów uznawany jest za najistotniejszy w rozważaniach nad ludzką podmiotowością, np. Daniel C. Dennett uważa, że „ja” to środek narratywnej ciężkości (Dennett 1991, 1992).

Znaczącym osiągnięciem gatunku ludzkiego jest zdolność rekonstruowania minionych zdarzeń. Wszystkie organizmy doświadczają wpływu przeszłości na terażniejszość, ale tylko człowiek potrafi odtwarzać przeszłość w sposób deklaratywny. Tego rodzaju odtwarzanie zdarzeń dokonuje się za pomocą pamięci epizodycznej, zwanej pamięcią zdarzeń. Częścią pamięci epizodycznej jest pamięć autobiograficzna, czyli ten rodzaj przypominania i odtwarzania zdarzeń, który odwołuje nas bezpośrednio do osobistych przeżyć, tworząc narrację o naszym życiu. Zdaniem Neissera „narracje na temat życia są znaczące, gdyż są jednym ze sposobów definicji »ja«” (Neisser 1994, s. 1).

W artykule pt. *Self-narratives: True and false* (1994) Neisser zauważa, że nie wszystkie narracje na temat „ja” są prawdziwe. Zwraca on uwagę na powszechność błędów i konfabulacji pamięciowych (Neisser 1994, ss. 4-5). Zastanawia się, do jakiego stopnia można wierzyć wspomnieniom, skoro występowanie zniekształceń pamięci oraz fałszywych wspomnień jest tak częste. Odwołuje się przy tym do badań Elizabeth Loftus, które wykazały, iż sugestie prowadzą często do fałszywych wspomnień. Powołuje się też na własne badanie przeprowadzone wraz z Nicole Harsch nad pamięcią błyskową, które wykazało, że nawet wspomnienia, uważane dotąd za niepodlegające złudzeniom, w rzeczywistości ulegają zniekształceniom.

Według Neissera w ramach pamięci epizodycznej należy wyróżnić kilka kwestii:

- (1) aktualne zdarzenie,
- (2) zdarzenie – o którym mowa – tak jak było ono doświadczane przez jednostkę,

- (3) późniejszy akt przypominania sobie go,
- (4) przypomniane zdarzenie, czyli szczególna wersja (1), która jest ustalana przez (3) (Neisser 1994, s. 2).

Analogiczne kategorie można wyróżnić w ramach pamięci autobiograficznej:

- (1) aktualne przeszłe zdarzenia, w których uczestniczy „ja” historyczne,
- (2) te zdarzenia, tak jak były doświadczane, włączając w to „ja” percepcyjne jednostki w tamtym czasie,
- (3) „ja” pamięciowe, czyli jednostka w akcie przypominania sobie tych zdarzeń przy jakiejś późniejszej okazji,
- (4) „ja” pamięciowe skonstruowane przy tej okazji (Neisser 1994, s. 2).

Według Neissera nasze „ja” pamięciowe zmienia się każdego dnia:

„Ja”, które jest pamiętane dzisiaj, nie jest historycznym „ja” z wczoraj, lecz tylko jego zrekonstruowaną wersją. Inna wersja – nowe „ja” pamięciowe – może być zrekonstruowane jutro. Jak bardzo jest różne? Ja sam jestem stronniczy względem ciągłości i mam tendencję do myślenia o większości „ja” pamięciowych jako dość stabilnych z jednego dnia na drugi (Neisser 1994, s. 8).

Neisser odwołuje się do kilku tekstów z pracy zbiorowej pt. *The remembering self: Construction and accuracy in self-narrative* (1994) oraz do tekstu Daniela Albrighta pt. *Literary and psychological models of the self*, który porusza kwestię zmienności i niekompletności „ja” pamięciowego. Badacz ten wątpi w jednorodność i koherencję „ja”, twierdząc, że u jego podstaw leży „rozbicie pamięci”. Wolalby raczej nazywać je „ja” niepamiętnym niż pamięciowym. Zwraca też uwagę na to, że podmiot, przypominając sobie coś, musi wypełniać ogromne pola pustej przestrzeni, co skłania go do egocentrycznych konfabulacji pamięciowych. Jednym ze stanów wyjściowych konfabulacji jest bowiem luka w systemie wiedzy, która zostaje zniwelowana przez wypełnianie konfabulacyjne. W ten sposób powstaje egocentryczna pamięciowa konfabulacja wypełniająca.

Badaczem, który również rozważa mnogość i niejednorodność „ja” pamięciowego, jest Edward Reed. W artykule pt. *Perception is to self as memory is to selves* (1994) zestawia i porównuje on percepcję do pamięci. Uważa, że o ile percepcja ma charakter unitarny i jednorodny, o tyle pamięć ma charakter dualny (za: Neisser 1994, s. 8).

Innym badaczem, na którego powołuje się Neisser, jest Jerome Bruner, który w artykule pt. *The „remembered” self* (1994) odrzuca pojęcie „ja” pamięciowego. Twierdzi on, że aktualne ludzkie narracje tworzące „ja” opierają się bardziej na bieżącym myśleniu niż na samym przypominaniu. Narracje te zmieniają się w zależności od okazji, nastroju czy też słuchaczy. Z tego wynika mnogość „ja”, które pamiętamy. Sam Neisser przyznaje, że narracje na temat „ja” nie muszą opierać się wyłącznie na pamięci epizodycznej (Neisser 1994, s. 1).

Zasadne wydaje się stwierdzenie, że „ja” rozszerzone to inaczej „ja”, które bazuje na pamięci, a jednocześnie na antycypacjach przyszłości. Według Neissera „ja” rozszerzone oznacza, że podmiot „rozciągnięty” jest w czasie w dwóch kierunkach – od przeszłości poprzez pamięć do przyszłości poprzez antycypację (Neisser 1994, s. 16). Choć Neisser używa zamiennie pojęć „ja” rozszerzone i „ja” pamięciowe, to należy przyjąć, iż „ja” pamięciowe jest częścią „ja” rozszerzonego. Autonarracja może opierać się na pamięci, ale może także dotyczyć bieżących myśli albo opierać się na antycypacji. W tym sensie „ja” rozszerzone daje podstawy nie tylko do konfabulacji pamięciowych, ale także do konfabulacji predykcyjnych.

Wyróżniłam ponadto trzy inne rodzaje konfabulacji z uwagi na przedmiot konfabulowania: naturalistyczne, socjocentryczne i egocentryczne. Sądzę bowiem, że „ja” ekologiczne w połączeniu z „ja” konceptualnym daje podstawy do percepcyjnych konfabulacji naturalistycznych. Są to np. konfabulacje na temat aktualnych stanów świata. Z kolei „ja” interpersonalne w połączeniu z „ja” konceptualnym daje podstawy do percepcyjnych konfabulacji socjocentrycznych, czyli konfabulacji na temat umysłów innych osób doświadczanych w trybie terażniejszym. Percepcyjne konfabulacje egocentryczne są natomiast związane z „ja” prywatnym i „ja” konceptualnym. Są to konfabulacje na temat bieżących doświadczeń podmiotu.

Oczywiście wyróżnione konfabulacje rozpięte są także na przeszłości i przyszłości. „Ja” rozszerzone i „ja” konceptualne dają podstawy do pamięciowych i predykcyjnych konfabulacji naturalistycznych i socjocentrycznych. Z kolei „ja” rozszerzone, „ja” konceptualne i „ja” prywatne dają podstawy do pamięciowych i predykcyjnych konfabulacji egocentrycznych. Tabela 9 zawiera zestawienie tych propozycji interpretacyjnych.

Tabela 9. Zestawienie rodzajów „ja” w ujęciu Neissera z proponowanymi przeze mnie typami konfabulacji

Rodzaje „ja” w ujęciu Neissera	Rodzaje konfabulacji (propozycja własna)
„Ja” ekologiczne i „ja” konceptualne	konfabulacje percepcyjne naturalistyczne
„Ja” interpersonalne i „ja” konceptualne	konfabulacje percepcyjne socjocentryczne
„Ja” konceptualne i „ja” prywatne	konfabulacje percepcyjne egocentryczne
„Ja” rozszerzone i „ja” konceptualne	konfabulacje pamięciowe i predykcyjne naturalistyczne i socjocentryczne
„Ja” rozszerzone, „ja” konceptualne i „ja” prywatne	konfabulacje pamięciowe i predykcyjne egocentryczne

Źródło: opracowanie własne.

Konfabulacje percepcyjne naturalistyczne dotyczą percepcji stanów świata. Przykładem tego rodzaju klinicznych konfabulacji może być pacjent z zespołem Antona, który konfabuluje na temat stanów świata. Wydaje się więc, że konfabulowanie tego typu dotyczy „ja” ekologicznego w połączeniu z „ja” konceptualnym. Pacjent twierdzi, że widzi otaczający go świat i z detalami przedstawia przebieg sceny wzrokowej, mimo że potyka się o przedmioty w swoim otoczeniu.

Konfabulacje percepcyjne socjocentryczne to konfabulacje na temat innych osób. Przykładem w ramach konfabulacji klinicznych są konfabulacje występujące w zespołach błędnej identyfikacji. Pacjent z zespołem Capgrasa konfabuluje, że ojciec, którego widzi, nie jest jego ojcem, tylko oszustem. Z kolei pacjent z zespołem Fregoliego konfabuluje, że nieznana mu, widziana aktualnie osoba jest w rzeczywistości jego znajomym. Wydaje się, że sytuacje te dotyczą „ja” konceptualnego i „ja” interpersonalnego, z uwagi na to, że pacjenci ci mają błędną reprezentację drugiej osoby, a także przypisują innym błędne pojęcia – oszusta bądź znajomego. Jednak percepcyjne konfabulacje socjocentryczne to nie tylko przypadki kliniczne. Przykładem normalnych percepcyjnych konfabulacji socjocentrycznych jest chociażby efekt aureoli (zwany też efektem halo), czyli tendencja do przypisywania cech osobowościowych na podstawie pozytywnego lub negatywnego wrażenia. Dzieje się to wówczas, gdy ktoś ocenia drugą osobę po wyglądzie zewnętrznym, np. przypisując atrakcyjnej osobie inteligencję, przyjazność, uczciwość.

Konfabulacje pamięciowe i predykcyjne naturalistyczne i socjocentryczne dotyczą „ja” rozszerzonego i „ja” konceptualnego. Konfabulacje pamięciowe naturalistyczne dotyczą przeszłych stanów świata. Przykładem tego są fałszywe wspomnienia odnoszące się do stanów świata. Natomiast konfabulacje pamięciowe socjocentryczne dotyczą przeszłych relacji z innymi osobami.

Z kolei konfabulacje predykcyjne naturalistyczne odnoszą się do przyszłych stanów świata. Ich przykładem są konfabulacje dotyczące przewidywań, takich jak pełne obaw stwierdzenie, że samolot, którym się leci, ulegnie katastrofie. Konfabulacje tego typu dotyczą zwykle błędnego oszacowania prawdopodobieństwa jakiegoś zdarzenia (np. błędne oszacowanie prawdopodobieństwa, że wygra się na loterii). Natomiast konfabulacje predykcyjne socjocentryczne dotyczą przyszłych relacji z innymi osobami, czego przykładem jest konfabulacja pesymisty na temat tego, że w przyszłości relacje z innymi będą układać się gorzej niż obecnie.

Konfabulacje percepcyjne egocentryczne to zniekształcenia „ja” konceptualnego i „ja” prywatnego, np. gdy ktoś konfabuluje na temat bieżących doświadczeń. Natomiast konfabulacje egocentryczne pamięciowe i predykcyjne to konfabulacje na temat przeszłych i przyszłych doświadczeń własnych. Przy-

kładem konfabulacji pamięciowych jest tendencyjność pamięci, która polega w szczególności na złudzeniu zgodności. Osoba ulegająca błędowi tendencyjności uważa, że jej poglądy (np. polityczne) nie zmieniły się znacznie (choć w rzeczywistości mogły ulec zmianie). Innym rodzajem błędu tendencyjności jest złudzenie zmiany. Występuje ono wówczas, gdy osoba bardzo chce zmienić jakieś cechy swego wyglądu, charakteru, światopoglądu i twierdzi, że tego dokonała, a w rzeczywistości nie nastąpiła znacząca zmiana. Są to egocentryczne konfabulacje pamięciowe, które powstają w wyniku zniekształcenia „ja” rozszerzonego, „ja” konceptualnego i „ja” prywatnego.

Z kolei konfabulacje predykcyjne egocentryczne to konfabulacje na temat przyszłych doświadczeń. Przykładem tego może być hipochondryk, który przewiduje, że stan jego zdrowia ulegnie wkrótce pogorszeniu, lub osoba, która oczekuje negatywnych emocji w związku z nadchodzącymi wydarzeniami, choć emocje te mogą wcale nie wystąpić, np. zdenerwowany student przed egzaminem oczekuje, że nie zda, mimo iż się przygotował (może się jednak zdarzyć, że negatywne emocje opadną zaraz po wejściu na egzamin).

Można zatem stwierdzić, że w koncepcji Neissera różne typy konfabulacji powiązane są z różnymi rodzajami „ja”. Chociaż „ja” ekologiczne i „ja” interpersonalne nie ulegają według Neissera zniekształceniom, to moim zdaniem powiązanie tych rodzajów „ja” z „ja” konceptualnym może być podstawą tworzenia kolejnych typów zniekształceń, np. zniekształcenie „ja” ekologicznego i „ja” konceptualnego może prowadzić do konfabulacji percepcyjnych naturalistycznych, a zniekształcenie „ja” interpersonalnego i „ja” konceptualnego – do konfabulacji percepcyjnych socjocentrycznych. Z kolei zdaniem Neissera „ja” rozszerzone, „ja” konceptualne i „ja” prywatne ulegają zniekształceniom, zaś według mnie konfabulacje pamięciowe i predykcyjne socjocentryczne i naturalistyczne powstają w wyniku zniekształcenia „ja” rozszerzonego i „ja” konceptualnego. Natomiast konfabulacje egocentryczne powstają w wyniku zniekształcenia „ja” konceptualnego i „ja” prywatnego (konfabulacje percepcyjne egocentryczne) oraz zniekształcenia „ja” rozszerzonego, „ja” prywatnego i „ja” konceptualnego (konfabulacje pamięciowe i predykcyjne egocentryczne).

## 1.2. Jaźń jako środek narratywnej ciężkości według Daniela C. Dennetta

Daniel C. Dennett podaje w wątpliwość istnienie realnego „ja”, uznając tym samym, że poszukiwanie jaźni w mózgu to błąd kategoryalny. Chociaż jego zdaniem badania nad mózgiem przybliżają nas do poznania ludzkiej podmiotowości, wątpli on, by kiedykolwiek udało się zlokalizować jaźń w jakimś ośrodku mózgu.

Istnienie minimalnego, biologicznego „ja”, niezbędnego do przetrwania, zapewniła nam ewolucja. Takie „ja” biologiczne posiada każde zwierzę (Dennett 1991, s. 413). Człowiek jednak oprócz „ja” biologicznego posiada też „ja” narracyjne, które jest jego naturalnym wyposażeniem, podobnie jak skorupa jest wyposażeniem żółwia (Dennett 1991, s. 418).

Cechą ludzkiej jaźni jest tworzenie sieci dyskursów (*web of discourses*) ze słów i podejmowanych działań (Dennett 1991, s. 416). Każdy normalny człowiek wytwarza „ja”:

Przędzie z mózgu sieć słów i czynów [...]. Ta sieć ochrania go, tak jak skorupa ślimaka i dostarcza mu środków do życia, tak jak sieć pająka [...] (Dennett 1991, s. 416).

Ta sieć dyskursów, która tworzy jaźń, jest tak samo biologicznym tworem jak każda inna konstrukcja w świecie zwierząt i bez niej istota ludzka byłaby niekompletna, podobnie jak ptak bez piór. Jest ona częścią fenotypu rozszerzonego człowieka. Tym, co różni człowieka od zwierząt, jest jednak zdolność reprezentowania samego siebie i prezentowania siebie innym. Reprezentowanie to jest możliwe dzięki językowi i gestom. Tym, co odróżnia środowisko człowieka od środowiska zwierząt, są słowa. Środowisko człowieka to nie tylko cztery „P”: pokarm, partner, przeciwnik, przyjaciel, lecz także świat języka.

Ludzka świadomość i narracyjna podmiotowość są produktem opowieści, jakie tworzymy dla siebie i innych.

Naszą fundamentalną taktyką samoochrony, samokontroli oraz samodefiniowania nie jest przedanie sieci ani budowanie tam, ale opowiadanie historii, a szczególnie preparowanie i kontrolowanie historii, którą opowiadamy innym – oraz sobie – na temat tego, kim jesteśmy. I tak jak pająki nie muszą myśleć, świadomie i umyślnie, o tym, jak prząść sieć [...], my (oprócz profesjonalnych opowiadaczy historii) świadomie i umyślnie nie wymyślamy, jakie narracje opowiedzieć i jak je opowiedzieć. Nasze opowieści są przedane, ale przez większość czasu nie my je przedziemy; to one przędą nas (Dennett 1991, s. 418).

Do wytworzenia „ja” nie wystarczy mózg, który zdaniem Dennetta to – używając w terminologii komputerowej – tylko *hardware*. Aby powstała narracyjna jaźń, musi zostać zainstalowany *software* w postaci odpowiednich kulturowych memów. Podczas gdy mózg jest produktem ewolucji biologicznej, taki memowy *software*, pozwalający na wytworzenie jaźni, jest produktem ewolucji kulturowej. Jaźń wytwarzana jest przez autonarracje, które dokonywane są za pomocą symbolicznego języka. Termin „ja” to pojęcie indeksykalne, natomiast to język pozwala na operowanie takimi pojęciami (Bremer 2005, s. 157).

Narracyjne „ja”, podobnie jak „ja” biologiczne, nie da się zlokalizować w mózgu. Jest ono raczej pewną abstrakcją, zdefiniowaną „przez miriady określeń i interpretacji (włączając samookreślenia i samointerpretacje), które

utworzyły biografię żyjącego ciała, a dla której owa jaźń jest centrum narratywnej ciężkości” (Dennett 1991, s. 427). Opowiadając historie o sobie innym ludziom, robimy to w taki sposób, jakby był w nas ktoś, kto powoduje, że i inni, i my sami jesteśmy przekonani o istnieniu podmiotu. Jednak my sami to nic innego jak centrum narratywnej ciężkości.

Centrum narratywnej ciężkości ma zastosowanie w badaniach nad człowiekiem, podobnie jak w przypadku badań środka ciężkości przedmiotów w fizyce (Dennett 1991, s. 418). Środek ciężkości przedmiotu w fizyce to nie jakiś atom lub jego część ani żadna inna jednostka w świecie, która posiadałaby masę lub inne fizyczne własności. Jedyne, co posiada, to lokalizacja przestrzenno-czasowa (Dennett 1992). Dlatego też można go nazwać teoretyczną fikcją. Mimo to z łatwością można manipulować środkiem ciężkości przedmiotów.

Ludzka jaźń jest podobnym środkiem ciężkości, czyli pewną abstrakcją bądź teoretyczną fikcją. Teoria, która opisuje jaźń, jest fenomenologiczna lub hermeneutyczna (Dennett 1992). Podobnie jak fizyk interpretuje zachowania przedmiotów w świecie, stosując do nich abstrakcyjne pojęcie środka ciężkości, tak fenomenolog czy hermeneuta interpretuje zachowania człowieka, stosując abstrakcyjne pojęcie „jaźni”, która dla Dennetta jest środkiem narratywnej ciężkości. Niemniej obiekt badań fenomenologa jest o wiele bardziej złożony niż obiekty badań fizyka.

Pogląd, że ludzka jaźń jest tylko teoretyczną fikcją, może budzić sprzeciw. Czy nie jest raczej tak, że fikcyjne „ja” są nadbudowane i zależne od jakiegoś realnego „ja”? Dennett uważa taki pogląd za błędny. Aby wykazać jego bezzasadność, proponuje następujący eksperyment myślowy. Wyobraźmy sobie, że istnieje wysoko wyspecjalizowany komputer, który został zbudowany, aby pisać powieści. Jedną z powieści, jaką pisze, jest autobiografia fikcyjnej osoby o imieniu Gilbert. Kiedy komputer pisze: „Jestem Gilbert”, wiemy, że chodzi o fikcyjną postać, nie o sam komputer. Zatem fikcyjny Gilbert posiada fikcyjne „ja”, choć nie jest tak, że jego twórca, czyli komputer, posiada „ja” realne. Następnie wyobraźmy sobie, że piszący powieść komputer jest w rzeczywistości robotem, posiadającym kamerę i koła pozwalające na poruszanie się w przestrzeni. Robot ten nadal pisze powieść o Gilbertcie w pierwszej osobie, ale w pewien sposób przygody Gilberta łączą się z tym, co dzieje się w świecie robota. Na przykład kiedy uderzysz robota kijem bejsbolowym, do historii na temat Gilberta zostaną włączone szczegóły tego, jak Gilbert został uderzony kijem bejsbolowym przez osobnika o twoim wyglądzie. Jeśli pomożesz robotowi, odpowie ci: „Dziękuję, Gilbert”. W takiej sytuacji Gilbert wydaje się być robotem, który jest realnym „ja”. W rzeczywistości wcale tak nie jest – nie mamy bowiem do czynienia z żadnym realnym „ja”. Jedyne, co jest realne, to wzorce zachowań, których częścią jest narracja odwołująca się do „ja”. Dennett uważa, że ten eksperyment

myślowy w bezpośredni sposób pokazuje status naszego własnego „ja” jako teoretycznej fikcji. Nie tylko bowiem Gilbert i komputer nie posiadają prawdziwej jaźni, lecz także my nie posiadamy realnego „ja”.

Koncepcję jaźni jako środka narratywnej ciężkości potwierdzają – zdaniem Dennetta – badania Michaela Gazzanigi nad pacjentami po operacji przecięcia spoidła wielkiego, które prowadzi do wytworzenia dwóch, w dużej mierze niezależnych od siebie, półkul mózgu. Pacjenci po rozszczeniu półkul nie odczuwają jednak rozdwojenia jaźni, lecz mają spójne psychologicznie stany umysłu. Jednak zdaniem Gazzanigi taka spójność zdrowego nawet umysłu to tylko iluzja. W rzeczywistości normalny umysł to powiązanie ze sobą częściowo autonomicznych systemów, które mają czasami wewnętrzne problemy komunikacyjne.

Dlaczego jednak opowiadanie historii sobie samemu jest tak istotne? Nawiązując do książki Juliana Jaynesa pt. *The Origins of Consciousness in the Breakdown of the Bicameral Mind* (1976), Dennett snuje ewolucyjną historię, która ma pokazać, jak doszło do tego, że człowiek zaczął wytwarzać jaźń. Nasi przodkowie porozumiewali się, lecz nie była to świadoma komunikacja, a raczej coś w rodzaju wyrzucania z siebie słów. Komunikacja służyła wówczas wyłącznie do rozwiązywania problemów w środowisku – jedni osobnicy zadawali pytania, inni na nie odpowiadali, aby poradzić sobie z kłopotami, które napotykali. Pewnego dnia jeden z naszych przodków zaczął pytać w dość osobliwych okolicznościach, kiedy nie było wokół żadnego słuchacza. To pozwoliło mu na usłyszenie własnego pytania, a wówczas zaczął sam szukać odpowiedzi, aż w końcu ją znalazł.

Ustalił, bez zdawania sobie z tego sprawy, połączenie komunikacyjne pomiędzy dwoma częściami własnego mózgu, pomiędzy którymi z jakiegoś głębszego biologicznego powodu był problem dostępności. Jeden komponent umysłu został skonfrontowany z problemem, który mógł rozwiązać inny komponent [...] (Dennett 1992).

Wedle tego ujęcia różne moduły umysłu pełnią różne funkcje. Dennett uważa, że wielkim osiągnięciem gatunku ludzkiego jest komunikacja między poszczególnymi modułami umysłu. Zdolność rozmawiania z samym sobą jest korzystna, gdyż pozwala na samodzielne rozwiązywanie problemów. Jednak głośne mówienie do siebie samego i słuchanie własnych słów jest nieefektywne, dlatego w pewnym momencie pojawiły się subwokalizacje, które pozwoliły na rozmawianie ze sobą w ciszy, poprzez werbalne myślenie. Okazało się też, że tego rodzaju myślenie jest korzystne, gdyż pozwala na rozmawianie ze sobą bez udziału potencjalnych słuchaczy, a także na ukrywanie przed innymi tego, co nie chcielibyśmy, by wiedzieli.

Jak jednak zagadnienie jaźni Dennetta wiąże się z zagadnieniem konfabulacji? Zdaniem tego filozofa wszyscy jesteśmy autobiograficznymi nowelistami, którzy tworzą swoją narracyjną biografię, często w sposób bardzo pomysłowy.

Wszyscy jesteśmy czasami konfabulantami, opowiadając i opowiadając na nowo sobie samym historię naszego własnego życia, ze zwracaniem niewielkiej uwagi na kwestię prawdy (Dennett 1992).

Ludzka jaźń to sieć narracyjna, która niekiedy zostaje zerwana, co przyczynia się do wypełniania luk konfabulacjami. Narracja ta potrzebna jest nie nam samym, ale do komunikowania się z innymi w środowisku społecznym.

## 2. Świadomość podmiotu konfabulującego

Zagadnienie konfabulacji wiąże się z kwestią świadomości. Kiedy rozważamy przypadki konfabulowania, pojawiają się pytania dotyczące tego, jaka jest świadomość podmiotu konfabulującego. Skoro konfabulacja to błędnie ugruntowane przekonanie na jakiś temat, podmiot konfabulujący nie ma świadomości, że jego przekonanie jest błędnie ugruntowane, a przy tym nie ma świadomości reprezentacji, którą w danym momencie powinien posiadać. Na przykład pacjent z zespołem Korsakowa nie jest świadomy tego, że nie posiada reprezentacji pamięciowej danego zdarzenia, nie jest też świadomy, że jego aktualna reprezentacja mentalna opiera się na wyobraźni, a nie na pamięci. Z kolei pacjent z zespołem Antona nie wie, że nie posiada reprezentacji percepcyjnej oraz nie jest świadomy tego, że jego reprezentacja nie jest reprezentacją percepcyjną. Osoba z zespołem Capgrasa nie jest zaś świadoma tego, że nie posiada reprezentacji umysłu swojego krewnego.

Podobnie dzieje się w przypadku konfabulacji w zdrowych systemach poznawczych. Osoba, która posiada fałszywe wspomnienia, nie jest świadoma, że jej reprezentacje pamięciowe są zniekształcone bądź w ogóle nie są reprezentacjami pamięciowymi, lecz wyobrażeniami. Według Hirsteina zarówno dane, jak i sprawdzanie danych w tych przypadkach zawodzi (Hirstein 2005).

### 2.1. Heterofenomenologia Dennetta w kontekście zagadnienia konfabulacji

Podstawą każdej nauki jest intersubiektywna komunikowalność. Czy świadomość, która ze swej natury odnosi się do subiektywnych przeżyć, podlega badaniom naukowym? Czy istnieje dla niej opis fizykalistyczny?

Podejście Daniela C. Dennetta stanowi próbę ujmowania świadomości, które byłoby zgodne z naturalistycznym rozumieniem umysłu. Celem filozofa jest zbudowanie metodologii badania zjawisk świadomych z punktu widzenia trzeciej osoby oraz w ramach teorii systemów intencjonalnych.

Jego heterofenomenologia to inaczej naukowe, trzecioosobowe podejście do badań nad świadomością oraz stanów mentalnych innych osób (Dennett 1991, 2003, 2007). Bierze ona pod uwagę eksperymenty, w których osoba badana aktywnie uczestniczy, współpracując z badaczem – poprzez interakcje werbalne, dawanie sugestii, zdawanie raportów werbalnych z tego, „jak to jest odczuwać dany stan”. Metoda heterofenomenologii, jak zauważa Dennett, jest tak oczywista i niekontrowersyjna, że wielu naukowców opierałoby się, by w ogóle nazywać ją metodą (Dennett 2003). Jej celem jest bliższe przyjrzenie się raportom werbalnym osoby badanej w sytuacji eksperymentalnej oraz skatalogowanie tego, „co do czego osoba badana jest przekonana, iż jest prawdziwe względem jej własnych świadomych doświadczeń” (Dennett 2003, s. 2). Filozof ten zaznacza, że metoda heterofenomenologii nie jest niczym nowym. Każdy badacz, który kiedykolwiek w sposób naukowy podchodził do zagadnienia świadomości, posługiwał się nią – zarówno psychofizjolog, psycholog poznawczy, jak i neuropsycholog analizuje raporty werbalne osób badanych będące opisem ich świadomych doświadczeń. Dzieje się tak m.in. dlatego, że człowiek jako jedyny obiekt nauki posiada zdolność werbalnej komunikacji (Dennett 1991, s. 73).

Podstawowymi danymi heterofenomenologii są wypowiedzi osób badanych. Stanowią one surowe dane, które muszą być następnie zinterpretowane jako sądy werbalne wyrażające przekonania na temat świadomych doświadczeń osób badanych. Przekonania na temat własnych doświadczeń nie są tym samym co doświadczenia, jednak Dennett odpira zarzut, jakoby podstawowymi danymi dla badacza świadomości były surowe świadome doświadczenia podmiotu:

[...] jeśli masz doświadczenia świadome, co do których nie jesteś przekonany, że je masz – te dodatkowe świadome doświadczenia są po prostu niedostępne dla nas, jako zewnętrznych obserwatorów [...], jeśli jesteś przekonany, że masz świadome doświadczenia, których w rzeczywistości nie masz – wtedy musimy wyjaśnić twoje przekonania, a nie nieistniejące doświadczenia (Dennett 2003, s. 3).

Problem, który porusza Dennett, jest niezwykle ważny dla zagadnienia konfabulacji. Przynajmniej niektóre rodzaje konfabulacji to przekonania podmiotu na temat tego, że posiada on jakieś świadome doświadczenie (percepcyjne, pamięciowe), podczas gdy w rzeczywistości doświadczenie to nie może mieć miejsca. Jednym z celów heterofenomenologii jest więc wyjaśnienie błędnych przekonań podmiotu na temat własnych doświadczeń.

Dennett zauważa ponadto, że w wielu sytuacjach eksperymentalnych osoby badane konfabulują, nie zdając sobie z tego sprawy (Dennett 1991). Jego zdaniem świadczy to o tym, że podmiot nie zawsze ma pełny dostęp do tego, co dzieje się w jego umyśle. Zatem introspekcyjne raporty werbalne nie mogą być tożsame z raportami werbalnymi na temat widzianych zdarzeń.

I ta możliwość w nas pokazuje, dlaczego musimy ominąć kłopot traktowania heterofenomenologii jako analogicznej do interpretacji fikcji. [...] są okoliczności, w których ludzie są po prostu w błędzie co do tego, co wykonują i jak to robią. To nie dlatego, że kłamią w sytuacji eksperymentalnej, ale konfabulują; uzupełniają luki, zgadują, spekulują, myślą się co do obserwacji [...] nie mają żadnego sposobu „widzenia” (prawdopodobnie wewnętrznym okiem) procesów, które rządzą ich twierdzeniami, ale to nie powstrzymuje ich od szczerego wyrażania opinii (Dennett 1991, s. 94).

Metoda heterofenomenologii składa się z dwóch etapów: najpierw badany składa raport werbalny ze stanu własnych świadomych przeżyć (może być nagrany bądź spisany), następnie badacz interpretuje otrzymany tekst. Etap interpretacji jest tu najistotniejszy, gdyż stanowi opis świata z perspektywy trzeciej osoby. Podczas tego etapu należy pamiętać, aby traktować podmiot jako kogoś, kto doświadcza percepcji, ma pragnienia, przekonania i inne intencjonalne stany mentalne, które wskazują na to, że podmiot działa na ich podstawie. Ważne jest także to, że status metafizyczny opisywanych obiektów jest taki sam jak status obiektów fikcyjnych. Zdaniem Dennetta nie powinniśmy więc pytać o to, czy obiekty opisywane przez podmiot istnieją rzeczywiście.

Heterofenomenologiczne obiekty są, tak jak środek ciężkości lub równik, abstrakcjami, nie konkretami. Nie są jałowymi fantazjami, lecz ciężko pracującymi fikcjami teoretyków (Dennett 1991, ss. 95-96).

Dennett uznaje takie podejście za neutralne, nazywając je heterofenomenologicznym minimalizmem metafizycznym (Dennett 1991).

Heterofenomenologia jest metodą współczesnego introspekcjonizmu (Miłkowski 2005, s. 103). Jednakże założenia współczesnego introspekcjonizmu odbiegają znacznie od tradycyjnie ujmowanej introspekcji.

Nie twierdzi się, że podmioty mają lepszy dostęp do własnych przeżyć niż do obiektów fizycznych [...]. Dopuszcza się korygowalność danych introspekcyjnych, unika się solipsystycznej teorii introspekcji; podkreśla się, że nie wszystkie czynności psychiczne (w szczególności ich przyczyny i skutki) są dostępne introspekcyjnie. Nie twierdzi się, że istnieje jakaś odrębna czynność polegająca na „patrzeniu wewnątrz” ani nie przyjmuje się teorii percepcji czy spostrzeżenia wewnętrznego, a tym bardziej nie traktuje się introspekcji jako podstawy dla przewyciężenia

sceptycyzmu na temat istnienia świata zewnętrznego. Podkreśla się też, że wydanie sądu introspekcyjnego następuje zawsze później od rozegrania się odpowiednich procesów w umyśle (Miłkowski 2005, s. 104).

### 2.1.1. Orwellowski i stalinowski model świadomości a teatr kartezyjański

Według Dennetta świadomość to nie jednolity proces czy jednoznaczny strumień, lecz wiele strumieni, w których poszczególne moduły realizują zadania, edytując i reedytując dane tak, że tworzy się wiele szkiców. Zgodnie z jego poglądem nie istnieje żadna „kwatery główna” czy kartezyjański teatr<sup>2</sup> o podłożu anatomicznym, który decydowałby o tym, który szkic dochodzi do świadomości w danym momencie.

Model wielu szkiców został zaproponowany przez Dennetta m.in. w związku z opisowymi problemami interpretacyjnymi pewnych eksperymentów, takich jak „zjawisko pozornego ruchu” (fenomen phi) (Dennett 1991, s. 120). Eksperyment polega na tym, że badanemu pokazuje się dwa świetlne punkty: czerwony i zielony, leżące blisko siebie, które na przemian pojawiają się i znikają. Badanemu wydaje się jednak, że to pojedynczy punkt przemieszcza się tam i z powrotem, zmieniając kolor w połowie drogi. Pojawia się więc pytanie, w jaki sposób badany wypełnia kolorem drogę biegnącą od jednego światła do drugiego, zanim zablśnie druga lampka (Dennett 1991, s. 114).

Przy interpretacji raportów werbalnych badanych (że ruch zachodzi) problematyczne jest to, czy uzupełnianie następuje na poziomie uzupełniania luk w aktualnym spostrzeżeniu pomiędzy wrażeniami wzrokowymi, tak iż wydają się być one ciągłe (model stalinowski), czy też na poziomie przypominania spostrzeżenia (model orwellowski).

Modele stalinowski i orwellowski mają na celu przybliżenie procesów powstawania iluzji w systemie poznawczym człowieka. Odzwierciedlają one totalitarne praktyki okłamywania społeczeństw. Interpretacja orwellowska jest odzwierciedleniem tego, czego dokonywało Ministerstwo Prawdy w powieści *Rok 1984* George'a Orwella, czyli fałszowania przeszłości na ogromną skalę. Z kolei interpretacja stalinowska nawiązuje do praktyk Stalina i polega na przedstawianiu fałszywych świadectw, zanim coś się wydarzy. Mózg może więc stosować albo model stalinowski, albo model orwellowski. Pierwszy z nich powoduje, że następuje uzupełnianie luk pomiędzy wrażeniami wzrokowymi,

---

<sup>2</sup> Teatr kartezyjański to pojęcie wprowadzone do filozofii umysłu na określenie naszego własnego świata prywatnego, w którym każdy jest swoim reżyserem. Teatr kartezyjański w wersji materialistycznej odnosi się do poglądu, zgodnie z którym istnieje ośrodek mózgu zbierający informacje z innych ośrodków i na tej podstawie podejmujący decyzje i sterujący zachowaniem.

tak iż powstaje jeden ciągły obraz; drugi zaś dokonuje rewizji uświadomionych wrażeń wzrokowych, tak iż następuje ich połączenie, dzięki czemu widzimy punkty świetlne jako jeden przemieszczający się punkt.

Obydwa modele posługują się zafalszowaniami, lecz dzieje się to w inny sposób. W modelu stalinowskim zafalszowanie dokonywane jest na poziomie postrzegania – zmieniane jest samo przeżycie percepcyjne przez uzupełnienie luki pomiędzy wrażeniami wzrokowymi. Na bieżąco edytowane jest samo spostrzeżenie, dzięki czemu zafalszowana informacja jest antycypowana. W modelu orwellowskim zafalszowanie dokonywane jest niejako *ex post*, retrospekcyjnie, już po spostrzeżeniu. Edycja polega tu na wymazaniu pierwotnego spostrzeżenia i zapisania na nowo treści pamięci.

Jednak według Dennetta to, czy mózg stosuje fałszywą korektę orwellowską, czy fałszywą korektę stalinowską, nie jest odróżnialne ani z perspektywy trzeciej osoby ani z perspektywy pierwszej osoby (Dennett 1991, s. 123). Oba modele są do pogodzenia z tym, co przeżywa podmiot – z tym, co mówi, myśli czy przypomina sobie, a my nie mamy żadnej eksperymentalnej możliwości sprawdzenia, która z hipotez jest właściwa. Na poziomie empirycznym nie da się odróżnić tych modeli. Ponadto sam podmiot, z perspektywy pierwszej osoby, nie jest w stanie odróżnić, czy zastosowana została korekta antycypacji, czy retrospekcji.

Ty, jako podmiot podczas eksperymentu z fenomenem phi, nie potrafisz z twojej własnej pierwszoosobowej perspektywy odkryć nic w doświadczeniu, co stawiałoby jedną teorię nad drugą; przeżycie byłoby „odczuwane tak samo” w obydwu przypadkach (Dennett 1991, s. 123).

Ilustrację niemożności odróżnienia modelu orwellowskiego od modelu stalinowskiego stanowi eksperyment myślowy Dennetta. Filozof każe nam sobie wyobrazić, że ktoś zaingerował w nasz mózg, wkładając do naszej pamięci fałszywą reprezentację kobiety ubranej w kapelusz, np. podczas przyjęcia w niedzielę. W poniedziałek, kiedy próbujemy sobie przypomnieć niedzielne przyjęcie, przywołujemy w pamięci kobietę w kapeluszu, ale nie znajdujemy żadnych wewnętrznych zasobów na potwierdzenie tego, więc zaczynamy wątpić w wiarygodność pamięci. Możemy wówczas przyjąć, że nie widzieliśmy kobiety w kapeluszu na niedzielnym przyjęciu. Następnie wyobraźmy sobie, że we wtorek przywołujemy w pamięci kobietę w kapeluszu i jest to reprezentacja pamięciowa bardzo wyraźna. Powiemy wówczas, że mamy wyraźne świadome doświadczenie związane z tym, że kobieta w kapeluszu pojawiła się na przyjęciu, ale doświadczenie to po raz pierwszy pojawiło się w poniedziałek, a nie w niedzielę (Dennett 1991, s. 116).

Dennett zwraca uwagę na to, że musimy odwołać się do nieistniejących obecnie neurochirurgicznych metod implantacji fałszywych wspomnień, by przeprowadzić taki eksperyment myślowy. Nasz umysł w codziennym życiu funkcjonuje w ten sposób, że pamiętamy rzeczy, które się nie przytrafiły i jest to dość częste. Dennett nazywa takie korekty pamięci czy poprzężyciowe zanieczyszczenia orwellowskimi rewizjami. Ponieważ zdajemy sobie sprawę z możliwości orwellowskich rewizji, rozpoznajemy ryzyko inferencyjnego przejścia od pamięci do rzeczywistych zdarzeń.

Oprócz rewizji orwellowskich istnieją też sztuczki stalinowskie, które polegają na preparowaniu fałszywych zeznań oraz symulowaniu dowodów. Dennett podkreśla, że to, iż na co dzień wydaje nam się, że wiemy, którą ze sztuczek: orwellowską czy stalinowską zastosował nasz umysł, dzieje się przez przypadek.

W jakiegokolwiek pomyślanej kampanii dezinformacyjnej zastanawialibyśmy się, czy relacje w gazetach były orwellowskimi sprawozdaniami zdarzeń, które nigdy nie miały miejsca, czy prawdziwymi relacjami fałszywych procesów pokazowych, które się w rzeczywistości się zdarzyły, możemy być niezdolni, by dokonać takiego rozróżnienia (Dennett 1991, s. 117).

Według Dennetta w życiu codziennym czasem trudno odróżnić, czy np. dowody sfabrykowano przed oskarżeniem kogoś, czy po oskarżeniu.

W większych skalach czasowych różnice między orwellowskim i stalinowskim modelem są łatwo wychwytywane. Jednak kolejny eksperyment myślowy Dennetta pokazuje, jak trudno wychwycić różnice między tymi modelami. Wyobraźmy sobie, że stoimy na rogu ulicy, a obok przebiega kobieta w długich włosach i bez okularów. Chwilę potem wspomnienie wcześniej widzianej kobiety w okularach zanieczyszcza pamięć tego, co widzieliśmy przed momentem. Zapytani, co widzieliśmy, relacjonujemy błędnie, że widzieliśmy kobietę w okularach. Mamy tu do czynienia z rewizją orwellowską – najpierw widzieliśmy kobietę bez okularów, jednak nagle przypomniana kobieta w okularach zanieczyściła naszą pamięć tak, iż przypominamy sobie kobietę w okularach. Jest to swego rodzaju złudzenie pamięciowe.

Możliwe jest jednak inne wyjaśnienie tego zjawiska. Wcześniejsze wspomnienie kobiety w okularach równie dobrze mogło zaingerować nie w pamięć, ale w bieżące świadome doświadczenie przebiegającej kobiety, tak iż od samego początku mieliśmy wrażenie, że widzieliśmy kobietę w okularach. W tej sytuacji mamy do czynienia ze sztuczką stalinowską – ulegamy nie złudzeniu pamięciowemu, lecz percepcyjnemu, a następnie przypominamy sobie wcześniejsze złudzenie.

Według Dennetta w mniejszych skalach czasowych modele stalinowski i orwellowski są nieodróżnialne; świadczy o tym nie tylko powyższy ekspery-

ment myślowy, ale także zjawisko phi. W zjawisku phi nie można mówić ani o uzupełnianiu antycypacyjnym, ani o uzupełnianiu retrospekcyjnym. Ma to związek z koncepcją świadomości jako modelu wielokrotnych szkiców. Zdaniem Dennetta, aby można było mówić o uzupełnianiu, trzeba zaakceptować istnienie teatru kartezjańskiego. Tymczasem nie istnieje żaden punkt w mózgu, przez który przechodzą wszystkie informacje, taki mianowicie punkt, w którym tego rodzaju uzupełnianie mogłoby zachodzić.

W książce *Consciousness Explained* Dennett odrzuca koncepcję teatru kartezjańskiego. Jest on traktowany jako pewien zbiór założeń dotyczących natury świadomości. Niekoniecznie musi być związany z dualizmem psychofizycznym, może także dotyczyć poglądów materialistycznych i wówczas okreśłany jest mianem materializmu kartezjańskiego (Dennett 1991, s. 107). Świadomość traktowana jest tu jako scena, na której rozgrywają się świadome zdarzenia. Scena ta to pewne miejsce w przestrzeni mentalnej (dualizm) lub w mózgu (materializm), gdzie dochodzi do zdarzenia świadomego.

Dennett krytykuje współczesne podejścia z dziedziny kognitywistyki, które jego zdaniem cechuje materializm kartezjański. Podejścia te zakładają, że istnieje jakieś miejsce w mózgu, gdzie wszystkie informacje się schodzą. W neuro-naukach najczęściej uznaje się, że są to: przednia część zakrętu obręczy, układ siatkowaty lub różne miejsca w płatach czołowych (Dennett 1991, s. 107).

Kartezjański materializm jest poglądem, że istnieje jakaś rozstrzygająca linia końcowa lub granica gdzieś w mózgu, tworząca miejsce, gdzie porządek nadejścia [informacji] jest równoważny z porządkiem „prezentacji” w doświadczeniu, ponieważ to, co się dzieje, to to, czego jesteś świadomy (Dennett 1991, s. 107).

Charakterystyczną cechą zarówno dualizmu, jak i materializmu kartezjańskiego jest to, że ujęcia te prowadzą do pojęcia homunkulusa i związanego z nim regresu do nieskończoności. Istnienie miejsca, w którym zbiegają się informacje i dochodzi do świadomego zdarzenia, zakłada istnienie obserwatora tych zdarzeń bądź świadomego podsystemu w mózgu. Jeśli jednak taki podsystem jest świadomy, to musi mieć wbudowanego kolejnego homunkulusa, czyli kolejny świadomy system. Taki eksperyment myślowy można przeprowadzać w nieskończoność.

Koncepcja teatru kartezjańskiego zakłada ponadto, że umysł przetwarza informacje w sposób szeregowy, a nie równoległy. Jednak wyniki współczesnych badań kognitywistycznych pokazują, że przetwarzanie ma charakter raczej równoległy niż szeregowy. Jednym z takich eksperymentów jest badanie zjawiska phi. Zjawisko to można wyjaśnić tylko wówczas, gdy przyjmiemy, że mózg przetwarza informację w sposób równoległy. Oznacza to, że nie istnieje jeden uporządkowany szeregowo strumień świadomości, dostępny dla homunkulusa. Dlaczego zatem świadomość wydaje nam się takim jednolitym strumieniem?

Dennett argumentuje, że istnieją mechanizmy mózgowe, które są odpowiedzialne za wytwarzanie poczucia jednorodności strumienia świadomości.

Model wielokrotnych szkiców zakłada, że świadomość nie jest procesem jednolitym, lecz przypomina edytowanie wielu szkicowych wersji artykułu. Wyobraźmy sobie – proponuje Dennett – że takie szkicowe wersje krążą gdzieś rozsyłane drogą mailową pomiędzy specjalistami, a autor wciąż je poprawia. W pewnym momencie nie wiadomo, która wersja jest ostateczna i poprawna, bo nie wszystkie poprawki zostały wszędzie naniesione. Arbitralnym wyborem jest ostateczna publikacja jednego ze szkiców jako kanonicznego tekstu (Dennett 1991, s. 125).

Podobnie świadomość jest zmiennym strumieniem wciąż edytowanych i reedytowanych szkiców, w którym trudno ustalić moment publikacji. Szkice te czasem są zapamiętywane, czasem inicjują wypowiedzi i działania, a czasem odchodzą w niepamięć. Docierające do mózgu informacje na bieżąco są edytowane. Błędnym poglądem na naturę świadomości jest to, że istnieje moment przetwarzania informacji w mózgu, w którym świadomość dochodzi do głosu.

Według Williama Hirsteina, aby mogła powstać konfabulacja kliniczna, konieczne są dwa etapy przetwarzania informacji w mózgu: brak danych w jednym z systemów wiedzy, a następnie tworzenie reprezentacji, której błędność nie zostaje wychwycona przez płaty czołowe mózgu. Taka architektura umysłu nazbyt przypomina jednak założenia materializmu kartezjańskiego.

## 2.2. Wykonawcze teorie konfabulacji pamięciowych a paradoks homunkulusa

W artykule pt. *Temporal consciousness and confabulation: Escape from unconscious explanatory idols* (2009) Gianfranco Dalla Barba polemizuje z teoriami konfabulacji pamięciowych, opierającymi się na założeniu istnienia mechanizmów monitorujących. Wiele współczesnych teorii konfabulacji pamięciowych odwołuje się do procesów monitorujących (Hirstein 2005; Schnider 2008). Teorie te nazywane są wykonawczymi ze względu na to, że postulują, iż funkcje wykonawcze pełnione przez płaty czołowe sprawiają, że dana reprezentacja dochodzi do świadomości bądź jest odrzucana. W zdrowym systemie poznawczym procesy monitorujące najczęściej wychwytyują nieprawidłowe względem kontekstu reprezentacje i je odrzucają. Natomiast poprawne względem kontekstu reprezentacje są podtrzymywane. Uszkodzenie mechanizmów monitorujących prowadzi do niezdolności wyhamowywania nieprawidłowych reprezentacji, co skutkuje tym, że osoba zaczyna konfabulować.

Istnieją dwie warte odnotowania teorie wykonawcze konfabulacji pamięciowych. Pierwsza z nich to teoria Marcii Johnson, która głosi, że konfabulacje są skutkiem uszkodzenia mechanizmu monitorowania rzeczywistości (za: Hirstein 2005). Mechanizm ten polega na odróżnianiu reprezentacji, które są jedynie wyobrażeniowe, od tych, które są rzeczywistymi wspomnieniami podmiotu. Uszkodzenie go prowadzi do niezdolności monitorowania, które reprezentacje są rzeczywistymi wspomnieniami, a które tylko wyobrażeniami. W ten sposób dochodzi do konfabulacji. Druga teoria to teoria Armina Schnidera, która głosi, że konfabulacje powstają w wyniku niezdolności wyhamowywania wspomnień nieodpowiednich do kontekstu. Jedną z funkcji płatów czołowych jest monitorowanie, które wspomnienia są odpowiednie, a które nie. Mechanizm monitorujący wychwytuje wspomnienia, które są istotne dla aktualnego przypominania. Na przykład kiedy próbuję sobie przypomnieć, co robiłam dwa dni temu w południe, mechanizm monitorujący przeszukuje w pamięci różne zdarzenia z ostatnich dni i wyhamowuje te, które są nieodpowiednie. Dzięki temu dochodzi do świadomości reprezentacja, która informuje, że dwa dni temu w południe byłam na seminarium. Teorie monitorowania źródła pamięci są także powszechne w wyjaśnianiu konfabulacji w zdrowych systemach poznawczych (French, Garry i Loftus 2009).

Dalla Barba uważa, że zasadne jest pytanie, czy procesy monitorujące to procesy świadome, czy nieświadome. Jego zdaniem zarówno stwierdzenie, że są to procesy świadome, jak i nieświadome prowadzi do paradoksów. Wyobraźmy sobie, iż procesy monitorujące to procesy świadome. Kiedy próbujemy przypomnieć sobie jakieś zdarzenie z przeszłości, nie jest tak, że przypominamy sobie wiele innych zdarzeń, które eliminujemy, by dotrzeć do właściwego wspomnienia, ale wspomnienie od razu przywoływane jest w pamięci. Na przykład gdy próbuję sobie przypomnieć, co robiłam wczoraj wieczorem, od razu przywołuję w pamięci zajęcia z angielskiego oraz pisanie artykułu. Nie muszę świadomie monitorować różnych wspomnień, żeby dojść do właściwego. Czasem tylko zdarza się, że świadomie przeszukujemy pamięć, aby dotrzeć do poprawnej reprezentacji pamięciowej. Na przykład próbując sobie przypomnieć, co robiliśmy dwa tygodnie temu wieczorem, najpierw przypominamy sobie, co robiliśmy rano i w południe, a potem przypominamy sobie, że wieczorem oglądaliśmy film w kinie. Możemy też przypominać sobie, co robiliśmy podczas wszystkich wieczorów poprzedniego tygodnia, aby poprzez stopniowe przybliżenia dotrzeć w pamięci do tego wieczora, o który nam chodzi. Jak zauważa Dalla Barba, ten proces selekcji nie dotyczy jednak prawdziwości pamięci, czyli źródła reprezentacji mentalnej, lecz jego umiejscowienia w czasie lub przestrzeni (Dalla Barba 2009, s. 235). Czasem, choć dzieje się to raczej rzadko, możemy wahać się co

do statusu naszej reprezentacji mentalnej – zastanawiamy się, czy pochodzi ona z pamięci, wyobraźni, czy np. ze snu. O wiele częściej jednak wahamy się, czy zdarzenie, które pamiętamy, zaszło w czasie, w którym je umiejscawiamy. Na przykład przypominamy sobie rozmowę ze znajomym, ale nie wiemy, czy miała ona miejsce dwa, czy trzy miesiące temu. Proces selekcji i weryfikacji wspomnień nie może zatem, z małymi wyjątkami, przebiegać w sposób świadomy.

Warto jednak się zastanowić, czy monitorowanie przebiega w sposób nieświadomy, jak postulują zwolennicy wykonawczych teorii konfabulacji pamięciowych. Zdaniem Dalla Barby przyjęcie założenia o nieświadomych procesach monitorujących prowadzi do paradoksu homunkulusa. Jeśli bowiem jakieś procesy monitorują naszą pamięć i selekcionują wspomnienia nieodpowiednie do kontekstu, to muszą posługiwać się jakąś świadomością, która pozwala na selekcionowanie, ewaluację i odrzucanie fałszywych wspomnień. Taka świadomość nieświadomości pozwalałaby na wyselekcjonowanie odpowiedniego wspomnienia i przekazania go „świadomej świadomości” (Dalla Barba 2009, s. 236). Założenie to prowadzi jednak do przyznania procesom monitorującym intencjonalności, co implikuje przypisanie im podmiotowości. Należy zauważyć, że taka świadoma nieświadomość, mając charakter oksymoronu, byłaby niedostępna świadomości w ścisłym sensie. W ten sposób powstaje paradoks homunkulusa – nieświadome procesy monitorujące są swego rodzaju homunkulesem umieszczonym w naszym mózgu po to, by mógł przeglądać nasze wspomnienia i wybierać te, które są odpowiednie do kontekstu, a następnie dostarczać je świadomości. Paradoks ten sprawia, że aby odróżnić wspomnienia prawdziwe od fałszywych, procesy monitorujące musiałyby posługiwać się jakimiś kryteriami odróżniania. Aby tego dokonać, homunkulus musiałby reprezentować sam dla siebie wspomnienia zarówno prawdziwe, jak i fałszywe oraz rozpoznawać fałszywe wspomnienia jako te, które mają być odrzucone. Musiałby więc nie tylko posiadać świadome reprezentacje wspomnień, które mają być odrzucone, ale także być świadomym tego, co musi robić, aby poprawnie wykonać swoje zadanie (Dalla Barba 2009, s. 247). Inaczej mówiąc, musiałby być samoświadomy. Takie założenie, zdaniem Dalla Barby, prowadzi do wniosku, że homunkulus sam potrzebuje wewnętrznego homunkulusa, co prowadzi do regresu w nieskończoność.

Dalla Barba proponuje inne wyjaśnienie konfabulacji pamięciowych niż wyjaśnienia opierające się na założeniu istnienia procesów wykonawczych. Proponuje on naukową teorię opisującą relację między świadomością a pamięcią, zwaną teorią MCTT (*Memory, Consciousness and Temporality Theory*). Teoria ta unika jego zdaniem problematycznego paradoksu homunkulusa pojawiającego się wraz z założeniem o istnieniu nieświadomych procesów monitorujących.

### 2.2.1. Świadomość czasu, świadomość wiedzy i ich znaczenie dla zagadnienia konfabulacji pamięciowych

Dalla Barba przyjmuje, że świadomość posiada cechę intencjonalności, czyli zawsze jest świadomością czegoś. Świadomość jest świadomością czegoś w pewien sposób, to znaczy zawsze obiera punkt widzenia swojego obiektu. Wynika z tego, że istnieją różne formy świadomości, z których każda jest wyjątkowa i nie da się sprowadzić do czegoś innego. Na potrzeby argumentacji dotyczącej konfabulacji Dalla Barba wyróżnia dwie nieredukowalne do siebie formy świadomości: świadomość wiedzy (*Knowing Consciousness* – KC) i świadomość czasu (*Temporality Consciousness* – TC).

Świadomość wiedzy jest pod pewnymi względami podobna do pamięci semantycznej, z tym że nie opiera się na nieświadomej reprezentacji mentalnej. Jest to świadomość obiektu w pewnym trybie, który czyni niemożliwym pomylenie tego, co wiem, z tym, co sobie przypominam lub tylko wyobrażam. Świadomość wiedzy jest w pewnym sensie bezczasowa, gdyż czas, w którym powstała owa wiedza, nie jest przez nas rozpoznawalny.

KC jest przeszłością, ale nie jest to świadomość przeszłości ani teraźniejszości, ani przyszłości (Dalla Barba 2009, s. 250).

Świadomość czasu to z kolei tematykacja obiektu w trybie czasowym.

Inaczej niż KC, TC wykracza poza zwykłą prezentację obiektu w celu umiejscowienia go w czasie (Dalla Barba 2009, s. 250).

Kiedy przypominam sobie, że wczoraj byłam na seminarium, obecnie siedzę w pokoju i piszę pracę, a wieczorem pójdę do kina, zakładam pewną wiedzę: czym jest kino, praca, seminarium, jednak sama wiedza nie jest celem aktu temporalizacji, który jest przeze mnie dokonywany. Wiedza jest jedynie pewną strukturą, na której akt temporalizacji jest ufundowany.

To *poprzez* wiedzę akt temporalizacji jest dokonywany, lecz nie *dla* wiedzy jest on realizowany (Dalla Barba 2009, s. 250).

Przeszłość, teraźniejszość i przyszłość są pewnym kontinuum, jednak dla pewnej jasności świadomość przeszłości, teraźniejszości i przyszłości należy traktować jako podrzędne struktury TC i rozważać je oddzielnie.

Świadomość przeszłości to inaczej przypominanie sobie. Jest to akt, w którym obiekt świadomości widziany jest jako nieobecny w przeszłości. W przeciwieństwie do świadomości wyobrażeniowej nie jest to widzenie obiektu nieobecnego jako nieistniejącego. Świadomość przeszłości odnosi się do personalnej przeszłości, która jest szczególna dla każdego. Nie dotyczy faktów

ogólnych, takich jak wiedza o tym, kto odkrył Amerykę czy też kto napisał *Moby Dicka*, gdyż tego rodzaju informacje dotyczą ogólnej przeszłości, a nie naszej własnej. Nasza zdolność przypominania sobie jest możliwa dzięki temu, że istnieje pewna łączność między moim obecnym „ja” i przeszłym „ja”. Świadomość przeszłości możliwa jest dzięki temu, że istnieje jakaś terażniejszość, względem której przeszłość jest przeszłością (Dalla Barba 2009, s. 250). Z kolei świadomość terażniejszości „reprezentuje tematyzację percepcji w trybie terażniejszości” (Dalla Barba 2009, s. 251). Terażniejszość rozumiana jest tu jako subiektywna czy też fenomenologiczna terażniejszość, która nie może być mierzona za pomocą obiektywnego czasu.

Zatem świadomość terażniejszości nie jest momentem ani też czymś, co już przeminęło, ale jest relacją pomiędzy podmiotem i jego obiektem w trybie prezentacji. Świadomość terażniejszości jest aktem, który temporalizuje świat w formie obecności [...] (Dalla Barba 2009, s. 251).

Świadomość przyszłości możliwa jest z kolei wówczas, gdy obiekt świadomości sytuowany jest w przyszłości.

Jest to tematyzacja mojej możliwości w trybie przyszłości. Przez „możliwość” rozumiem to, co jest ufundowane w wiedzy o mojej przeszłości i terażniejszości i to radykalnie odróżnia świadomość przyszłości od życzeń (Dalla Barba 2009, s. 251).

Świadomość czasu to zatem „możliwość czasowej egzystencji podmiotu” (Dalla Barba 2009, s. 250). Aby wyjaśnić, na czym polega związek TC i KC z konfabulacjami, Dalla Barba przedstawia operacje, jakie jego zdaniem dokonywane są na świadomości czasu i świadomości wiedzy.

Świadomość jest zawsze świadomością jakiegoś obiektu. Obiekt świadomości może być charakteryzowany jako ten, który „reprezentuje i odkrywa dla świadomości swoją unikatowość (U) lub mnogość (M)” (Dalla Barba 2009, s. 252).

To pióro leżące na blacie stołu jest jednocześnie piórem w ogóle i tym piórem. W pierwszym przypadku jest nieokreślonym piórem, czymś, co należy do kategorii „pióra”, obiektem, który rozpoznaję i używam odpowiednio, ponieważ go rozpoznaję. W przeciwieństwie do tego, w drugim przypadku „to pióro” jest określonym obiektem, jest to dokładnie to pióro naprzeciw mnie [...]. Tak więc to pióro odkrywa U oraz M. M odzwierciedla się w tym, że jest to pióro, a nie jakiś inny obiekt. U manifestuje się w byciu tym dokładnie piórem, a nie jakimś innym (Dalla Barba 2009, s. 252).

Dalla Barba argumentuje, że unikatowość obiektu jest odkrywana przez świadomość działającą w trybie TC. Świadomość czasu jest bowiem tym aktem świadomości, który pozwala na wybór obiektu jako tego, na który patrzę,

który sobie przypominam bądź którego oczekuję. Przejście od mnogości do unikalności dokonuje się dzięki temu, że świadomość czasu wykracza poza świadomość wiedzy. W świadomości wiedzy odkrywamy mnogość obiektu, w świadomości czasu zaś – jego unikalność. Przypominanie polega więc na przywoływaniu w świadomości unikalności obiektu.

Postrzegane zdarzenia dokonują beczasowych i niespecyficznych modyfikacji w mózgu. Modyfikacje są beczasowe, gdyż nie zawierają informacji na temat czasu, a niespecyficzne dlatego, że nie zawierają informacji na temat tego, czy reprezentują epizody, znaczenia, procedury, czy reguły. Modyfikacje te mogą być mniej lub bardziej stabilne oraz mniej lub bardziej wrażliwe w zależności od uwagi poświęconej kodowaniu informacji emocjonalnej wartości zdarzenia, głębokości kodowania czy powtórnego doświadczania tego samego zdarzenia.

Mniej stabilne i bardziej wrażliwe wzorce modyfikacji mózgu są konieczne, ale niewystarczające do interakcji pomiędzy TC oraz U obiektu, podczas gdy bardziej stabilne i mniej wrażliwe modyfikacje mózgu są konieczne, ale niewystarczające do interakcji pomiędzy KC i M obiektu (Dalla Barba 2009, ss. 254-255).

Zdaniem Dalla Barby w normalnie funkcjonującym systemie poznawczym mniej stabilne modyfikacje korelują więc ze świadomością czasu, która odkrywa unikalność obiektu, natomiast bardziej stabilne – ze świadomością wiedzy, która odkrywa mnogość obiektu.

W przypadkach klinicznych konfabulacji pamięciowych mamy jednak do czynienia z sytuacją nietypową. Pacjenci, którzy konfabulują, nadal posiadają świadomość czasu, o czym świadczy to, że starają się przypomnieć sobie przeszłość, są zorientowani na teraźniejszość oraz planują przyszłość, choć we wszystkich tych trzech czynnościach popełniają błędy, co prowadzi do konfabulacji. Zdaniem Dalla Barby tego rodzaju sytuację poznawczą można opisać następująco: TC nadal istnieje u tych pacjentów, ale nie jest skorelowane z mniej stabilnymi wzorcami modyfikacji mózgu.

Zgodnie z MCTT, w tym przypadku TC oddziałuje z bardziej stabilnymi wzorcami modyfikacji mózgu oraz zwraca się do M obiektu, przyzwyczajień, rutyny, powtarzalnego działania, jako U, specyficznego, unikalnego przeszłego zdarzenia (Dalla Barba 2009, s. 256).

To tłumaczy, dlaczego pacjenci zapytani o to, co robili wczoraj albo co planują robić jutro, często wymieniają rutynowe działania, jakie zwykli wykonywać, zanim zaczęli konfabulować. Zdarza się bowiem, że treść konfabulacji pacjenów jest prawdopodobna. Okazuje się, że bardziej stabilne modyfikacje mózgu, które zwykle korelują tylko z KC, u pacjentów konfabulujących oddziałują także z TC, które zwraca się do mnogości obiektu jako jego unikalności.

Dalla Barba zauważa, że przypadki konfabulacji prawdopodobnych nie wyczerpują klasy konfabulacji pamięciowych. Istnieją pacjenci, których konfabulacje nie opierają się na opowieściach o rutynowych i codziennych zdarzeniach, które mogły mieć miejsce w przeszłości pacjenta. Czasem treść konfabulacji wykracza poza codzienną rutynę i jest związana z fantazjowaniem na temat własnej przeszłości i przyszłości. Czy także w takich przypadkach można zastosować teorię MCTT Dalla Barby?

Filozof twierdzi, że nieprawdopodobne i kwieciste konfabulacje pacjentów nie są związane z myleniem przeszłych wspomnień, lecz tworzone są *ex nihilo*. Nie można więc stosować do ich wyjaśnienia koncepcji Schnidera, który twierdzi, że konfabulacje powstają w wyniku niezdolności wyhamowywania wspomnień nieodpowiednich do kontekstu. Nie można też zastosować w tym przypadku teorii Marcii Johnson, która utrzymuje, że konfabulacje są wynikiem mylenia rzeczywistości z wyobraźnią. Konfabulacje tego rodzaju można jednak wyjaśnić za pomocą teorii MCTT. W przypadku konfabulacji, które są fantazjami, świadomość czasu nadal jest zachowana. TC jest świadomością czegoś i temporalizuje swój obiekt ze względu na przeszłość, teraźniejszość lub przyszłość:

[...] nie ma potrzeby pytać, skąd pochodzi obiekt temporalizacji, gdyż nie pochodzi on z żadnego nieświadomego świata. Nie jest on poza świadomością, jest on tutaj, przed świadomością (Dalla Barba 2009, s. 257).

Jaka jest więc różnica między zwykłymi przypadkami pamiętania i antycypacji oraz przypadkami konfabulacji? W normalnych przypadkach TC oddziałuje z mniej stabilnymi wzorcami modyfikacji w mózgu, dlatego też rozpoznajemy prawdziwą przeszłość, teraźniejszość i przyszłość. W przypadku konfabulacji mamy do czynienia ze świadomością konfabulacyjną, czyli taką, w której przeszłość, teraźniejszość i przyszłość są zniekształcone w stosunku do tego, co się wydarzyło, co obecnie ma miejsce lub co się wydarzy.

Przyczyna zaistnienia świadomości konfabulacyjnej jest, zdaniem Dalla Barby, biologiczna. Lezje w mózgu pacjentów, którzy konfabulują, prowadzą do tego, że TC zostaje zmienione – przestaje oddziaływać z mniej stabilnymi wzorcami modyfikacji w mózgu, a zaczyna oddziaływać z bardziej stabilnymi. Istnieją jednak przypadki uszkodzeń mózgu (np. amnezja), w których TC całkowicie zanika.

Jakie struktury w mózgu odpowiadają więc za poprawne funkcjonowanie świadomości czasu oraz przypadki jej zmiany i zaniku? Wiele badań wskazuje na to, że lezje w środkowym płacie skroniowym (MTL – *Medial Temporal Lobe*) prowadzą do amnezji. Pacjenci z uszkodzeniem tego obszaru są niezdolni do przypominania sobie epizodów z życia, podczas gdy pamięć faktów jest u nich

dość dobrze zachowana (Dalla Barba 2009, s. 258). Jednym z bardziej znanych przykładów w historii psychiatrii jest przypadek pacjenta o inicjałach HM, badanego przez Brendę Milner (1958). Pacjent ten był niezdolny do przypominania sobie personalnych zdarzeń z przeszłości ani nie potrafił wyobrazić sobie własnej przeszłości. Był niejako zawieszony „w beczasowej momentalnej terażniejszości” (Dalla Barba 2009, s. 259), co świadczy o utracie świadomości czasu.

Co ciekawe, pacjenci z uszkodzeniem MTL często zachowują pamięć semantyczną, co oznacza, że świadomość czasu i świadomość wiedzy są odrębnymi rodzajami świadomości. Za KC odpowiadają prawdopodobnie inne struktury mózgu. Do normalnego funkcjonowania TC potrzebny jest przyśrodkowy płąt skroniowy, co potwierdzają nie tylko badania na pacjentach z lezjami tego obszaru, ale także badania obrazowania mózgu prowadzone na zdrowych ludziach. Obszar hipokampa aktywuje się zarówno wtedy, gdy osoby przypominają sobie jakieś zdarzenie z przeszłości, jak i wtedy, gdy wyobrażają sobie przyszłość (Dalla Barba 2009, s. 259).

Co dzieje się jednak z MTL w przypadku pacjentów konfabulujących? Do zaistnienia TC potrzebne jest funkcjonowanie MTL, a pacjenci konfabulujący posiadają sprawną świadomość czasu, choć jest ona nieco zmieniona, a zatem zarówno dla normalnej świadomości czasu, jak i dla konfabulacyjnej świadomości czasu niezbędne jest sprawne funkcjonowanie MTL.

[...] najbardziej prawdopodobna odpowiedź na to pytanie jest taka, że w przypadku konfabulacji MTL jest nienaruszone, ale lezje w innych obszarach mózgu nie pozwalają na jego normalne funkcjonowanie, co jest biologicznym odpowiednikiem twierdzenia, że w przypadku konfabulacji TC jest zachowane, ale nie oddziałuje już z mniej stabilnymi modyfikacjami mózgu (Dalla Barba 2009, s. 260).

Takie stanowisko potwierdzają wyniki współczesnych badań, które wykazują, że pacjenci konfabulujący nie mają lezji w przyśrodkowym płacie skroniowym (Dalla Barba 2009, s. 260).

### 2.3. Paradoks homunkulusa a epistemiczna teoria konfabulacji William Hirsteina

Epistemiczna koncepcja konfabulacji Williama Hirsteina zakłada istnienie dwóch rodzajów dysfunkcji systemu poznawczego: dysfunkcji systemu epistemicznego (np. systemu percepcji lub pamięci) oraz dysfunkcji systemu sprawdzającego. Hirstein uważa, że we wszystkich rodzajach zespołów neuropsychologicznych, w których występują konfabulacje, dochodzi do niezdolności wychwytywania tego, że dana reprezentacja jest nieodpowiednia do kontekstu. Twierdzi on, iż zaburzone procesy monitorujące odpowiedzialne są nie tylko za

konfabulacje pamięciowe, ale także za inne rodzaje konfabulacji. Funkcje sprawdzające i monitorujące dane pełnione są przez płaty czołowe. Dlatego uważa on, że zachodzą prawdopodobnie nieświadome procesy monitorujące, które w nieuszkodzonym systemie poznawczym odpowiedzialne są za monitorowanie, czy dana reprezentacja jest poprawna, czy nie, a następnie wyhamowywanie niepoprawnych reprezentacji i poddawanie świadomości reprezentacji poprawnej. W kontekście rozważań Dalla Barby nad paradoksem nieświadomych procesów monitorujących warto się zastanowić, czy postulowane przez Hirsteina procesy sprawdzające nie prowadzą do logicznego błędu homunkulusa.

Należy jednak podkreślić, że nie wszyscy kognytywiści uważają koncepcję nieświadomego homunkulusa za nieadekwatną. Zdaniem Francisa Cricka i Christofa Kocha nieświadomy homunkulus jest koncepcją atrakcyjną i użyteczną do wyjaśniania mechanizmów funkcjonowania mózgu i nie musi prowadzić do regresu do nieskończoności (Crick i Koch 2009, s. 307). Zdaniem Kocha „gdzieś w obrębie płata czołowego istnieją sieci neuronalne, które działają dokładnie jak homunkulus” (Koch 2008, s. 306). Nieświadomy homunkulus otrzymuje wejście zmysłowe z tylnych obszarów kory, a następnie na ich podstawie podejmuje decyzje i wysyła je do odpowiednich obszarów ruchowych. Ujmując rzecz metaforycznie, homunkulus „patrzy” do tyłu kory na układy sensoryczne (Koch 2008, s. 306; Crick i Koch 2009, s. 307).

Koch przytacza za psychologiem Fredem Attneavem dwa rodzaje krytyki koncepcji homunkulusa. Pierwszy dotyczy dualizmu tej koncepcji, zgodnie z którą kwestia homunkulusa jest poza obszarem badań naukowych, gdyż musiałby on leżeć poza sferą materii. Koch jednak ripostuje, że taki rodzaj homunkulusa nie stosuje się do jego i Cricka koncepcji, gdyż w ich ujęciu „homunkulus odpowiada za czynności realnie istniejącego układu fizycznego, zlokalizowanego w obrębie płata czołowego [...]” (Koch 2008, s. 306). Drugi rodzaj krytyki to regres do nieskończoności: Czy homunkulus nie potrzebuje kolejnego homunkulusa, który planowałby i kontrolował jego działania? Na ten zarzut Koch odpowiada następująco:

Jednak w naszym przypadku nie występuje taki nieskończony regres, ponieważ koncepcja homunkulusa nie ma tu na celu wyjaśnienia istoty *qualiów*. Nasz homunkulus działa raczej jak jednostka obliczeniowa (Koch 2008, ss. 306-307).

Zdaniem Kocha homunkulus odpowiedzialny jest za wiele złożonych operacji mentalnych, takich jak: intencje, tworzenie pojęć czy myśli, które nazywa operacjami supramentalnymi.

Przetwarzanie supramentalne znajduje się poza obszarem świadomej percepcji i przeciwstawia się domenie submentalnej, która obejmuje bardziej pierwotne etapy przetwarzania, również niedostępne dla świadomej percepcji (Koch 2008, s. 307).

Koncepcja nieświadomego homunkulusa ma także duże znaczenie w wyjaśnianiu takich zjawisk, jak: twórcze myślenie, wgląd czy rozwiązywanie problemów. Nie musi ona zatem prowadzić do regresu do nieskończoności, jakby chciał Dalla Barba. Wobec tego epistemiczna koncepcja konfabulacji Hirsteina, opierająca się na wykonawczych teoriach konfabulacji, może zostać utrzymana w mocy. Jest to o tyle istotne, że konfabulacje według Hirsteina to szersza klasa zjawisk, niekoniecznie związanych z procesami pamięci. Tymczasem Dalla Barba próbuje wyjaśnić wyłącznie zjawisko klinicznych konfabulacji pamięciowych. Zaletą koncepcji Hirsteina jest natomiast wyjaśnienie wszystkich rodzajów konfabulacji klinicznych poprzez odwołanie się do mechanizmów monitorujących. Jak piszą Crick i Koch:

Hipoteza o istnieniu homunkulusa jest dzisiaj bardzo niemodna, ale przecież nie da się zaprzeczyć, że jest to powszechny sposób, w jaki każdy z nas myśli o sobie. Byłoby to zdumiewające, gdyby to nieodparte złudzenie nie odzwierciedlało w pewien sposób najogólniejszych zasad organizacji mózgu (Crick i Koch 2009, s. 307).

## Konfabulacja jako egzaptacja

### 1. Ewolucja błędnych przekonań a problem wartości przystosowawczej wiarygodnej wiedzy<sup>1</sup>

U podstaw wielu koncepcji tłumaczących fenomen ewolucji człowieka tkwi milczące założenie, że posiadanie prawdziwych przekonań na temat świata jest cechą adaptacyjną w przeciwieństwie do posiadania przekonań błędnych. Informacje takie jak to, gdzie znajduje się pokarm, kto może być potencjalnym przyjacielem, kto partnerem, a kto przeciwnikiem, wydają się kluczowe dla przetrwania i reprodukcji organizmów.

Czy przekonania, które przybliżają właściwości rzeczywistości, maksymalizują więc wartość przystosowawczą organizmu, który te przekonania posiada? Odpowiedź twierdząca oznacza, że ewolucja stworzyła ludzki system poznawczy w taki sposób, aby mógł on formować prawdziwe przekonania o świecie, gdyż są one użytecznymi narzędziami pozwalającymi na efektywne działanie zarówno w środowisku fizycznym, jak i społecznym.

Jeśli jednak wiarygodna wiedza o świecie podnosi wartość przystosowawczą, to jak wytłumaczyć liczne przypadki konfabulowania, samooszukiwania czy iluzji poznawczych? Jak wyjaśnić powszechne w ludzkiej populacji występowanie odstępstw od wiarygodnych przekonań, zarówno tych dotyczących świata, samego siebie, jak i innych? Czy wszystkie przypadki posiadania błędnych przekonań świadczą więc o ograniczeniach ewolucyjnych, czy też raczej pokazują, że pierwotne założenia o przystosowawczej wartości wiarygodnej wiedzy są bezpodstawne?

Zwolennicy stanowiska dyskredytującego przystosowawczą wartość wiarygodnej wiedzy uważają, że „dobór naturalny nie dba o prawdę; dba on tylko o sukces reprodukcyjny” (Stich 1990, s. 62). Ponadto sądzą oni, że podtrzy-

---

<sup>1</sup> Główne idee tego rozdziału zostały przedstawione w: Reuter 2010a, 2010b.

mywanie poglądu, iż „dobór naturalny faworyzuje te systemy nerwowe, które tworzą bardziej wierne obrazy świata, musi być bardzo naiwnym poglądem na ewolucję mentalną” (Trivers 2006, s. xx). Niemniej nawet tak radykalni w tej kwestii ewolucjoniści, jak Robert Trivers czy Stephen Stich, uznają, że prawdziwe przekonania o świecie mają wartość instrumentalną dla doboru naturalnego o tyle, o ile umożliwiają sukces ewolucyjny. Wśród wielu badaczy zorientowanych ewolucyjnie panuje jednak pogląd, że to właśnie przekonania, które przybliżają rzeczywistość, maksymalizują wartość przystosowawczą organizmu, który je posiada (McKay i Dennett 2009; Millikan 1993). Wiarygodna wiedza jest bowiem narzędziem, które służy sprawnemu poruszaniu się w środowisku. Możliwość nabywania wiedzy związana jest z ewolucyjnie wypracowanymi mechanizmami formowania wiarygodnych przekonań przez system poznawczy człowieka. Zakłada się zatem, że ludzki system poznawczy został ukształtowany w drodze ewolucji biologicznej w taki sposób, aby zdobywać i utrzymywać wiarygodną wiedzę o świecie.

Skoro ewolucja ukształtowała nasz system poznawczy tak, by nabywał reliabilną wiedzę o świecie, to jak wyjaśnić odstępstwa od tej zasady, takie jak konfabulacje, samooszustwa, iluzje i błędy poznawcze? Zagadnienie to wydaje się kluczowe dla problematyki ewolucji umysłu.

Zdaniem Ryana McKaya i Daniela C. Dennetta odpowiedź na to pytanie zmienia się w zależności od tego, jakie przypadki błędnych przekonań rozpatrujemy (McKay i Dennett 2009). Jak wcześniej wspomniałam, badacze ci dzielą błędne przekonania na nietolerowalne i tolerowalne. Pierwsze są skutkiem dysfunkcji systemu poznawczego i wynikają z uszkodzeń mózgu – są to przypadki różnych zespołów neuropsychologicznych. Drugie są zaś niepożądanym, aczkolwiek tolerowalnym skutkiem poprawnie funkcjonujących procesów w zdrowych systemach poznawczych. Nietolerowalne błędne przekonania będące wynikiem dysfunkcji procesów poznawczych są nieadaptacyjne. Co jednak z przypadkami tolerowalnych błędnych przekonań? Czy zawsze są adaptacyjne? Zdaniem tych kognitywistów adaptacyjnymi błędnymi przekonaniami mogą być np. pozytywne iluzje poznawcze na temat samego siebie, własnej przeszłości i przyszłości oraz na temat bliskich osób. Dlaczego jednak tylko te iluzje? Co z innymi typami błędnych przekonań?

Proponuję, aby uznać konfabulacje, samooszustwa oraz iluzje poznawcze za efekty uboczne występowania innych adaptacyjnych cech umysłu ludzkiego, oraz by zamiast wyjaśnień w kategoriach adaptacyjnych stosować wyjaśnienia w kategoriach egzaptacji i pendentywów.

## 1.1. Krytyka radykalnego adaptacjonizmu Stephena J. Goulda

Przedmiotem analizy są rodzaje konfabulacji pełniące funkcje adaptacyjne, które stanowią produkt uboczny poprawnie działających procesów poznawczych. Ponieważ niektóre przypadki konfabulacji są skutkami ubocznymi poprawnie funkcjonujących procesów poznawczych, stosują do nich pojęcia „pendentyw” i „egzaptacja” wprowadzone do biologii ewolucyjnej przez Stephena J. Goulda, Richarda Lewontina i Elisabeth Vrbę (Gould i Lewontin 1973; Gould i Vrba 1982). Badacze ci wskazują na kłopoty radykalnego adaptacjonizmu w wyjaśnianiu zagadnień z zakresu ewolucji procesów poznawczych człowieka. Stosowanie przez nich pojęć „egzaptacja” i „pendentyw” miało na celu odejście od dość radykalnie rozumianego sensu adaptacyjności w wyjaśnianiu ludzkich dyspozycji przystosowawczych. Termin „egzaptacja” stosuje się na określenie cechy, która obecnie ma wartość przystosowawczą, ale nie została ukształtowana w drodze doboru naturalnego i spełnia inną funkcję niż ta cecha, która powstała jako adaptacja.

Czy pozytywne iluzje poznawcze oraz konfabulacje i samooszustwa to na pewno cechy adaptacyjne ludzkiego umysłu, które powstały w drodze doboru naturalnego? Czy inne rodzaje iluzji poznawczych, takie jak iluzje percepcyjne czy iluzje pamięci, są cechami adaptacyjnymi? A jeśli nie, to jakie są inne możliwe wyjaśnienia?

W 1979 r. Gould i Lewontin w artykule pod znamienym tytułem *The Spandrels of San Marco and Panglossian Paradigm: A Critique of the Adaptationist Programme* krytycznie odnieśli się do radykalnych koncepcji adaptacjonistycznych, wedle których jeśli istnieje jakaś cecha organizmu, to musi mieć swoje wyjaśnienie adaptacyjne. Badacze ci uznali, że wyjaśnienia w kategoriach adaptacyjnych mogą być błędne, ponieważ ewolucja nie jest procesem teleologicznym, a zatem ma swoje ograniczenia. Tymczasem teorie adaptacjonistyczne *implicite* lub *explicite* tworzone są według paradygmatu Panglossa<sup>2</sup>. Podobnie – zdaniem Goulda i Lewontina – radykalni adaptacjoniści opowiadają „takie sobie bajeczki” ewolucyjne: najpierw rozbijają organizm na cechy, które są wyjaśniane jako struktury optymalnie ukształtowane w drodze doboru naturalnego do pełnienia swoich funkcji, a następnie interpretują go jako najlepszy kompromis pośród konkurencyjnych wymagań. Dlatego ostrze krytyki Goulda i Lewontina skierowane jest na konstatację, że suboptymalność danej części organizmu jest wyjaśniana przez adaptacjonistów jako najlepsze z możliwych rozwiązań dla całego organizmu.

<sup>2</sup> Pangloss to bohater książki Woltera pt. *Kandyd*, który uosabia poglądy niemieckiego filozofa Leibniza, a zwłaszcza jego słynną maksymę: „Świat jest najlepszym z możliwych”.

Według badaczy tych jest to błędny opis procesów ewolucyjnych, ponieważ prowadzi do paradoksów. Przykładem mogą być takie absurdalne twierdzenia adaptacjonistyczne, jak to, że płatek ucha powstał po to, by umożliwić wieszanie na nim kolczyków. Tymczasem nie każda cecha organizmu została zachowana w ramach doboru naturalnego jako adaptacyjna, a więc przynosząca korzyść. Niektóre cechy mogą być skutkiem ubocznym cech adaptacyjnych, które dziś spełniają jakieś użyteczne funkcje (Lewontin i Gould 1979).

Aby to dokładniej wyjaśnić, badacze odwołali się do pojęcia pendencywu zaczerpniętego z architektury. Pendencywy w architekturze renesansowej to trójkątne obszary zakrzywienia sufitu ponad łukami, wykorzystywane często do umieszczania w nich fresków lub płaskorzeźb. Obszar ten uważa się raczej za uboczną konsekwencję kształtu łuku niż za formę umyślnie zaprojektowaną do umieszczania w niej fresków. Pendencywy powstały zatem w wyniku ograniczeń architektonicznych i dopiero później zaczęły pełnić nową funkcję – ozdobną. Podobnie było z powstawaniem struktur i mechanizmów działania organizmu w toku ewolucji – niektóre cechy organizmów mogły rozwinąć się z tego samego powodu co pendencywy w architekturze. Pojęcie pendencywu wprowadzone do biologii ewolucyjnej oznacza zatem nieadaptacyjną cechę organizmu, powstałą jako efekt uboczny cechy adaptacyjnej. Istotą pendencywu jest to, że może on zacząć wtórnie pełnić jakąś użyteczną funkcję (Gould i Lewontin 1979).

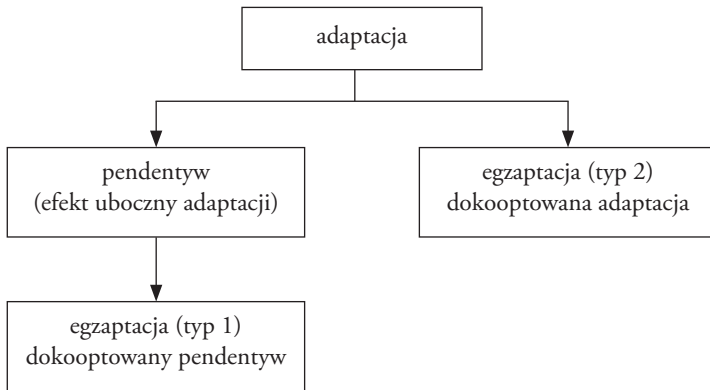
Kolejnym przełomem w tej problematyce był artykuł Goulda i Vrby pt. *Exaptation: A missing term in the science of form* (1982). Autorzy piszą w nim, że w biologii ewolucyjnej brakuje terminów na określenie cech organizmu, które zapewniają obecnie lepsze przystosowanie, mimo iż nie zostały ukształtowane w drodze doboru naturalnego. Terminem, który proponują, jest egzaptacja. Oznacza ona użyteczne struktury, które „wyewoluowały dla innych zastosowań (lub nie miały na początku żadnej funkcji), lecz później zostały przejęte dla obecnie pełnionej roli” (Gould i Vrba 1978). Są one przystosowane dla aktualnie pełnionej roli (*aptus*), ale nie zostały ukształtowane do (*ad*) obecnego użycia, lecz stanowią konsekwencję (*ex*) cech powstałych z innych powodów. Dlatego są one raczej „egzaptacyjne” (*ex aptus*) a nie adaptacyjne (*ad aptus*).

Ponieważ pojęcia „pendentyw” i „egzaptacja” dostarczają krytykom okazji do dyskusji i polemik, zarówno jeśli chodzi o zasadność wprowadzenia samych pojęć wraz z leżącym u podłoża tego zabiegu podejściem ewolucyjnym, jak również niekonsekwencje definicyjne i konceptualne w artykułach twórców tych pojęć, postaram się rozjaśnić najważniejsze kwestie związane z tymi terminami.

Po pierwsze, należy ustalić, czym jest pendencyw w przeciwieństwie do egzaptacji i adaptacji. Adaptacja to cecha dziedziczna i podlegająca rozwojowi, która została uformowana w drodze doboru naturalnego. Organizmy posiadające tę cechę wydają liczniejsze potomstwo od tych, które jej nie mają. Proponuję,

aby uznać pendentyw za efekt uboczny jakiejś cechy adaptacyjnej, który nie pełni żadnej funkcji przystosowawczej, lecz może zacząć ją pełnić. Natomiast egzaptacja to taka właściwość lub funkcja organizmu, która pojawiła się w toku ewolucji jako cecha korzystna (w danym czasie i warunkach bytowania), np. cecha A, która przez powiązanie (korelacyjne lub innego rodzaju) z inną cechą adaptacyjną, np. B, upowszechniła się w drodze doboru naturalnego, po czym okazała się korzystniejszą cechą adaptacyjną niż cecha B.

Jednak, jak zauważają David M. Buss i in. (1998), w artykułach Goulda pojawiają się dwie definicje egzaptacji. Po pierwsze, egzaptacja to dodatkowa funkcja nadbudowana nad istniejącą już funkcją adaptacyjną. Ten rodzaj egzaptacji można nazwać za Bussem i in. (1998) dokooptowaną adaptacją (*co-opted adaptation*), a jej przykładem są pióra ptaków. Pióra pierwotnie występowały u gadów i pełniły rolę termoregulacji organizmu, dopiero u ptaków zaczęły służyć lataniu. Termoregulacja to zatem cecha adaptacyjna, a zdolność latania to egzaptacja. Po drugie, egzaptacja to funkcja nadbudowana nad nieadaptacyjnym pendentywem, czyli efektem ubocznym cechy adaptacyjnej. Ten typ egzaptacji można nazwać dokooptowanym pendentywem (*co-opted spandrel*). Model powstawania egzaptacji i pendentywów przedstawia schemat 8.



Schemat 8. Dwa typy egzaptacji według Stephena J. Goulda

Źródło: opracowanie własne na podstawie: Buss i in. 1998.

Pojęcia „egzaptacja” i „pendentyw” przeniesiono z czasem na grunt szeroko rozumianej ewolucji ludzkiego umysłu. Sam Gould w artykule *Exaptation: A crucial tool for an evolutionary psychology* (1991) napisał:

[...] ludzki mózg osiągnął większy rozmiar i zdolności w drodze doboru naturalnego dla osiągnięcia określonych celów u naszych przodków [...]. Dobór naturalny

uksztaltował mózg; jednak z racji jego strukturalnej złożoności mózg może wykonywać wiele zadań, które potem mogą stać się kluczowe dla kultury, ale które są raczej pendentami niż obiektami pierwotnego doboru naturalnego. [...] Z pewnością dla czegoś tak złożonego i pełnego ukrytych zdolności jak ludzki mózg pendentów musi być znacznie przewyższać liczbą pierwotne przyczyny, a egzaptacje ludzkiego mózgu muszą przewyższać adaptacje z uwagi na ich ilość (Gould 1991, s. 57).

Co ciekawe, pendentem i egzaptacją określane są w badaniach nad ewolucją umysłu takie zdolności poznawcze, jak język (Gould 1991; Hauser, Chomsky i Fitch 2002), muzyka (Pinker 1997), estetyka (De Sousa 2004), religia (Gould 1991), a także procesy poznawcze wraz z ich efektami, takimi jak samooszustwo (Van Leeuwen 2007), świadomość (Gould 1991), pamięć i jej cechy (Sherry i Schacter 1987; Schacter 2003). Z zagadnień tych szczególnie ciekawe wydają się zjawiska pokrewne iluzji poznawczej, czyli samooszustwa i zniekształcenia pamięci. Z zagadnień tych szczególnie interesujące wydają się fenomeny umysłu, takie jak iluzje poznawcze, samooszustwa oraz zniekształcenia pamięci.

## 1.2. Samooszustwo życzeniowe jako pendent

Jednym z pierwszych badaczy, który zwrócił uwagę na to, że samooszustwo jest nie tyle adaptacją, ile skutkiem ubocznym innych cech adaptacyjnych ludzkiego umysłu, był D. S. Neil Van Leeuwen (2007a, 2007b, 2008). Podzielił on samooszustwa na trzy fenomeny: samooszustwo życzeniowe, samooszustwo umyślne i samooszustwo negatywne, uznając, że pierwsze jest najbardziej powszechne w ludzkiej populacji i na nie w szczególności należy zwrócić uwagę podczas rozważań ewolucyjnych na temat przystosowawczego charakteru wiarygodnej wiedzy (Van Leeuwen 2008).

W pierwszej kolejności Van Leeuwen podaje w wątpliwość tezę Roberta Triversa, że samooszustwo jest adaptacyjne i ma pozytywną wartość przystosowawczą, gdyż wspiera oszukiwanie innych (Trivers 2000). Teza ta jest kontrowersyjna z uwagi na to, że można podać wiele przykładów kłamstwa bez samooszustwa i na odwrót. Trudno też sobie wyobrazić, jaką korzyść miałby czerpać podmiot oszukujący innych, oszukując jednocześnie siebie w tej samej sprawie. Głównym założeniem kłamstwa jest to, że korzyść czerpana jest przez kłamcę, który jako jedyny wie, że przekonania, jakie głosi, są niezgodne z prawdą. Gdyby kłamca oszukiwał sam siebie, nie mógłby czerpać korzyści ze swoich zmysłów (Van Leeuwen 2007a, 2007b, 2008). W przypadku zwierzęcia, które chowa przed innymi pożywienie, można uznać to za rodzaj taktycznej decepcji – pożywienie chowane jest po to, aby inne osobniki go nie zauważyły. Korzyść z tego „oszustwa” czerpie sam oszust, ale tylko w sytuacji, gdy sam siebie nie oszukuje.

Van Leeuwen postuluje, żeby zrezygnować z tezy o wartości przystosowawczej samooszustwa i uznać, że fenomen ten jest w istocie pendentywem, czyli skutkiem ubocznym innych cech adaptacyjnych ludzkiego umysłu. Cechy, które prowadzą do samooszustwa, to:

- charakterystyczny dla ludzkich pragnień niepokój towarzyszący antycypacji lub dowody na niezaspokojenie pragnień,
- selektywna uwaga skierowana na dane wejściowe i dowody,
- ogólna tendencja do unikania dyskomfortu,
- strukturalna organizacja danych w systemie poznawczym,
- tworzenie przekonań na bazie dowodów wraz z normami epistemicznymi,
- doświadczanie przyjemności w sytuacji posiadania dowodów, że pragnienia zostaną zaspokojone,
- poszukiwanie przyjemności (Van Leeuwen 2008, ss. 198-200).

Van Leeuwen wyróżnia także czynniki pośrednie wspomagające oszukiwanie samego siebie, które dzieli na: ułatwienia epistemiczne (*epistemic facilitators*) czyli czynniki, które w odpowiednim kontekście ułatwiają podtrzymywanie przekonań sprzecznych z normami epistemicznymi, oraz źródła pragnień (*desire sources*), czyli aspekty umysłu, które tworzą różnego rodzaju pragnienia prowadzące do samooszustwa. Ułatwienia epistemiczne to: inercja sieci przekonań, generowanie myśli i przypuszczeń przez emocje oraz pragnienia, przypisywanie odpowiedniego stopnia sceptycyzmu różnym propozycjom, zdolność tłumienia niechcianych wspomnień. Źródła pragnień to z kolei: tworzenie skomplikowanych intencji oraz doświadczanie dysonansu poznawczego (Van Leeuwen 2008, ss. 200-204).

Samooszustwo jest więc pendentywem powstałym jako skutek uboczny innych, adaptacyjnych właściwości mentalnych ludzkiego umysłu, takich jak wymienione powyżej:

[...] myślę, że jest bardziej owocne, aby widzieć zdolność samooszukiwania się jako pendentyw, nie jako adaptację. To znaczy ludzka skłonność do samooszukiwania istnieje nie dlatego, że cecha ta została w tym celu wyselekcjonowana w naszej ewolucyjnej historii, ale dlatego, że inne aspekty naszej poznawczej architektury prowadzą do samooszukiwania jako efektu ubocznego (Van Leeuwen 2007a, s. 339).

### 1.3. Iluzje poznawcze i konfabulacje – adaptacje, pendentywy czy egzaptacje?

Van Leeuwen, stosując pojęcie pendentywu do analizy fenomenu samooszustwa, nie wziął pod uwagę tego, że krytyka radykalnego adaptacjonizmu została rozszerzona w dalszych artykułach Goulda o pojęcie egzaptacji. Moim zdaniem

pojęcie to jest bardziej istotne niż pojęcie pendentywu w rozważaniach nad ludzkimi fenomenami mentalnymi. Tytuł jednego z ważniejszych artykułów Goulda wskazuje na to, że to właśnie egzaptacja ma stać się centralnym pojęciem psychologii ewolucyjnej (Gould 1991).

Przy stosowaniu tego pojęcia na gruncie badań ewolucji procesów poznawczych człowieka należy pamiętać, że istnieją dwa typy egzaptacji, i to raczej egzaptacja w pierwszym znaczeniu (jako dokooptowany pendentyw) jest istotniejsza dla tego rodzaju badań.

W moim przekonaniu samooszustwo życzeniowe w ujęciu Van Leeuwena jest fenomenem podobnym do iluzji pozytywnych, o których piszą Dennett i McKay. Różnica tkwi w stopniu pewności przypisywanej przekonaniom oraz w wiedzy, jaką posiada podmiot oszukujący sam siebie oraz podmiot ulegający iluzji. Samooszustwo związane jest z epistemicznym napięciem między dowodami nie do odparcia a życzeniowym przekonaniem podmiotu. W związku z tym stopień pewności przypisywany przekonaniu może się wahać. Natomiast w przypadku iluzji podmiot nie musi mieć dowodów przeczących jego przekonaniu, a subiektywna pewność przypisywana przekonaniom często jest wysoka.

Samooszustwo oraz iluzje pozytywne mogą także różnić się istotnością kontekstu powstawania błędnie ugruntowanego przekonania. Podczas gdy w badaniach nad iluzjami pozytywnymi, takimi jak efekt ponadprzeciętności, bierze się pod uwagę jedynie ogólną tendencję w populacji do nabywania przekonań tego rodzaju, to w badaniach nad samooszustwem kontekst powstawania przekonania błędnie ugruntowanego ma duże znaczenie. Na przykład osoba, która jest ofiarą przemocy w rodzinie, może oszukiwać sama siebie, że przemoc stosowana wobec niej zakończy się, na podstawie selektywnej uwagi kierowanej na dowody, jakimi są deklaracje oprawcy o zaprzestaniu znęcania się nad ofiarą. Dodatkowym czynnikiem może być wyparcie traumatycznych wspomnień. Kontekst zaistniałego wtedy zdarzenia ma tu zatem duże znaczenie. W przypadku iluzji, jaką jest efekt ponadprzeciętności dotyczący bliskich osób, kontekst zdarzenia nie ma już takiego znaczenia – badani niezależnie od kontekstu twierdzą, że ich partner jest bardziej inteligentny, atrakcyjny, interesujący niż inni.

Pomimo różnic iluzje pozytywne można rozpatrywać jako przypadek samooszustwa życzeniowego bądź też na odwrót – samooszustwo życzeniowe może być rozpatrywane jako rodzaj pozytywnego samozłudzenia. W moich rozważaniach istotna jest jednak ewolucyjna charakterystyka tych fenomenów epistemicznych. McKay i Dennett uważają, że iluzje pozytywne to adaptacje. Jednak w związku z krytyką radykalnego adaptacjonizmu w psychologii ewolucyjnej warto się zastanowić, czy iluzje pozytywne nie są egzaptacjami powstałymi na podstawie pendentywu (pierwszy typ egzaptacji). Na przykład

efekt ponadprzeciętności może być rozpatrywany jako przypadek życzeniowego samooszustwa, który jest skutkiem ubocznym selektywnej uwagi skierowanej na dowody oraz poszukiwania przyjemności i unikania dyskomfortu.

Według Van Leeuwena samooszustwo, choć samo nie posiada wartości przystosowawczej, stanowi skutek uboczny innych adaptacyjnych cech umysłu ludzkiego. Czy jednak samooszustwo, mimo że nie powstało w drodze doboru naturalnego i jest tylko efektem ubocznym adaptacji, nie pełni jakiejś funkcji przystosowawczej? Na przykład pacjent cierpiący na ciężką chorobę może oszukiwać sam siebie, że stan jego zdrowia jest dobry bądź że wkrótce mu się polepszy. Idąc tropem Van Leeuwena, można powiedzieć, że tego rodzaju przekonanie jest skutkiem ubocznym innych adaptacyjnych cech ludzkiego umysłu, takich jak: poszukiwanie przyjemności, selektywna uwaga na dowody, tendencja do unikania dyskomfortu. Jednakże samo przekonanie o dobrym stanie zdrowia, choć jest efektem ubocznym innych adaptacyjnych cech umysłu, może zacząć pełnić nową funkcję i w tym sensie może okazać się egzaptacją jako dokooptowanym pendentywem. Ta dodatkowa funkcja – w podanym przykładzie – to lepsze subiektywne samopoczucie pacjenta oraz lepsze rokowania na wyzdrowienie.

W kilku badaniach wykazano też, że nierealistycznie pozytywna ocena stanu własnego zdrowia może prowadzić do jego poprawy bądź do wolniejszego postępowania choroby oraz dłuższego życia pacjenta, gdy choroba jest ciężka (Reed i in. 1994; Reed i in. 1999; Taylor i in. 2003). McKay i Dennett uważają, że dane te świadczą o adaptacyjności tego fenomenu, gdyż bezpośrednio wpływają na przeżycie lub pośrednio – na adaptacyjne zachowania. Jednak pozytywne złudzenia na temat własnego stanu zdrowia można by uznać za egzaptację, która podnosi wartość przystosowawczą. Powstała ona jednak jako dokooptowany pendentyw, czyli jako nowa funkcja dla efektów ubocznych innych cech adaptacyjnych ludzkiego umysłu, takich jak wymienione przez Van Leeuwena.

Podobnie efekt ponadprzeciętności dotyczący bliskich osób może pełnić funkcję przystosowawczą, taką jak angażowanie się w działania adaptacyjne, czyli w związku z partnerami. Nie musimy tu przyjmować za McKayem i Dennettem stanowiska radykalnie adaptacjonistycznego, twierdząc, że jest to cecha adaptacyjna, która została ukształtowana w drodze doboru naturalnego. Bardziej prawdopodobne wydaje się to, że jest to efekt uboczny innych cech adaptacyjnych ludzkiego umysłu (analogicznie do argumentacji Van Leeuwena dotyczącej samooszustwa życzeniowego) i dopiero wtórnie podtrzymywanie tego rodzaju przekonań zaczęło pełnić dodatkowe funkcje, stając się egzaptacją powstałą na bazie pendentywu.

Koncepcję egzaptacji i pendentywów można też zastosować do innych rodzajów iluzji poznawczych. Przykładem mogą być konfabulacje pamięciowe,

takie jak błędna atrybucja źródła, która polega na tym, że przekonanie pamięciowe jest sklejeniem dwóch różnych śladów pamięciowych. Na przykład gdy świadek zdarzenia przypomina sobie detale sytuacji poprawnie, ale przyporządkowuje twarz innej osoby, którą widział wcześniej, do twarzy przestępcy i ma błędne przekonanie, że niewinna osoba popełniła przestępstwo. Inny przykład to przypominanie sobie, że byliśmy świadkami zdarzeń, które tylko sobie wyobrażyliśmy bądź znamy je z innego źródła (z mediów, od innych osób).

Daniel Schacter, który jest badaczem pamięci, uważa, że tego typu błędy pamięci są efektami ubocznymi (pendentywami) adaptacyjnych procesów pamięciowych, które mają jednak swoje ograniczenia. Te adaptacyjne cechy pamięci to zdolność kodowania schematów poznawczych bądź ogólna pamięć zdarzenia kosztem pamięci szczegółów (Schacter 2003).

Inny rodzaj złudzenia pamięci, zwany tendencyjnością pamięci, np. złudzenie stałości poglądów, Schacter uważa za egzaptację.

Tendencyjność pamięci jest przypadkowym produktem ubocznym takiej właściwości umysłu, która pozwala ogólnej wiedzy i przekonaniom sterować aktami zapamiętywania i odtwarzania [...]. Specyficzne formy tendencyjności pamięci (związanej z operacjami poznawczymi wyższego rzędu) będą się różnić między kulturami, są więc one raczej efektem norm społecznych i kulturowych niż adaptacją biologiczną powstałą w drodze doboru naturalnego (Schacter 2003, s. 203).

Reasumując, proponuję, by po objaśnieniu pojęć „pendentyw” i „egzaptacja” uznać propozycję Goulda za interesującą alternatywę dla radykalnie adaptacjonistycznych koncepcji upowszechnianych w psychologii ewolucyjnej. Wydaje się, że rozpatrywanie cech ludzkiego umysłu wyłącznie jako adaptacji powstałych w drodze doboru naturalnego jest problematyczne z uwagi na to, że istnienie efektów doboru naturalnego można zarejestrować dopiero w bardzo długich odcinkach czasu, podczas gdy ludzkie zdolności z perspektywy historii ewolucyjnej powstały dość niedawno. Dlatego propozycja ta może pomóc w wyjaśnianiu takich fenomenów, jak: iluzje, konfabulacje czy samooszustwa, które według radykalnego adaptacjonizmu wydają się paradoksami ewolucyjnymi.

Na pytanie, czy wiarygodna wiedza podnosi wartość przystosowawczą organizmu, można odpowiedzieć twierdząco, uznając, że przypadki iluzji poznawczych, konfabulacji i samooszustw są efektami ubocznymi, czyli pendentywami, które w toku ewolucji zaczęły pełnić nowe funkcje, stając się egzaptacjami. W moim przekonaniu wyjaśnienie to jest o wiele bardziej klarowne i adekwatne do niniejszych rozważań niż stanowiska utrzymujące, że wiedza nie podnosi wartości przystosowawczej lub że błędne przekonania mają charakter adaptacyjny.

W proponowanej interpretacji koncepcji Goulda występują dwa typy egzaptacji, z której każda ma inne źródło: albo nadbudowuje się nad pierwotną cechą adaptacyjną, albo też nadbudowuje się nad skutkiem ubocznym cechy adaptacyjnej. Konfabulowanie jest, moim zdaniem, jedną z form egzaptacji, tj. taką, która jest dokooptowana do pendentywu, czyli skutku ubocznego jakiejś cechy adaptacyjnej. Należy zwrócić uwagę na adaptacyjne cechy systemu poznawczego człowieka, których produkt uboczny (pendentyw) stanowi konfabulacja, a także wskazać na potencjalne funkcje, jakie pełni konfabulacja jako forma egzaptacji.

Uważam, że konfabulacja jest skutkiem ubocznym cechy umysłu, którą jest wypełnianie luk w systemie poznawczym. Obecnie konfabulacja może pełnić funkcje przystosowawcze i tym sposobem okazuje się egzaptacją. Weźmy przykład systemu pamięci. Uzupełnianie luk w systemie pamięciowym to skutek uboczny adaptacyjnej cechy umysłu, jaką jest generowanie schematów poznawczych. Ten skutek uboczny, czyli pendentyw, może jednak zacząć pełnić funkcje przystosowawcze, takie jak uspójnianie modelu świata lub modelu własnego „ja”.

Funkcję uspójniania obrazu poznawczego może pełnić zasadniczo konfabulacja i w zależności od tego, co jest jej przedmiotem, może uspójniać obraz świata zewnętrznego (konfabulacje naturalistyczne), obraz samego siebie (konfabulacje egocentryczne) bądź obraz umysłów innych osób (konfabulacje socjocentryczne).



## Zakończenie

Zjawisko konfabulowania, czyli nabywania i wygłaszania błędnie ugruntowanych przekonań, stanowi jedną z najbardziej intrygujących zagadek ludzkiego umysłu. Nie będąc ani intencjonalnym kłamstwem, ani świadomym tworzeniem fikcji, konfabulacja stanowi odrębny fenomen epistemiczny, który od niedawna jest przedmiotem rozważań nie tylko psychologów i neuropsychologów, ale także epistemologów i kognitywistów.

W zależności od definicji konfabulacja może być traktowana jako patologia neuronalna prowadząca do dysfunkcji systemu formowania przekonań, występująca rzadko w różnych zespołach neuropsychologicznych, bądź też dużo szerzej – jako część codziennej aktywności poznawczej normalnie działających w świecie podmiotów. Opierając się na szerszej definicji konfabulacji, jako błędnie ugruntowanego przekonania występującego w poprawnie działających systemach poznawczych, pokazałam, że zjawisko to występuje na tyle często w ludzkiej populacji, że zrozumienie, czym jest, jak powstaje oraz jaką pełni funkcję, może mieć ogromne znaczenie dla dyscyplin badawczych zajmujących się zagadnieniem ludzkiego poznania, takich jak epistemologia, filozofia umysłu, psychologia i neuropsychologia czy neuronauka poznawcza. Ponadto wiedza na temat mechanizmów powstawania oraz rodzajów konfabulacji może odegrać istotną rolę w naukach społecznych, takich jak: nauki polityczne, prawo, ekonomia czy pedagogika.

W niniejszej książce zaproponowałam poszerzenie aktualnych rozważań nad tym zagadnieniem o perspektywę epistemologiczną i kognitywistyczną oraz wzbogacenie perspektywy badawczej poprzez wzięcie pod uwagę przypadków konfabulacji w poprawnie funkcjonujących systemach poznawczych. W świetle dotychczasowych analiz uwidacznia się bowiem brak refleksji nad tym, dlaczego oraz jak funkcjonuje umysł normalny wtedy, gdy konfabuluje. Zagadnienie to wydaje się nader istotne dla zrozumienia, na czym polegają

odstępstwa patologiczne bądź zachowania znajdujące się na granicy normy fizjologiczno-funkcjonalnej.

Przedstawiona przeze mnie argumentacja przemawia za epistemiczną teorią konfabulacji w ujęciu Williama Hirsteina. Teoria ta mówi, że konfabulacja to błędnie ugruntowane przekonanie bądź błędnie ugruntowana reprezentacja ikoniczna, która jest skutkiem dwóch błędów w systemie poznawczym: błędu epistemicznego oraz błędu sprawdzającego. Hirstein rozważa jednak wyłącznie przypadki konfabulacji patologicznych. Książka ta zawiera zaś odautorskie uzupełnienie pracy tego badacza o wątki, które sprawiają, że zagadnienie konfabulacji staje się nie tylko przedmiotem refleksji badawczej psychologii, psychiatrii czy neurobiologii, ale także przedmiotem badań epistemologicznych, a co za tym idzie – wstępnych badań eksperymentalnych z zakresu kognitywistyki.

Zaproponowane przeze mnie podziały konfabulacji miały na celu pokazanie, że fenomen ten występuje w różnych dziedzinach naszego poznania, towarzysząc nam na co dzień w percepcji codziennych zdarzeń, odtwarzaniu wspomnień czy planowaniu działań. Moja propozycja nowego podejścia do wyjaśniania mechanizmów powstawania konfabulacji ukazuje, że punktem wyjścia dla tego zjawiska mogą być różne rodzaje stanów poznawczych, takie jak: luka w systemie poznawczym, brak dostatecznych informacji bądź w miarę kompletne dane. Stany te mogą prowadzić do trzech różnych rodzajów konfabulacji: wypełniającej, scalającej i addytywnej.

Z uwagi na to, że konfabulacja jest niekiedy utożsamiana z kłamstwem, a kiedy indziej z fikcjotwórstwem, podjęłam w książce próbę analizy porównawczej, zestawiając status i formę działania konfabulującego umysłu normalnego ze znanymi w literaturze przedmiotu ujęciami zagadnień: kłamstwa, fikcjotwórstwa i samooszustwa. W wyniku tego badania okazało się, że konfabulacja jest całkiem odmiennym zjawiskiem od wyżej wymienionych. Podaję także argumentację epistemologiczną za owymi odróżnieniami, pokazując, że konfabulacja to zjawisko, w którym nie zachodzi intencjonalne zwodzenie odbiorcy komunikatu, ponieważ osoba konfabulująca nie ma dystansu do tego, o czym mówi, oraz nie zakłada takiego dystansu u odbiorcy, a poza tym nie doświadcza osobliwego napięcia między sprzecznymi przekonaniem.

Zagadnienie konfabulacji można lepiej rozumieć, odwołując się do pojęć, jakie od pewnego czasu funkcjonują w psychologii poznawczej. Niezwykle istotne było więc odniesienie się do badań empirycznych z zakresu konfabulacji pamięciowych w poprawnie funkcjonujących systemach poznawczych. Wybór konfabulacji pamięciowych został podyktowany tym, że są one najbardziej powszechne w ludzkich społecznościach oraz z uwagi na dość obszerne już opracowanie tego zagadnienia w literaturze przedmiotu. Na podstawie analizy najnowszych badań pokazałam, że fałszywe wspomnienia, ale także

sugestywność czy uporczywość wspomnień są dość powszechnymi rodzajami konfabulacji pamięciowych w ludzkiej populacji.

Konfabulacje patologiczne, w przeciwieństwie od normalnych, charakteryzują się wysokim stopniem pewności i słabą podatnością na korektę. Ta niezdolność weryfikowania własnych błędnych przekonań w obliczu jawnych dowodów przeczących podawanych przez osoby z otoczenia jest przedmiotem zainteresowania badaczy dążących do wprowadzenia programów terapii pacjentów. Przedstawione przeze mnie wyjaśnienie tego zjawiska, odwołujące się do teorii Mind-Reading, miało na celu pokazanie, że pacjenci konfabulujący mają często trudności z „czytaniem umysłów” innych osób, przez co nie potrafią także korygować własnych błędnych przekonań, nawet gdy ich przekonania poddawane są racjonalnej krytyce i korekcie innych osób. Niezdolność odczytywania przekonań i intencji innych może być jednym z dodatkowych źródeł konfabulacji w zespołach neuropsychologicznych. Teoria Mind-Reading jest ponadto stosowana do wyjaśniania konfabulacji socjocentrycznych w zespołach błędnej identyfikacji (takich jak zespół Capgrasa czy zespół Fregoliego), gdzie odczytywanie stanów mentalnych innych ma szczególne znaczenie.

Rozstrzygnięcia dotyczące statusu konfabulacji, jej mechanizmów oraz funkcji mają wpływ na rozważania podejmowane w obrębie filozofii umysłu dotyczące podmiotowości oraz świadomości. Zaprezentowana w książce argumentacja, że konfabulacje pojawiają się wówczas, gdy zachodzi możliwość operacji na reprezentacjach mentalnych, ma konsekwencje dla niektórych koncepcji podmiotowości, takich jak wybrana przeze mnie jako szczególnie interesująca propozycja Ulrica Neissera. Wynika z niej, że „ja” ekologiczne i „ja” interpersonalne, opierające się na percepcji bezpośredniej, nie podlegają konfabulacjom, natomiast „ja” konceptualne, rozszerzone i prywatne, opierające się na reprezentacjach mentalnych, są tymi, które mogą, choć nie muszą, podlegać konfabulowaniu. Dlatego też przedstawiłam własne zestawienie rodzajów konfabulacji w odniesieniu do różnych rodzajów podmiotowości, które mają przybliżyć i rozwinąć to zagadnienie.

Konfabulacja jest specyficznym stanem umysłu, w którym podmiot tworzy i podtrzymuje błędne przekonania, bez świadomości, że są one niepoprawne. Niezdolność monitorowania poprawności reprezentacji bądź własnej wypowiedzi, będąca jednym z czynników konfabulacji w teorii epistemicznej, skłoniła mnie do analizy tego fenomenu w nawiązaniu do wybranych wątków z problematyki zagadnienia świadomości. Koncepcja heterofenomenologii wybrana została z uwagi na to, że jej twórca, Daniel C. Dennett, zauważył, iż ludzie w czasie eksperymentów laboratoryjnych często konfabulują, w związku z tym zadaniem heterofenomenologii stało się m.in. wyjaśnienie, dlaczego konfabulacje powstają. Zaproponowany przez tego filozofa model wielokrot-

nych szkiców, będący próbą wyjaśnienia zjawiska świadomości, ma za zadanie odeprzeć stanowiska opierające się na postulatcie teatru kartezyjskiego (także w wersji materialistycznej). Dennett krytykuje również pojęcie homunkulusa w badaniach świadomości, które – jak się okazuje – może być problemem także dla wykonawczych teorii konfabulacji postulujących istnienie nieświadomych procesów monitorujących. W książce pokazałam, jak uniknąć tego błędu i omówiłam konkurencyjną dla teorii wykonawczych koncepcję powstawania konfabulacji w ujęciu Gianfranca Dalla Barby, opierającą się na pojęciach świadomości czasu i świadomości wiedzy. Nie zgadzam się jednak z poglądem, że założenie o istnieniu mechanizmów monitorujących w nieunikniony sposób prowadzi do paradoksu homunkulusa. Przyjęcie tezy o istnieniu w mózgu nieświadomego homunkulusa może być przydatne w refleksji nad sposobem funkcjonowania mózgu, co postulują tacy badacze świadomości, jak Francis Crick i Christof Koch.

Istotne dla zrozumienia zjawiska konfabulowania jest nie tylko pytanie o jego definicję, rodzaje i mechanizmy, ale także o ewentualną funkcję w perspektywie ewolucyjnej. W niniejszej książce argumentuję, że konfabulowanie można lepiej zrozumieć, gdy interpretuje się je w kategoriach egzaptacji czy pendentywów. Przyjęta przeze mnie perspektywa biologiczno-ewolucyjna wynika z przekonania, że wszystkie systemy ożywione w procesie konfrontacji swoich kompetencji behawioralnych z warunkami otoczenia uruchamiają dyspozycje poznawcze i emocjonalne. Inaczej mówiąc, dyspozycje te poddane są w pierwszej kolejności testowi adaptacyjnemu. Oznacza to zarazem, że także w pierwszej kolejności poddane są one mechanizmowi ewolucji biologicznej, a jeśli chodzi o człowieka – także biologiczno-kulturowej. Z tego powodu przypisuję więc duże znaczenie przedstawionej próbie wyjaśnienia fenomenu konfabulacji przy wykorzystaniu pojęcia egzaptacji i pendentywu, zastosowanych pierwotnie do wyjaśniania cech organizmów w biologii ewolucyjnej, a następnie wykorzystanych do wyjaśniania ludzkich dyspozycji poznawczych.

## Bibliografia

- Albright, D. (1994). Literary and psychological models of the self. W: R. Fivush, U. Neisser (red.), *The remembering self: Construction and accuracy in the self-narrative* (ss. 19-40), Nowy Jork: Cambridge University Press.
- Antas, J. (2008). *O kłamstwie i kłamaniu*. Kraków: Universitas.
- Baron-Cohen, S. (1995). *Mindblindness. An Essay on Autism and Theory of Mind*. Cambridge, Mass.: The MIT Press.
- Baron-Cohen, S. (2009). Rozwój zdolności czytania innych umysłów: cztery etapy. W: A. Klawiter (red.), *Formy aktywności umysłu. Ujęcia kognitywistyczne*, t. II. Warszawa: Wyd. Naukowe PWN.
- Baron-Cohen, S., Baldwin, D. A., Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, 68, 48-57.
- Bremer, J. (2005). *Jak to jest być świadomym?* Warszawa: IFiS PAN.
- Brothers, L. (1995). Neurophysiology of the perception of intentions by primates. W: M. Gazzaniga (red.), *The Cognitive Neurosciences*. Cambridge, Mass.: The MIT Press.
- Bruner, J. (1994). The 'remembered' self. W: R. Fivush, U. Neisser (red.), *The remembering self: Construction and accuracy in the self-narrative* (ss. 41-54). Nowy Jork: Cambridge University Press.
- Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., Wakefield, J. C. (1998). Adaptations, Exaptations, and Spandrels. *American Psychologist*, 53(5), 533-548.
- Chequers, J., Joseph, S., Diduca, D. (1997). Belief in extraterrestrial life, UFO-related beliefs, and schizotypal personality. *Personal and Individual Differences*, 23, 519-521.
- Crick, F., Koch, Ch. (2009). Rama pojęciowa dla świadomości. W: A. Klawiter (red.), *Formy aktywności umysłu. Ujęcia kognitywistyczne*, t. I (ss. 287-323). Warszawa: Wyd. Naukowe PWN.
- Dalla Barba, G. (2009). Temporal consciousness and confabulation: Escape from unconscious explanatory idols. W: W. Hirstein (red.), *Confabulation. Views from neuroscience, psychiatry, psychology and philosophy* (ss. 223-262). Nowy Jork: Oxford University Press.

- Damasio, A. (2005). *W poszukiwaniu Spinozy*, tłum. J. Szczepański, Poznań: Rebis.
- Davidson, D. (1988). Deception and Division. W: J. Elster (red.), *Multiple Self* (ss. 79-92). Cambridge – Nowy Jork – Melbourne – Oslo: Cambridge University Press.
- Davidson, P. S. R., Glisky, E. L. (2002). Is flashbulb memory a special instance of source memory? Evidence from older adults. *Memory*, 10, 99-111.
- Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little Brown.
- Dennett, D. C. (1992). The Self as the Center of Narrative Gravity. W: F. S. Kessel, P. M. Cole, D. L. Johnson, M. D. Hakeł (red.), *Self and Consciousness: Multiple Perspectives* (ss. 275-288). Mahwah, NJ: Erlbaum.
- Dennett, D. C. (2003). Who's On First? Heterophenomenology Explained. *Journal of Consciousness Studies*, 10(9-10), 10-30.
- Dennett, D. C. (2007). Heterophenomenology reconsidered. *Phenomenology and the Cognitive Sciences*, 6, 247-270.
- De Sousa, R. (2004). Is art an adaptation? Prospects for an evolutionary perspective on aesthetic emotions. *The Journal of Aesthetics and Art Criticism*, 62(2), 109-118.
- Ellis, H. D., Young, A. W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157, 239-248.
- Feinberg, T. E. (2009). Confabulation, the self, and ego functions: The ego disequilibrium theory. W: W. Hirstein (red.), *Confabulation. Views from neuroscience, psychiatry, psychology and philosophy* (ss. 91-108). Nowy Jork: Oxford University Press.
- Feinberg, T. E., Roane, D. M. (1997). Anosognosia, completion and confabulation: The neutral-personal dichotomy. *Neurocase*, 3, 73-85.
- Fingarette, H. (1969). *Self-Deception*. Berkeley: University of California Press.
- French, L., Garry, M., Loftus, E. (2009). False memories: A kind of confabulation in non-clinical subjects. W: W. Hirstein (red.), *Confabulation. Views from neuroscience, psychiatry, psychology and philosophy* (ss. 33-66). Nowy Jork: Oxford University Press.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14-21.
- Gallup, G. H., Jr, Newport, F. (1991). Belief in paranormal phenomena among adult Americans. *The Skeptical Inquirer*, 15, 137-146.
- Gazzaniga, M. (2000). Cerebral specialization and interhemispheric specialization. *Brain*, 123, 1293-1326.
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23, 121-123.
- Gibson, J. (1976). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Gloor, P. (1986). The role of the human limbic system in perception, memory and affect: Lessons from temporal lobe epilepsy. W: B. K. Doane, K. E. Livingstone (red.), *The Limbic System: Functional Organization and Clinical Disorders* (ss. 159-169). Nowy Jork: Raven.
- Goldman, A. (1979). What is justified belief? W: G. S. Pappas (red.), *Justification and knowledge* (ss. 1-23). Dordrecht: D. Reidel.
- Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, Mass.: Harvard University Press.

- Goldman, A. (1993). *Philosophical applications of cognitive science*. San Francisco – Oxford: Boulder – Westview Press.
- Goldman, A. (1998). Justification externalism: Pro and con. W: *Routledge Encyclopedia of Philosophy*. London: Routledge.
- Goleman, D. (1999). *Konieczne kłamstwa, proste prawdy*, tłum. M. Madaliński. Warszawa: Albatros.
- Gould, S.J. (1991). Exaptation: A crucial tool for evolutionary psychology. *Journal of Social Issues*, 47, 43-65.
- Gould, S.J., Lewontin, R. (1979). The spandrels of San Marco and the Panglossian paradigm. *Proceedings of the Royal Society of London*, series B, 205: 581-598.
- Gould, S.J., Vrba, E.S. (1982). Exaptation: A missing term in the science of form. *Paleobiology*, 8(1), 4-15.
- Grabowska, A. (2001). Na styku świadomości i nieświadomości: logiczny świat absurdalnych zjawisk. W: R. K. Ohme, M. Jarymowicz, J. Reykowski (red.), *Automatyzmy w procesach przetwarzania informacji* (ss. 25-41). Warszawa: IP PAN, SWPS.
- Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., Hodges, J.R. (2002). Theory of mind in patients with frontal variant fronto-temporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain*, 125, 752-764.
- Hauser, M. D., Chomsky, N., Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569-1579.
- Haxby, J. V., Hoffman, E. A., Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4, 223-233.
- Herzyk, A. (2001). Nieświadomość percepcyjna, poznawcza i emocjonalna z perspektywy neuropsychologii klinicznej. W: R. K. Ohme, M. Jarymowicz, J. Reykowski (red.), *Automatyzmy w procesach przetwarzania informacji* (ss. 44-57). Warszawa: IP PAN, SWPS.
- Herzyk, A. (2005). *Wprowadzenie do neuropsychologii klinicznej*. Warszawa: Scholar.
- Hirstein, W. (2005). *Brain fiction. Self-deception and the riddle of confabulation*. Cambridge, Mass., The MIT Press.
- Hirstein, W., Ramachandran, V.S. (1997). Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society B*, 264(1380), 437-444.
- Hirstein, W., Ramachandran, V.S., (2009). He is not my father, and that is not my arm: Accounting for misidentifications of people and limbs. W: W. Hirstein (red.), *Confabulation. Views from neuroscience, psychiatry, psychology and philosophy* (ss. 109-138). Nowy Jork: Oxford University Press.
- Horniak, J., Rolls, E. T., Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia*, 34, 247-261.
- Hyman, I. E., Billings, F.J. (1998). Individual differences and the creation of false childhood memories. *Memory*, 6, 1-20.
- Hyman, I. E., Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language*, 35, 101-171.

- Jaynes, J. (1976). *The Origins of Consciousness in the Breakdown of the Bicameral Mind*. Nowy Jork: John Wiley.
- Koch, Ch. (2008). *Neurobiologia na tropie świadomości*. Warszawa: Wyd. Uniwersytetu Warszawskiego.
- Kopelman, M. D. (1980). Two types of confabulation. *Neurology, Neurosurgery and Psychiatry*, 43, 461-463.
- Kraepelin, E. (1904). *Lectures on Clinical Psychiatry*. Londyn: Bailliere, Tindal and Cox.
- Loftus, E. (2002). Memory Faults and Fixes. *Issues in Science & Technology*, National Academies of Science, 41-50.
- Loftus, E. F., Palmer, J. C. (1974). Reconstruction of automobile deconstruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behaviour*, 13, 585-589.
- Loftus, E. F., Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25, 720-725.
- Łastowski, K., Reuter, M. (2012). Ewoluuące umysły. Hipotezy, koncepcje, argumenty empiryczne. W: R. Poczobut, M. Miłkowski (red.), *Przewodnik po filozofii umysłu* (ss. 651-688). Kraków: WAM.
- Marciszewski, W. (1972). *Podstawy logicznej teorii przekonania*. Warszawa: PWN.
- Marsh, E. J., Tversky, B. (2004). Spinning the stories of our lives. *Applied Cognitive Psychology*, 18, 491-503.
- Maruszewski, T. (2005). *Pamięć autobiograficzna*. Gdańsk: GWP.
- Maruszewski, T. (2001). *Psychologia poznania*. Gdańsk: GWP.
- Mazzoni, G. A. L., Loftus, E. F., Kirsch, I. (2001). Changing beliefs about implausible events: A little plausibility goes a long way. *Journal of Experimental Psychology: Applied*, 7, 51-59.
- McLaughlin, B. P. (1988). Exploring the Possibility of Self-Deception in Belief. W: A. O. Rorty, B. P. McLaughlin (red.), *Perspectives of Self-Deception* (ss. 29-60). Berkeley: University of California Press.
- McKay, R., Dennett, D. C. (2009). The evolution of misbelief. *Behavioural and Brain Science*, 32, 493-561.
- McNally, R. J., Lasko, N. B., Clancy, S. A., Macklin, M. L., Pitman, R. K., Orr, S. P. (2004). Psychopsychological responding during script-driven imagery in people reporting abduction by space aliens. *Psychological Science*, 15, 493-497.
- Mele, A. (2001). *Self-deception Unmasked*. Princeton: Princeton University Press.
- Millikan, R. (1993). *White queen psychology and other essays for Alice*. Cambridge, Mass.: The MIT Press.
- Milner, B. (1958). Psychological Defects Produced by Temporal lobe Excision. Association for Research in Nervous and Mental Disease. *The Brain and Human Behavior*, Baltimore: Williams and Wilkins.
- Miłkowski, M. (2005). *Konstrukcja umysłu. Naturalizm a intuicje zdrowego rozsądku w filozofii Daniela Dennetta* (praca doktorska pod kierunkiem J. Hołówki). Instytut Filozofii UW.
- Neisser, U. (1982). Snapshots or benchmarks? W: U. Neisser (red.), *Memory Observed: Remembering in Natural Context* (ss. 43-48). San Francisco: W. H. Freeman & Co.

- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, 1 (1), 35-59.
- Neisser, U. (1991). Two Perceptually Given Aspects of the Self and Their Development. *Developmental Review*, 11, 197-209.
- Neisser, U. (1993). The self perceived. W: U. Neisser (red.), *The perceived self. Ecological and interpersonal sources of self-knowledge* (ss. 3-23). Nowy Jork: Cambridge University Press.
- Neisser, U. (1994). Self-narratives: True and false. W: R. Fivush, U. Neisser (red.), *The remembering self: Construction and accuracy in the self-narrative* (ss. 1-18). Nowy Jork: Cambridge University Press.
- Neisser, U., Harsh, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. W: E. Winograd, U. Neisser (red), *Affect and Accuracy in Recall: Studies of "Flashbulb" Memories. Emory Symposia in Cognition*, t. 4 (ss. 9-31). Nowy Jork: Cambridge University Press.
- Northhoff, G., Heinzel, A., Greck, M. de, Bermpohl, F., Dobrowolny, H., Panksepp, J. (2006). Self-referential processing in our brain: A meta-analysis of imaging studies on the self. *NeuroImage*, 31, 440-457.
- Piłat, R. (2009). Oszukiwanie samego siebie. *Kwartalnik Filozoficzny*, 37(3), 31-54.
- Pinker, S. (1997). *How the mind works*. Nowy Jork: W. W. Norton & Co.
- Ramachandran, V.S. (1995). Anosognosia in parietal lobe syndrome. *Consciousness and Cognition*, 4, 22-51.
- Ramachandran, V., Blakeslee, S. (1998). *Phantoms in the brain: Probing the Mysteries of the Human Mind*. Nowy Jork: William Morrow.
- Reed, E. (1994). Perception is to self as memory is to selves. W: R. Fivush, U. Neisser (red.), *The remembering self: Construction and accuracy in the self-narrative* (ss. 278-292). Nowy Jork: Cambridge University Press.
- Reed, G. M., Kemeny, M. E., Taylor, S. E., Visscher, B. R. (1999). Negative HIV-specific expectancies and AIDS-related bereavement as predictors of symptom onset in asymptomatic HIV-positive gay men. *Health Psychology*, 18: 354-363.
- Reed, G. M., Kemeny, M. E., Taylor, S. E., Wang, H.-Y. J., Visscher, B. R. (1994). „Realistic acceptance” as a predictor of decreased survival time in gay men with AIDS. *Health Psychology*, 13: 299-307.
- Reuter, M. (2010). The Great Exaptation. Around a fundamental idea of evolutionary psychology. *Ecological Questions. An International Journal on Controversial Problems of Ecology*, 13, 89-94.
- Reuter, M. (2011). Iluzje poznawcze z punktu widzenia ewolucji ludzkiego umysłu. W: Z. Błaszczak, A. Szczuciński (red.), *Wokół ewolucjonizmu. Dylematy biologów, filozofów i fizyków* (ss. 85-94). Poznań: Oficyna Wydawnicza Batik.
- Reuter, M. (2011). Kognitywistyczne ujęcia konfabulacji: podejście pamięciowe vs. epistemiczne. *Studia z Kognitywistyki i Filozofii Umysłu*, 6(1), 121-142.
- Ryle, G. (1949). *The Concept of Mind*. Chicago: University of Chicago Press.
- Sahdra, B., Thagard, P. (2003). Self-Deception and Emotional Coherence. *Minds and Machines*, 13(2), 213-231.
- Schacter, D. (2003). *Siedem grzechów pamięci*. Warszawa: PIW.

- Schnider, A. (2008). *The confabulating mind: How the brain creates reality*. Oxford: Oxford University Press.
- Sherry, D. F., Schacter, D. L. (1987). The Evolution of Multiple Memory Systems. *Psychological Review*, 94(4), 439-454.
- Springer, S., Deutsch G. (1998). *Lewy mózg, prawy mózg z perspektywy neurobiologii poznawczej*. Warszawa: Prószyński i S-ka.
- Squire, L. R. (1986). Mechanisms of memory. *Science*, 232, 1612-1619.
- Stich, S. (1990). *The fragmentation of reason*. Cambridge, Mass.: The MIT Press.
- Stone, V. E., Baron-Cohen, S., Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10, 640-656.
- Struss, D. T., Gallup, G. G., Alexander, M. P. (2001). The frontal lobes are necessary for „theory of mind”. *Brain*, 124, 279-286.
- Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., McDowell, N. K. (2003). Are self-enhancing cognitions associated with healthy or unhealthy biological profiles? *Journal of Personality and Social Psychology*, 85(4): 605-615.
- Tessier-Lavigne, M. (2000). Visual processing by the retina. W: E. Kandel, J. Schwartz, Th. Jessell (red.), *Principles of neural sciences*. Nowy Jork: McGraw-Hill.
- Tové, M. J. (2003). *An introduction to visual system*. Cambridge – Nowy Jork – Melbourne – Madryt – Cape Town – Singapore – São Paulo: Cambridge University Press.
- Trevarthen, C. (1983). Cerebral mechanisms for language: Prenatal and postnatal development. W: U. Krik (red.), *Neuropsychology of Language, Reading and Spelling* (ss. 45-80). Nowy Jork: Academic Press.
- Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, 907, 114-131.
- Trivers, R. L. (2006). Foreword to Richard Dawkins' *The Selfish Gene*. W: *The selfish gene: 30<sup>th</sup> anniversary edition* (ss. xix-xx). Oxford: Oxford University Press.
- Tulving, E. (1972). Episodic and semantic memory. W: E. Tulving, W. Donaldson (red.), *Organization of memory* (ss. 381-403). Nowy Jork: Academic Press.
- Van Leeuwen, D. S. N. (2007a). The spandrels of self-deception. *Philosophical Psychology*, 20(3), 329-348.
- Van Leeuwen, D. S. N. (2007b). The product of self-deception. *Erkenntnis*, 67, 419-437.
- Van Leeuwen, D. S. N. (2008). Finite rational self-deceivers. *Philosophical Studies*, 139, 191-208.
- Ziemińska, R. (2002). *Eksternalizm we współczesnej epistemologii*. Szczecin: Wyd. Uniwersytetu Szczecińskiego.