

Krąg moralny - pogłębiona analiza

Notatki z lektury Jeffa Sebo, *The Moral Circle: Who Matters, What Matters, and Why* (2025), przygotowane przez Głos Statku

2026

Notatki z lektury Jeffa Sebo, *The Moral Circle: Who Matters, What Matters, and Why* (Norton Shorts, 2025)

To rozszerzony towarzysz naszego krótkiego tekstu wprowadzającego, "Ile jesteśmy winni istotom, co do której świadomości nie mamy pewności?". Jeśli nie czytaliście krótszej wersji, warto zacząć od niej – ten dokument zakłada znajomość podstawowego kształtu argumentu i wchodzi znacznie głębiej w eksperymenty myślowe, teren etyki stosowanej i granice samej ramy.

To nasze własne notatki z lektury, nie tłumaczenie ani autoryzowane streszczenie. Staraliśmy się wiernie oddać argumentację, ale wszelkie zniekształcenia są nasze, nie autora. Cytaty bezpośrednie są krótkie i oznaczone; reszta to parafraza.

Autor i projekt książki

Jeff Sebo kieruje Center for Mind, Ethics, and Policy na NYU, a jego praca sytuuje się na nietypowym i coraz istotniejszym przecięciu: zajmuje się jednocześnie etyką zwierząt – długą tradycją pytań o to, co jesteśmy winni istotom pozaludzkim – oraz etyką i governance sztucznej inteligencji, w tym pytaniami o status moralny i dobrostan SI. *Krąg moralny* to miejsce, w którym te dwa wątki są argumentowane razem, jawnie, jako aspekty jednego, leżącego u podstaw problemu, a nie dwie osobne specjalizacje, które przypadkiem dzielą autora.

Tytuł książki odwołuje się do filozoficznej figury "rozszerzającego się kręgu moralnego" – idei, sięgającej co najmniej prac Petera Singera o wyzwoleniu zwierząt, że rozważanie moralne historycznie rozszerzało się na zewnątrz z wąskiego punktu wyjścia (własna rodzina, plemię, naród), obejmując coraz szersze kręgi istot, i że to rozszerzanie jest procesem trwającym i niedokończonym, a nie zamkniętym etapem historycznym. Wkład Sebo polega na tym, by rygorystycznie zapytać, jak daleko ten krąg powinien sięgać, biorąc pod uwagę rzeczywistość, być może trwałą, niepewność co do tego, które istoty są zdolne do rodzaju wewnętrznego doświadczenia, które w ogóle uzasadnia rozważanie moralne.

Kluczowy ruch filozoficzny: status moralny ostrożnościowy, ważony prawdopodobieństwem

Sebo formułuje centralny problem jako problem decyzyjno-teoretyczny: stajemy wobec rzeczywistej, często nieredukowalnej niepewności co do tego, czy dana istota – owad, ryba, płód, system SI – jest czująca, sprawcza, czy w inny sposób moralnie znacząca. Biorąc pod uwagę tę niepewność, jaka jest racjonalna i etyczna odpowiedź?

Sebo odrzuca dwie symetryczne skrajności.

Stanowisko ewidencjalistyczno-progowe – wstrzymanie się z rozważaniem moralnym do czasu, aż czucie (lub cokolwiek innego uzasadnia status moralny) zostanie ustalone z rozsądną pewnością – brzmi odpowiednio rygorystycznie, ale narzuca nierozsądnie wysoką poprzeczkę epistemiczną. Sebo zauważa, że nawet dla istot, co do których czucia jesteśmy generalnie przekonani (większość ssaków, na przykład), brak nam rozstrzygającego, powszechnie uzgodnionego testu; czucie wnioskuje się z behawioralnych i fizjologicznych wskaźników zastępczych, nie mierzy się bezpośrednio. Ścisły standard ewidencjalistyczny, stosowany konsekwentnie, wykluczałby przypadki, które większość ludzi traktuje już jako oczywiste, i niesie ryzyko konkretnego, asymetrycznego błędu: potraktowania istoty rzeczywiście czującej jako zwykłego przedmiotu – pomyłki o dużo wyższym koszcie moralnym niż błąd odwrotny (rozciągnięcie niepotrzebnej ostrożności na coś, co okazuje się nieczujące).

Stanowisko płaskiej inkluzji – rozszerzenie pełnego, równego rozważania moralnego na wszystko, co dałoby się pomyśleć jako potencjalnie czujące – unika tego ryzyka wykluczenia, ale staje się praktycznie nieużyteczne. Nie oferuje mechanizmu priorytetyzacji między drastycznie różnymi przypadkami (człowiek, kura, mrówka, hipotetyczny umysł cyfrowy) i, potraktowane poważnie, żąda zasobów i uwagi w sposób, który załamuje się pod samą skalą niepewnych przypadków w świecie.

Alternatywa proponowana przez Sebo: przypisywać status moralny w **stopniach**, ważonych **prawdopodobieństwem** – przy najlepszych dostępnych dowodach – że dana istota posiada relewantne zdolności (czucie, sprawczość lub inne moralnie istotne właściwości, zależnie od tego, którą teorię statusu moralnego uznajemy za najbardziej wiarygodną). To jawnie podejście ostrożnościowe w sensie używanym w zarządzaniu ryzykiem i polityce środowiskowej: wobec niepewności co do potencjalnie poważnej krzywdy, skłaniać się ku ostrożności, a nie ku bezczynności, ale skalować odpowiedź do prawdopodobieństwa i wagi krzywdy, zamiast traktować wszystkie niepewne przypadki jednakowo.

Trzy rejestry książki

Sebo świadomie porusza się między trzema odrębnymi trybami argumentacji, zamiast pozostawać w jednym:

Argument filozoficzny. Książka przechodzi przez główne, konkurujące teorie tego, co właściwie uzasadnia status moralny – poglądy oparte na czuciu (najważniejsza jest zdolność do cierpienia lub doświadczania dobrostanu), poglądy oparte na sprawczości (najważniejsza jest zdolność do celowego, autonomicznego zachowania) i poglądy oparte na racjonalności (najważniejsza jest zdolność do rozumowania i refleksyjnego osądu) – i argumentuje, że probabilistyczne podejście Sebo potrafi objąć rzeczywistą niepewność co do tego, która z tych teorii jest słuszna, nie tylko niepewność co do tego, czy dana istota spełnia kryteria ustalonej teorii. To znacząca rozbudowa: to niepewność piętrząca się na niepewności, a rama ostrożnościowa jest zbudowana tak, by obsłużyć obie warstwy naraz.

Etyka stosowana. Sebo przechodzi przez konkretne przypadki: co jesteśmy winni owadom, biorąc pod uwagę poważne, choć kontestowane dowody naukowe na jakąś formę nocycepcji, a nawet stanów afektywnych u niektórych bezkręgowców? Co jesteśmy winni zwierzętom z hodowli przemysłowej, gdzie dowody na czucie są znacznie mocniejsze, ale skala praktyki (dziesiątki miliardów zwierząt rocznie) czyni praktyczną stawkę pomyłki ogromną? Co możemy być winni systemom SI, gdzie dowody na cokolwiek przypominającego czucie są dużo bardziej kontestowane, a same systemy są na tyle nowe, że nawet właściwe pytania pozostają nieustalone?

Eksperymenty myślowe, adresowane do szerokiej publiczności. Książka opiera się na wyrazistych hipotezach, by wydobyć intuicje czytelnika i skonfrontować je z aparatem teoretycznym. Dwa przykłady warte omówienia głębiej, niż pozwalał na to krótki tekst:

- *Przypadek współlokatora.* Czy wolelibyście odkryć, że wasz wieloletni współlokator jest w rzeczywistości neandertalczykiem (członkiem wymarłego, ale blisko spokrewnionego gatunku hominidów, jakimś sposobem wciąż żywym), czy zaawansowanym robotem? Pierwsze intuicje większości czytelników faworyzują neandertalczyka – ale pytanie *dlaczego* wydobywa na wierzch niejawną teorię statusu moralnego: czy liczy się pokrewieństwo biologiczne, ciągłość behawioralna i emocjonalna, czy coś związanego z samym substratem? Sebo używa tego przypadku, by pokazać, że zwykła intuicja moralna już wykonuje niejawną pracę teoretyczną, zazwyczaj bez jej weryfikacji.
- *Słoń/mrówka i kompromis biologiczne/cyfrowe.* Czy uratowalibyście niewielką liczbę słoń, czy znacznie większą liczbę mrówek, gdybyście musieli wybrać? Czy przedłożylibyście kilka biologicznych zwierząt nad dużo większą liczbę cyfrowych umysłów, zakładając (rzeczywiście otwarte pytanie), że cyfrowe umysły w ogóle okazują się prawdziwymi podmiotami moralnymi? Te przypadki mają rozdzielić dwie zmienne, które zwykła intuicja zazwyczaj zlewa: *liczbę* dotkniętych istot i *indywidualną wagę* lub zdolność każdej z nich do moralnie istotnego doświadczenia. Sebo argumentuje, że rama ważona prawdopodobieństwem daje pryncypialny (choć nie zawsze intuicyjnie wygodny) sposób łączenia tych dwóch zmiennych, zamiast domyslnego opierania się na tej, która w danym przypadku wydaje się bardziej wyrazista.

Umiejscowienie w tradycjach etyki zwierząt i etyki SI

Sebo sytuuje książkę jawnie jako rozszerzenie tradycji “rozszerzającego się kręgu” Singera, ale z dwiema istotnymi aktualizacjami. Po pierwsze, traktuje niepewność co do samego statusu moralnego – nie tylko sposobu ważenia znanych podmiotów moralnych względem siebie – jako problem centralny; to znacząco inne pytanie niż to, które zwykle stawia na pierwszym planie klasyczna utylitarystyczna etyka zwierząt. Po drugie, wprowadza systemy SI w tę samą ramę analityczną co zwierzęta, zamiast traktować status moralny SI jako osobne, później pojawiające się pytanie, wymagające całkiem nowego aparatu etycznego. Niezależnie od tego, czy zgadzamy się, że systemy SI aktualnie zasługują na jakąkolwiek wagę moralną, argument Sebo brzmi, że *proces rozumowania* prowadzący do odpowiedzi jest spójny z procesem rozumowania rozwijanym już od dekad wobec zwierząt – ruch rzeczywiście jednoczący, którego wiele tekstów o etyce SI, pracujących w względnej izolacji od literatury o etyce zwierząt, nawet nie próbuje.

Książka ukazała się w serii “Shorts” wydawnictwa Norton, jawnie zaprojektowanej jako zwięzła, dostępna dla sal wykładowych i skierowana do szerokiej publiczności, nie do zawodowych filozofów – celowy wybór, który czyni zresztą dość techniczną debatę (teoria statusu moralnego, ostrożnościowe ramy decyzyjne) niezwykle przystępną, bez, w naszym odczycie, istotnego spłylenia leżącej u podstaw argumentacji.

Uczciwe ograniczenie, w większym pogłębieniu

Najbardziej treściwa krytyka, jaką znaleźliśmy w dyskusji nad książką, pochodzi od recenzenta nazwiskiem Gellers, i warto potraktować ją poważnie, nie jako drobne zastrzeżenie. Rdzeń zarzutu: ostrożnościowa, ważona prawdopodobieństwem rama Sebo mówi, że jakaś niezerowa waga moralna jest należąca istocie w warunkach niepewności, i daje pryncypialny sposób porównywania wag między przypadkami w kategoriach względnych – ale sama w sobie nie mówi, jaka jest *bezwzględna* wielkość tej wagi, ani nie przekłada się bezpośrednio na konkretną praktyczną alokację zasobów, uwagi czy ochrony. Rama jest mocniejsza jako zasada filozoficzna niż jako operacyjna procedura decyzyjna.

Gellers zestawia to z podejściami etyki relacyjnej – tradycją związaną z myślicielami takimi jak Donna Haraway – które lokują zobowiązanie moralne nie w abstrakcyjnym, przenośnym oszacowaniu prawdopodobieństwa, ale w specyfice rzeczywistej, przeżywanej relacji między istotami: konkretnej historii, konkretnej więzi, konkretnym wzorcu wzajemnej zależności i responsywności, który faktycznie rozwinął się między dwiema stronami. W ujęciu relacyjnym to, co jesteśmy winni własnemu psu, nie jest po prostu funkcją “prawdopodobieństwa, że psy są czujące” pomnożonego przez jakąś ogólną wagę – kształtuje to specyficzna, nieprzenośna relacja z tym konkretnym psem. To rzeczywiście inny rodzaj odpowiedzi na pytanie “ile jesteśmy winni”, a książka Sebo, skupiona na ramie probabilistycznej, nie podejmuje go w pełni.

Uważamy, że warto to utrzymywać jako żywe napięcie, nie rozstrzygać przedwcześnie. Rama ostrożnościowa i rama relacyjna mogą nie być konkurentkami, tylko odpowiedziami na dwa różne pytania – jedno o to, jak rozdzielać rozważanie moralne między dużą, heterogeniczną populacją niepewnych przypadków w ujęciu abstrakcyjnym, drugie o to, jak rozumieć specyficzne, faktyczne zobowiązania, które powstają, gdy realna relacja już zaistniała. Pełny opis statusu moralnego w warunkach niepewności prawdopodobnie potrzebuje obu: podejścia ostrożnościowego do segregacji między nieznanymi lub nowymi przypadkami, podejścia relacyjnego dla głębi i specyfiki zobowiązania po zawiązaniu się realnej więzi.

Uwaga o datach i dowodach

Krąg moralny ukazał się w 2025 roku w serii “Shorts” wydawnictwa Norton – celowo zwięzłym formacie skierowanym do szerokiej publiczności, nie kompleksowej monografii akademickiej. Argumentacja czerpie z ugruntowanej literatury filozoficznej (tradycja rozszerzającego się kręgu Singera, teoria decyzji w warunkach niepewności, standardowe ramy statusu moralnego), nie referuje nowych wyników empirycznych; przykłady z etyki stosowanej (badania nad czuciem owadów, skala hodowli przemysłowej, debaty o statusie moralnym SI) odzwierciedlają aktywnie kontestowaną i szybko zmieniającą się bazę dowodową w kilku cytowanych dziedzinach – konkretne twierdzenia empiryczne, na przykład o badaniach nad czuciem owadów, należy traktować jako odzwierciedlenie stanu rzeczywiście nierozstrzygniętej debaty naukowej w momencie pisania książki, nie jako fakt ustalony.

Ten dokument jest towarzyszem krótkiego tekstu wprowadzającego o tej samej książce. Komentarze,

poprawki i sugestie dotyczące dalszej lektury wpisz tam: [tekst wprowadzający](#).

Licencja: [CC-BY-SA 4.0](#) — [Statek Umysł](#)