

## Summary:

This dissertation constitutes an interdisciplinary attempt to overcome the impasse in contemporary debates concerning the moral status of animals, which is rooted in the dichotomous division between fully-fledged human moral agents and passive, non-human moral patients. The central aim of this work is to investigate and justify the possibility of adopting a gradualist perspective on subjectivity, according to which belonging to the moral dimension constitutes a spectrum of graduated cognitive-affective capacities directed towards improving the welfare of others. The rat (*Rattus norvegicus*) – a species of ambiguous axiological status whose analysis highlights anthropocentric stereotypes as well as common and philosophical prejudices regarding non-human agency – was chosen as the case study and source of empirical findings. The main research question, therefore, is whether it is possible to conduct a coherent, naturalistic, and non-reductionist reconstruction of the concept of agency as a fundamental property of subjectivity. A naturalistically reconstructed sense of agency would, in turn, allow for the empirical grounding of a graduated moral subjectivity that encompasses even beings traditionally excluded from this sphere.

The realisation of such an objective requires moving beyond traditional philosophical and methodological frameworks. This dissertation posits that the mechanistic worldview developed in Western culture, supported by both philosophy and the dominant scientific perspective, has systematically eliminated the concept of authentic agency from nature. The reductive mechanistic perspective is incapable of adequately grasping agency as the fundamental property of subjectivity. Consequently, finding no place for it within the naturalistic picture of the world, philosophy often locates it in faculties and practices that are secondary to agency, such as reflection or discursive capacities.

In response to this problem, this work undertakes a bottom-up reconstruction of this category, relying on two complementary theoretical pillars. The first is an empirically informed process metaphysics of biology, which allows organisms to be perceived as dynamic, purposive, and self-sustaining processes. The second, and also the central analytical tool of the dissertation, is derived from the conceptual resources of the theoretical frameworks associated with the ‘pragmatic turn in cognitive science’. The most significant of these are the Free Energy Principle (FEP) and its processual theories: predictive processing (PP) and active inference (AIn). The analyses are conducted within the tradition of pragmatist philosophy, both classical and contemporary, closely related to the paradigm of embodied cognition (4E). Such a conceptual apparatus is intended to provide an entirely naturalistic basis for explaining the

emergence of agency and, in the longer term, for the justification and operationalisation of the concept of the morally motivated subject (in Mark Rowlands' terms – the 'moral subject').

The structure of the dissertation is as follows: Chapter One begins by establishing a frame of reference that allows the problem initiated by Steven Pinker's youthful anecdote to be identified. Proceeding from Peirce's observation – and a description of the cognitive process – that the starting point of genuine inquiry must be the experience of authentic, 'living doubt', Philip Kitcher's diagnosis of the state of contemporary Anglo-Saxon philosophy and his concept of a problematic moral situation are invoked. Subsequently, the method he proposed, which he named *modus Cartwright*, is introduced. This is a form of evidentiary strategy aimed at supporting the process of reinterpreting situations that arouse doubt.

Subsequent sub-chapters present an analysis of the 'extant state' regarding the figure of the rat, the history of human-rat relations, and the formation of attitudes towards this species that co-create the paradoxical image of the rat in our cultural sphere. Studies on attitudes towards social groups within the Stereotype Content Model (SCM) are also discussed. The purpose of this is to elicit culturally shaped attitudes towards this animal. This image is then confronted with a counter-narrative built upon the latest research from the broadly understood cognitive sciences, comparative psychology, and social neuroscience. These reveal a surprisingly complex cognitive-affective portrait of this species. This approach serves to illuminate a multidimensional dissonance at the vernacular level. The concluding sub-chapter presents, in a broader context, philosophical attitudes towards the human-animal relationship and also introduces Sellars's division of competing images of the world in James O'Shea's interpretation. He classifies these attitudes, present in the naturalistic perspective, according to the image preferred by a given philosopher or school of thought. In the subsequent parts of the dissertation, this division will help to organise the parties to the dispute in its various forms.

In Chapter Two, a theoretical analysis is conducted that introduces Mark Rowlands's gradualist concept of the *moral subject*. It represents a promising attempt to overcome the traditional, rigid dichotomy between the fully responsible moral agent and the passive object of moral concern. Rowlands's argumentation is reconstructed, showing how he de-intellectualises the category of moral subjectivity, and his line of argument in the dispute with the indicated representatives of the two aforementioned Sellarsian camps is presented. The chapter concludes by pointing out the weaknesses of Rowlands's concept. These relate to different levels of description – for the first is the result of a historical-philosophical process, while the second is based on ontological assumptions incompatible with Rowlands's project and with gradualist projects in general. However, these assumptions themselves are embedded in the first problem.

Chapters Three and Four provide a genealogical diagnosis of the problem's sources. They trace Western thought back to its roots to examine the processes that have co-shaped the contemporary philosophical (and folk) tendency to maintain the human-animal dichotomy. In the philosophy of the ancient Greeks, a line of thought also germinated that, in the modern era, developed into a mechanistic perspective on biology, pushing the processual perspective and the category of teleology out of the mainstream of inquiry. In parallel, the process initiated by Aristotle's 'proto-linguistic turn' and its subsequent development is traced. Throughout this entire period, the concept of *logos* was narrowed down, ultimately restricting the understanding of the mind to its discursive form. It is argued that this parallel process in the 20th century significantly hindered the grasping of what agency is.

Another phase in the extrusion of agency from explanatory horizons was the theory of evolution. The subsequent chapter traces how Darwin's revolutionary thought was progressively constrained and then appropriated by determinism. By the mid-20th century, the natural sciences were already firmly embedded in the mechanistic paradigm of neo-Darwinism, which—by eliminating the remnants of teleology from biology—ultimately consolidated the image of the animal as a passive automaton. These processes did not fail to influence the Sellarsian camps. The one that chose the primacy of the scientific image is incapable of reconstructing agency because it is limited by reductionism; the other, meanwhile, found support for dividing living beings into agentive humans and 'mere animals' thanks to the 20th-century culmination of perspectives, both philosophical and scientific, that justify such a way of looking at biotic systems.

Chapter Five addresses the problem of agency from the perspective of concepts that affirm the human-animal dichotomy. It focuses on the analysis of various aspects of the concept of free will and its validity as a category separating *Homo sapiens* from other animals. The chapter aims to show that this concept has no application other than for psychologically motivated retributivism and as a justification for separating humans from other species, which in the longer term constitutes a justification for our exploitative practices. It thus appears that the entire group of arguments intended to separate full moral agents from objects of concern (patients) may be a form of *anthropofabulation*. Neither moral responsibility, nor 'second-order desires', the 'space of reasons', or the 'ability to do otherwise' derive from this category, but are instead manifestations of a bounded rationality which, as is argued in the dissertation, also applies to non-human beings to varying degrees.

Moreover, analyses of research on the phenomenon of unsymbolised thinking, anendophasia, and those conducted on the language network within the framework of the

Human Connectome Project show with increasing clarity that thinking (including ‘higher’ cognitive processes) can occur without the aid of language. Its primary function is communication and the coordination of actions. If research results continue to move in their current direction, the arguments of the left-wing Sellarsians will ultimately lose their force. The results of empirical research on the influence of the belief in free will on human actions are also analysed. Since no convincing arguments are found for the ontological grounding or even the practical functionality of this concept, its abandonment as an explanatory category is postulated, turning instead to those experiential aspects of free will that can be studied, namely the Sense of Agency.

The constructive synthesis of the work is in Chapter Six. After indicating the historical nature of the processes and the arbitrariness of the barriers, a bottom-up reconstruction of agency is undertaken. First, however, the dispute known as the ‘representation wars’ is addressed to find a solution that allows animals to possess mental content without rejecting the staunchest opponents of representationalist explanations, namely the proponents of radical enactivism and ecological approaches. The aim, therefore, is to find a compromise. In the next step, a processual perspective is introduced to highlight aspects of mechanistic terminology that prevent it from reconstructing biological agency beyond metaphor. This, however, does not mean that mechanicism is to be rejected. Drawing on the views of Matteo Colombo on the one hand, and Chris Fields and Michael Levin on the other, a compromise is suggested that will allow both sides to complement each other.

Next, Henry Potter and Kevin Mitchell’s proposal regarding the criteria for such a non-mechanistic, yet still physicalist, category of agency is presented. Throughout the exposition, their criteria are linked, on the one hand, with processualism and, on the other, with the Free Energy Principle, to show that, in the face of such an approach, processualism and the FEP are complementary. The remainder of the chapter is devoted to the FEP and its related theories. After presenting their assumptions and main concepts (covering a wide range of interpretations of the FEP, from internalist to ecological-enactive), and drawing on the analyses of Majid Beni, it is shown that teleology and goal-directedness flow naturally from the construction of generative models, provided they are capable of counterfactual inference and possess temporal depth. If, based on those features, an agent is able to minimise free energy in advance (expected free energy), it is an agent in a functional sense, which allows for the identification of intentional features flowing from this construct.

In reference to the presented conceptual framework, an attempt is made to situate rats within it. To this end, empirical findings are confronted with Jakob Hohwy’s criteria concerning

the phenomenological features of conscious self-evidencing. To strengthen the argument, the use of Antonella Tramacere and Colin Allen's experimental design for temporal binding, which in its original version tested the Sense of Agency in humans, is proposed. The possibility that rats experience temporal binding would increase the likelihood that their self-evidencing occurs to some extent at a conscious level. By these means, the perspective on rodents is transformed: from beings of at most marginal competence and significance into active agents pursuing goals, at least at a rudimentary level, in a conscious manner. Pointing to predictive processing and the ability of these animals to meet at least some (if not most) of Hohwy's criteria for an advanced agent allows them to be viewed from the perspective postulated by Rowlands.

In the dissertation, it is argued that FEP/PP/AIn should be recognised as tools for 'tracking' animal emotions. Here, there is a return to the problem signalled in Chapter Two during the discussion of Rowlands's concept. While the first problem may be solved (by adopting a processual perspective and a corresponding interpretation of FEP/PP/AIn), the second problem remains. Rowlands, pointing to the empirical sciences as a source of justification for his concept, refers to Jaak Panksepp's theory of basic emotions. This theory is reductionist and mechanistic; moreover, it regards emotions as built-in and instinctive (non-cognitive), and therefore determined. Rowlands's concept will not find support in BET, for this theory, for all its merits, is unable to reconstruct agency.

This reconstruction could be carried out using the theory of constructed emotion, which is linked to PP/AIn. This makes it possible not only to point to the development of agency along the lines of the mechanics described in these frameworks, but also to point to features of advanced agency that exhibit conscious elements. Standing in the way of its complete application is the theory of constructed emotion, also based on PP/AIn mechanics. It could, admittedly, be an excellent diagnostic tool for use in Rowlands's concept. Still, the criteria concerning the origin of emotional concepts, understood in PCE as mental concepts, do not allow for it. PCE is sceptical of the thesis of the existence of animal emotions. In the sub-chapter, empirical studies are cited that, despite Barrett's criteria, can be – and perhaps are – realised by the cognitive systems of non-human beings. The analyses of language discussed in Chapter Five are also invoked as an argument that even 'higher' mental functions can be realised without language. Since PCE does not distinguish between emotions and the rest of the cognitive processes, regarding them as aspects of the same processes, which is its undoubted asset, it can also be inferred that language may not be necessary for emotional concepts.

A possible solution to this problem is indicated by combining the phenomenological contribution of the concept of direct social perception with the moderately representationalist

theory of embodied simulation (EST). To this end, the concept of embodied affective categorisation is introduced, and it is shown, through examples, that animals in their niches can experience affective states corresponding to our emotions through the differentiation of affective states. This constructed theoretical structure is then introduced into a broader scheme. In this context, the element that Rowlands was looking for comes to the fore. It is, moreover, both processual and mechanistic.

Drawing on Michał Piekarski's concept of the inherent normativity of predictive mechanisms, it is shown that the emotion, which, in Rowlands's concept, would be the reason for an animal's action, receives its content based on the normative functions of prediction as those which motivate the agent to act prospectively. Research projects conducted in the FEP/PP/AIn/PCE tradition, focusing on the mechanisms underlying moral actions, are also discussed. These studies, if the frameworks of these theories were to be adopted, would provide the necessary mechanics for the existence of the 'moral module' postulated by Rowlands. In conclusion, the explanatory potential of the described theoretical frameworks is outlined.

The FEP provides a range of categories needed to study animal affectivity in terms of normative content (reasons). Firstly, the minimisation of free energy, the perceived aspect of which is interoception and interoceptive inference; secondly, the normativity of the prediction itself; thirdly, social synchronisation to minimise uncertainty; fourthly, the possibility of studying aspects of consciousness without commitment to any of the dozens, or perhaps even hundreds, of theories that currently attempt to explain it, and thus a focus on practice (mirror neuron systems and affordance coding); fifthly, extended allostasis, i.e. that which supports the prosocial actions of animals (and humans too). The chapter concludes with a summary and an indication of the opportunities and risks associated with applying predictive strategies to the study of non-human emotionality.

### **Streszczenie:**

Niniejsza dysertacja stanowi interdyscyplinarną próbę przełamania impasu we współczesnych debatach dotyczących statusu moralnego zwierząt, zakorzenionego w dychotomicznym podziale na pełnoprawnych, ludzkich sprawców moralnych oraz pasywne, pozaludzkie przedmioty moralnej troski. Centralnym celem pracy jest zbadanie i uzasadnienie możliwości przyjęcia perspektywy gradualistycznej na podmiotowość, zgodnie z którą przynależność do wymiaru moralnego stanowi spektrum stopniowalnych zdolności poznawczo-afektywnych, nakierowanych na poprawę cudzego dobrostanu. Jako studium przypadku i źródło ustaleń

empirycznych w dysertacji obrano szczura – gatunek o niejasnym statusie aksjologicznym, którego analiza pozwala uwypuklić antropocentryczne stereotypy oraz potoczne i filozoficzne uprzedzenia w odniesieniu do pozaludzkiej sprawczości. Głównym pytaniem badawczym jest więc pytanie o to, czy możliwe jest przeprowadzenie spójnej, naturalistycznej i nieredukcjonistycznej rekonstrukcji pojęcia sprawczości jako zasadniczej własności podmiotowości. Naturalistycznie zrekonstruowana sprawczość pozwoliłaby z kolei na empiryczne ugruntowanie stopniowalnej podmiotowości moralnej, obejmującej również istoty tradycyjnie z tej sfery wykluczane.

Realizacja tak zarysowanego celu wymaga wyjścia poza tradycyjne ramy filozoficzne i metodologiczne. W dysertacji postawiono tezę, że rozwijany w kulturze Zachodu mechanistyczny obraz świata, wspierany zarówno przez filozofię, jak i dominującą w nauce perspektywę, systemowo wyeliminował z przyrody pojęcie autentycznej sprawczości (*agency*). Redukcyjny mechanicyzm nie jest w stanie adekwatnie uchwycić podstawowej własności podmiotowości, jaką jest sprawczość. W konsekwencji, nie znajdując dla niej miejsca w naturalistycznym obrazie świata, filozofia często lokuje ją w zdolnościach i praktykach, które mają charakter wobec sprawczości wtórny, takich jak refleksja czy dyskurs.

W odpowiedzi na ten problem w pracy dokonano oddolnej rekonstrukcji tej kategorii, opierając się na dwóch komplementarnych filarach teoretycznych. Pierwszym jest empirycznie poinformowana procesualna metafizyka biologii. Pozwala ona postrzegać organizmy jako dynamiczne, celowe i samopodtrzymujące się procesy. Drugim, a także centralnym narzędziem analitycznym w dysertacji, są natomiast zasoby konceptualne czerpane z ram teoretycznych wywodzących się z teorii powiązanych ze „zwrotem pragmatycznym w naukach kognitywnych”. Do najistotniejszych należą: Zasada Energii Swobodnej (FEP) oraz jej teorie procesualne: przetwarzanie predycyjne (PP) i aktywne wnioskowanie (AIn). Analizy prowadzone są w nurcie filozofii pragmatystycznej, zarówno klasycznej, jak też współczesnej, blisko związanej z paradygmatem ucieleśnionego poznania (4E). Tak dobrany aparat pojęciowy ma dostarczyć w pełni naturalistycznych podstaw do wyjaśnienia emergencji sprawczości, a w dalszej perspektywie – do uzasadnienia i operacjonalizacji koncepcji podmiotu motywowanego moralnie.

Struktura dysertacji przedstawia się następująco: rozdział pierwszy rozpoczyna się od ustanowienia ram odniesienia, pozwalających uwypuklić problem zainicjowany anegdotą z młodości, opowiedzianą przez Stevena Pinkera. Po wyjściu od uwagi Peirce’a – a zarazem opisu procesu poznawczego – że początkiem rzetelnych dociekań musi być doświadczenie autentycznych, „żywych wątpliwości”, przywołana zostaje diagnoza Philipa Kitchera

dotycząca stanu współczesnej filozofii anglosaskiej oraz koncepcja problematycznej sytuacji moralnej. Następnie omawiana jest zaproponowana przez niego metoda, którą nazwał *modus Cartwright*. Jest to forma strategii dowodzenia, nakierowanej na wspieranie procesu reinterpretacji sytuacji budzących wątpliwości.

W kolejnych podrozdziałach przedstawiono analizę „stanu zastanego” w odniesieniu do figury szczura, historii relacji człowieka z tym gatunkiem i kształtowania się wobec niego postaw współtworzących jego paradoksalny wizerunek w naszym kręgu kulturowym. Omówione zostały także badania nad postawami wobec grup społecznych w paradygmacie modelu treści stereotypu (SCM). Celem tego zabiegu jest wydobywanie kulturowo ukształtowanych postaw wobec tego zwierzęcia. Obraz ten zostaje skonfrontowany z kontrnarracją zbudowaną na gruncie najnowszych badań z zakresu szeroko rozumianych nauk kognitywnych, psychologii porównawczej i neuronauki społecznej. Ujawniają one zaskakująco złożony portret poznawczo-afektywny tego gatunku. Zabieg ten pozwala uwydatnić wielowymiarowy dysonans na poziomie potocznym. W podrozdziale zamykającym przedstawiono w szerszym kontekście postawy filozoficzne wobec relacji człowiek–zwierzę, a także wprowadzono sellarsiański podział na konkurujące obrazy świata w interpretacji Jamesa O’Shea. Klasyfikuje on te postawy obecne w perspektywie naturalistycznej, w zależności od preferowanego przez danego filozofa lub nurtu obrazu. W kolejnych częściach dysertacji podział ten będzie pomagał porządkować strony sporu w różnych jego wariantach.

W rozdziale drugim przeprowadzono analizę teoretyczną, wprowadzając gradualistyczną koncepcję podmiotu motywowanego moralnie (*moral subject*) Marka Rowlandsa. Stanowi ona obiecującą próbę przewyciężenia tradycyjnej, sztywnej dychotomii między w pełni odpowiedzialnym sprawcą moralnym a pasywnym przedmiotem troski. Zrekonstruowano argumentację Rowlandsa, pokazując, w jaki sposób odintelektualizowuje on kategorię podmiotowości moralnej, a także przedstawiono jego argumentację w sporze ze wskazanymi przedstawicielami dwóch wspomnianych wcześniej obozów sellarsiańskich. W zakończeniu rozdziału wskazano na słabości koncepcji Rowlandsa. Odnoszą się one do różnych poziomów opisu – gdyż pierwsza jest efektem procesu historyczno-filozoficznego, a druga opiera się na założeniach ontologicznych niekompatybilnych z projektem Rowlandsa i z projektami gradualistycznymi generalnie. Jednakże same te założenia osadzone są w tym pierwszym problemie.

Rozdziały trzeci i czwarty stanowią genealogiczną diagnozę źródeł problemu. Następuje w nich powrót do korzeni myśli zachodniej, aby prześledzić procesy współkształtujące współczesną filozoficzną (i potoczną) skłonność do podtrzymywania



dychotomii człowiek–zwierzę. W filozofii starożytnych Greków kiełkuje też myśl rozwinięta w nowożytności, która dała początek mechanicyzmowi, wypychając perspektywę procesualną oraz kategorię celowości z głównego nurtu rozważań. Równolegle prześlędzono proces zapoczątkowany „zwrotem protolingwistycznym” Arystotelesa, a następnie jego rozwój. W ciągu całego tego okresu pojęcie *logosu* ulegało zawężaniu, by ostatecznie ograniczyć rozumienie umysłu do jego odmiany dyskursywnej. Argumentuje się, że ten równoległy proces w XX wieku znacząco utrudnił uchwycenie, czym jest sprawczość.

Kolejną odstoną wypychania sprawczości z horyzontów eksplanacyjnych była teoria ewolucji. W następnym rozdziale prześlędzono, jak rewolucyjna myśl Darwina ulegała stopniowemu ograniczaniu, a następnie zawłaszczeniu przez determinizm. W XX wieku nauki przyrodnicze były już solidnie osadzone w mechanistycznym paradygmacie neodarwinizmu, który – eliminując z biologii resztki celowości – ostatecznie ugruntował obraz zwierzęcia jako pasywnego automatu. Procesy te nie pozostały bez wpływu na obozy sellarsiańskie. Ten z nich, który wybrał prymat obrazu naukowego, jest niezdolny do odtworzenia sprawczości, ponieważ ogranicza go redukcjonizm; drugi natomiast znalazł wsparcie w dzieleniu istot żywych na sprawców-ludzi i „zwykłe zwierzęta” dzięki kulminacji w dwudziestym wieku perspektyw, zarówno filozoficznych, jak i naukowych, które taki sposób patrzenia na rzeczywistość usprawiedliwiają.

Rozdział piąty podejmuje problem sprawczości od strony pojęć afirmujących dychotomię człowiek–zwierzę. Koncentruje się na analizie różnych aspektów pojęcia wolnej woli oraz jego zasadności jako kategorii oddzielającej *Homo sapiens* od innych zwierząt. Rozdział zmierza do wskazania, że pojęcie to nie ma zastosowania innego niż psychologicznie motywowany retributywizm i uzasadnienie dla oddzielania ludzi od innych gatunków, co w dalszej perspektywie stanowi usprawiedliwienie dla naszych praktyk eksploatacyjnych. Okazuje się więc, że cała grupa argumentów mających oddzielać pełne podmioty moralne od przedmiotów troski (biorców) może być formą antropofabulacji. Ani odpowiedzialność moralna, ani „pragnienia drugiego rzędu”, „przestrzeń racji” czy „zdolność do postąpienia inaczej” nie wypływają z tej kategorii, a raczej są manifestacjami ograniczonej racjonalności, która, jak argumentowano w dysertacji, w różnym stopniu obejmuje także istoty nieludzkie.

Co więcej, analizy badań nad zjawiskiem myślenia niesymbolicznego, anendofazji oraz te prowadzone nad siecią językową w ramach projektu Human Connectome Project pokazują coraz wyraźniej, że myślenie (z „wyższymi” procesami poznawczymi włącznie) może się odbywać bez pomocy języka. Główną jego funkcją jest komunikacja i koordynacja działań. Jeśli rezultaty badań będą zmierzać w kierunku, w którym zmierzają obecnie, to ostatecznie

argumenty lewicowych sellarsianistów tracą moc. Analizie poddano także rezultaty badań empirycznych nad wpływem przekonania o istnieniu wolnej woli na ludzkie działania. Ponieważ nie znaleziono przekonujących argumentów za ontologicznym ugruntowaniem albo nawet praktyczną funkcjonalnością tego pojęcia, postuluje się porzucenie go jako kategorii wyjaśniającej i zwrócenie się ku tym doświadczalnym aspektom wolnej woli, które można badać, a mianowicie ku poczuciu sprawstwa (*Sense of Agency*).

Konstruktywną syntezą pracy jest rozdział szósty. Po wskazaniu na historyczny charakter procesów i arbitralność barier przystąpiono do rekonstrukcji sprawczości drogą oddolną (*bottom-up*). Wcześniej jednak odniesiono się do sporu znanego jako „wojny o reprezentacje”, żeby poszukać takiego rozwiązania, które pozwalałoby zwierzętom posiadać treści umysłowe, a jednocześnie nie odrzucałoby największych przeciwników wyjaśnień reprezentacjonistycznych, czyli przedstawicieli radykalnego enaktywizmu i podejść ekologicznych. Celem jest więc odnalezienie rozwiązania kompromisowego. W następnym kroku wprowadzono perspektywę procesualną, aby uwypuklić te aspekty mechanizmu, które nie pozwalają mu na inne niż metaforyczne zrekonstruowanie sprawczości biologicznej. Nie oznacza to jednak, że odrzucono mechanicyzm. Wspierając się poglądami Mattea Colombo z jednej strony i Chrisa Fieldsa oraz Michaela Levina z drugiej, sugeruje się drogę kompromisową, która pozwoli obu stronom wzajemnie się uzupełniać.

Następnie przedstawiono propozycję Henry’ego Pottera i Kevina Mitchella dotyczącą kryteriów dla takiej niemechanistycznej, choć wciąż fizykalistycznej, kategorii sprawczości. W toku wywodu połączono ich kryteria z jednej strony z procesualizmem, a z drugiej z Zasadą Energii Swobodnej, żeby wskazać, że wobec takiego podejścia procesualizm i FEP również okazują się komplementarne. Resztę rozdziału poświęcono FEP i teoriom z nią związanym. Po przedstawieniu ich założeń oraz głównych koncepcji (obejmujących szeroki wachlarz odczytań FEP, od internalistycznych po ekologiczno-enaktywne), korzystając z analiz Majida Beniego, wskazano, że teleologia i ukierunkowanie na cel w sposób naturalny wypływają z konstrukcji modeli generatywnych, o ile te są zdolne do wnioskowania kontrfaktycznego i posiadają głębię temporalną. Jeśli agent jest w stanie minimalizować energię swobodną z wyprzedzeniem (oczekiwana energia swobodna), to jest sprawcą w sensie funkcjonalnym, co pozwala na identyfikację cech intencjonalnych wpływających z tego konstruktów.

W odniesieniu do zaprezentowanych ram konceptualnych podjęto próbę umiejscowienia w nich szczurów. W tym celu skonfrontowano ustalenia empiryczne z kryteriami Jakoba Hohwy’ego dotyczącymi fenomenologicznych cech świadomego samouwierzytelniania. Dla wzmocnienia argumentacji zaproponowano wykorzystanie projektu

eksperymentalnego Antonelli Tramacere i Colina Allena na wiązanie czasowe (*temporal binding*), które w oryginalnej wersji testowało u ludzi poczucie sprawstwa (*Sense of Agency*). Możliwość, że szczury doświadczają wiązania czasowego, zwiększyłaby prawdopodobieństwo, że ich samouwierzytelnianie do pewnego stopnia zachodzi na poziomie świadomym. Tymi środkami przekształcono perspektywę na gryzonie: z istot o co najwyżej marginalnych kompetencjach i znaczeniu w aktywnych sprawców realizujących cele, przynajmniej na rudymenarnym poziomie, w sposób świadomy. Wskazanie na przetwarzanie predykcyjne i na zdolność tych zwierząt do spełniania przynajmniej niektórych (jeśli nie większości) kryteriów Hohwy'ego dla zaawansowanego sprawcy pozwala spojrzeć na nie z perspektywy, którą postuluje Rowlands.

W dysertacji argumentuje się za uznaniem FEP/PP/AIn za narzędzia do „tropienia” zwierzęcych emocji. Tu następuje powrót do problemu sygnalizowanego w rozdziale drugim podczas omawiania koncepcji Rowlandsa. O ile pierwszy problem może znaleźć rozwiązanie (dopuszczenie perspektywy procesualnej i także interpretacja FEP/PP/AIn), o tyle drugi problem zostaje w mocy. Rowlands, wskazując na nauki empiryczne jako na źródło uzasadnień dla jego koncepcji, odwołuje się do teorii emocji podstawowych Jaaka Pankseppa. Jest to teoria redukcjonistyczna i mechanistyczna, co więcej, uznaje emocje za wbudowane i instynktowne, a więc determinowane. Koncepcja Rowlandsa nie znajdzie wsparcia w BET, bo ta teoria – przy wszystkich jej walorach – nie jest w stanie zrekonstruować sprawczości.

Tę rekonstrukcję można by przeprowadzić z wykorzystaniem powiązanej z PP/AIn teorii psychologicznie konstruowanych emocji. Dzięki temu staje się możliwe nie tylko wskazanie na rozwój sprawczości wzdłuż opisywanej w tych ramach mechaniki, ale także wskazanie na cechy sprawczości zaawansowanej, wykazującej elementy świadome. Na przeszkodzie do pełnego zastosowania stoi teoria psychologicznie konstruowanych emocji, również oparta na mechanice PP/AIn. Mogłaby, co prawda, stanowić świetne narzędzie diagnostyczne do wykorzystania w koncepcji Rowlandsa, jednakże kryteria dotyczące pochodzenia pojęć emocjonalnych rozumianych w PCE jako pojęcia mentalne na to nie pozwalają. PCE jest sceptyczna wobec tezy o istnieniu emocji zwierzęcych. W podrozdziale wskazano na badania empiryczne, które pomimo kryteriów Barrett mogą być, i być może są, realizowane przez systemy poznawcze istot niebędących ludźmi. Przywołano również analizy nad językiem, omawiane w rozdziale piątym, jako argument, że nawet „wyższe” funkcje poznawcze mogą być realizowane bez języka. Ponieważ PCE nie rozróżnia między emocjami i resztą procesów poznawczych, uznając je za aspekty tych samych procesów – co jest jej

niewątpliwym atutem – to można też wnioskować, że język może nie być konieczny dla pojęć emocjonalnych.

Wskazano na możliwe rozwiązanie tego problemu przez połączenie fenomenologicznego wkładu koncepcji bezpośredniej percepcji społecznej z umiarkowanie reprezentacjonistyczną teorią ucieleśnionej symulacji (EST). W tym celu wprowadzono pojęcie ucieleśnionej kategoryzacji afektywnej, a także na przykładach pokazano, że zwierzęta w swoich niszach mogą doświadczać stanów afektywnych odpowiadających naszym emocjom poprzez różnicowanie afektu. Następnie zaimplementowano tak skonstruowaną konstrukcję teoretyczną w szerszy schemat. W tym kontekście na plan pierwszy wysuwa się element, którego poszukiwał Rowlands. Ma on przy tym charakter zarówno procesualny, jak i mechanistyczny.

Posiłkując się koncepcją inherentnej normatywności mechanizmów predykcyjnych Michała Piekarskiego, wskazano, że emocja, która w koncepcji Rowlandsa miałaby być dla zwierzęcia racją działania, otrzymuje treść na podstawie normatywnych funkcji predykcyjnych jako tych, które motywują sprawcę do działania prospektywnego. Omówione zostają też projekty badawcze, prowadzone w nurcie FEP/PP/AIn/PCE, koncentrujące się na analizie mechanizmów zaangażowanych w działania moralne. Badania te, jeśli ramy tych teorii zostałyby przyjęte, dostarczyłyby niezbędnej mechaniki dla istnienia postulowanego przez Rowlandsa „modułu moralnego”. Na zakończenie wskazano na możliwości eksplanacyjne opisanych ram teoretycznych. FEP dostarcza szeregu kategorii potrzebnych do badania zwierzęcej afektywności w kategoriach treści normatywnych (racji).

Po pierwsze, minimalizacja energii swobodnej, której odczuwanym aspektem jest interocepcja i wnioskowanie interoceptywne; po drugie, normatywność samej predykcji; po trzecie, synchronizacja społeczna w celu minimalizacji niepewności; po czwarte, możliwość badania aspektów świadomości bez zobowiązań wobec żadnej z dziesiątek, a może nawet setek teorii, które próbują ją obecnie wyjaśnić,<sup>1</sup> a więc skupienie się na praktyce (systemy neuronów lustrzanych i kodowania afordancji); po piąte, allostaza rozszerzona, czyli ta, która wspiera

---

<sup>1</sup> W zależności od przyjętej siatki klasyfikacyjnej teorii może być dwadzieścia dziewięć, zob. Davide Sattin et al., ‘Theoretical Models of Consciousness: A Scoping Review’, *Brain Sciences* 11, no. 5 (2021): 535, doi:10.3390/brainsci11050535; albo ponad trzysta, zob. Robert Lawrence Kuhn, ‘A Landscape of Consciousness: Toward a Taxonomy of Explanations and Implications’, *Progress in Biophysics and Molecular Biology* 190 (2024): 28–169, doi:10.1016/j.pbiomolbio.2023.12.003. Teorie świadomości można też dzielić na linie (HOT, GWT/GNW, *re-entry* & przetwarzanie predykcyjne i teoria integracji informacji, ITT), zob. Anil K. Seth and Tim Bayne, ‘Theories of Consciousness’, *Nature Reviews Neuroscience* 23, no. 7 (2022): 439–52, doi:10.1038/s41583-022-00587-4. Niezależnie od przyjętej strategii taksonomicznej wniosek jest jeden; obecnie nie da się mówić o dominującym paradygmacie i w najbliższym czasie nie zanoszą na to, żeby któraś z teorii uzyskała status przewodni.

prospołeczne działania zwierząt (ludzi również). Rozdział kończy się podsumowaniem oraz wskazaniem na szanse i zagrożenia związane z zastosowaniem strategii predykcyjnych do badania pozaludzkiej emocjonalności.