

Heinzelmann (2024): Replik zu den Kommentaren, *Zeitschrift für philosophische Forschung*

Patricia Rich und Philipp Brüllmann danke ich für ihre klugen und konstruktiven Kommentare zu *Weakness of Will and Delay Discounting*. Im Folgenden werde ich nacheinander besonders auf ihre Kritikpunkte eingehen.

Philipp Brüllmann bezweifelt „in drei Schritten“ (S.2), dass die Monografie eine Erklärung von Willensschwäche liefert, und fordert mich zur Präzisierung meines Projekts auf. Dem versuche ich hiermit nachzukommen. Brüllmanns ersten Schritt gehe ich mit: Diachrone Diskontierungsmodelle können synchrone Fälle von Willensschwäche nicht erfassen und sie höchstens „wegerklären“ (Brüllmann S.4). Dagegen hat er *per se* nichts einzuwenden, worin wir also auch einig sind. Denn eine „Wegerklärung“ scheint mir nur dann problematisch zu sein, wenn sie eine Konzeption für ein nachweislich existierendes Phänomen verschwinden ließe, d.h., wenn wir unsere Terminologie dort einschränken würden, wo wir sie brauchen. Aber das tut die Monografie ja nicht, im Gegenteil: WW schlägt genau deswegen die Konzeption der willensschwachen Diskontierung vor, weil die empirischen Wissenschaften mit ihr das Phänomen der Willensschwäche beschreiben und erforschen.

Allerdings hält Brüllmann – zweitens – diese Konzeption für unabhängig problematisch.

Willensschwache Präferenzumkehr oder Diskontierung wie im Donut-Beispiel unterscheidet sich darin von nicht-willensschwacher Präferenzumkehr wie im Peter-Beispiel (vgl.

Heinzelmann 2023, Kap. 5.5), dass nur im ersten Fall der*die Akteur*in einer Versuchung unterliege. Der Begriff der Versuchung impliziere „eine Form der Gleichzeitigkeit“ (Brüllmann S.5), ergo könne ein diachrones Diskontierungsmodell sie nicht beschreiben. Ob Versuchung ausschließlich synchron ist, sei hier einmal dahingestellt. Ob Willensschwäche Versuchung voraussetzt, ist umstritten – während sie in einigen Ansätzen wie etwa Holtons

(Heinzelmann 2023, Kap. 3.5) vorkommt, fehlt sie in anderen, so bei Davidson (Kap. 3.3) und Mele (Kap. 3.4). Dass das diachrone Modell vermeintlich synchrone Willensschwäche nicht beschreiben kann, möchte ich aber bestreiten. So sind die scheinbar synchronen willensschwachen Handlungen, auf die sich Brüllmann konzentriert, wie womöglich alle Handlungen zeitlich ausgedehnt und damit überhaupt nur diachron beschreibbar. Außerdem erlaubt uns die Konzeption willensschwacher Diskontierung als kognitive Verzerrung (engl. „cognitive bias“), die relevanten Fälle anhand eines spezifischen psychologischen Mechanismus zu identifizieren und von anderen zu unterscheiden. Dieses Vorgehen mag uns Philosophierenden fremd erscheinen, die wir im Lichte unseres kartesischen Erbes generell unserer Intuition und Introspektion vertrauen. Tückischerweise aber sind kognitive Verzerrungen weder Introspektion noch Intuition noch Selbstwissen allein zugänglich. Vielleicht halten sich implizite Vorurteile gegen Minderheiten deswegen hartnäckig in der akademischen Philosophie. Wie Brüllmann (S.5) schreibt, könnten bei Peter irgendwelche kognitive Verzerrungen vorliegen; allerdings scheint für seinen Präferenzwechsel nicht diejenige ursächlich zu sein, die spezifisch der willensschwachen Diskontierung zugrunde liegt. Für diesen spezifischen Präferenzwechsel ist weiterhin nicht allein „die verzerrte Abwertung des Apfels“ entscheidend (*pace* Brüllmann S.5), sondern die Präferenz bezüglich *beider* Optionen bzw. ihre (diskontierten) Werte *relativ zueinander*: Zunächst ist der Apfel *wertvoller als* der Donut, dann ist es umgekehrt. Der wesentliche Unterschied zwischen Peter und der Donutesserin, der auch Bestandteil unserer alltäglichen Beispiele ist, wird also erfasst: Er besteht in den unterschiedlichen psychologischen Mechanismen, welche die jeweiligen Präferenzwechsel verursachen.

Drittens geht Brüllmann auf die im ersten Teil meiner Monografie beschriebenen drei Merkmale ein, „die sich mit dem Konzept der Willensschwäche verbinden“ (S.6):

Willensschwäche ist rätselhaft, ein Konflikt und ein Fehler oder Defekt (allerdings *besteht* der Konflikt nicht im Handeln wider eigene Standards, *was wiederum* rätselhaft und irrational ist). Brüllmann kritisiert (ebd.): „Bei der Anwendung der Kriterien auf den eigenen Vorschlag scheint Heinzelmann diesen Zusammenhang aber aufzulösen, indem sie einen unabhängigen Begriff der Irrationalität einführt.“ Dass der Begriff der Irrationalität von demjenigen der Willensschwäche unabhängig ist, wie es die Debatten über rationale Akrasia veranschaulichen, dürfte wohl kaum ein Problem sein. Vielmehr scheint es Brüllmann (S.6) problematisch zu finden, dass irrationale „Diskontierung in einem anderen Sinn irrational ist als die Willensschwäche“. Diese Behauptung erscheint mir allerdings zweifelhaft. Natürlich sind die Begriffe von Irrationalität, Diskontierung und Willensschwäche voneinander „unabhängig“ (ebd.). Kapitel 7 geht der Frage nach, ob der Defekt der Willensschwäche als ein Defekt der Rationalität präzisiert werden kann. Dafür ziehe ich Theorien der Rationalität heran. Weil es nicht „*einen* unabhängigen Begriff der Irrationalität“ gibt (Brüllmann S.6, meine Hervorhebung), sondern eine Vielzahl miteinander konkurrierender Theorien, kann sich WW nur mit einer Auswahl derselben befassen. Wenn Rationalität beispielsweise eine Art von Inkohärenz mentaler Zustände ist, so wäre willensschwache Diskontierung irrational, weil sie mit (diachron) inkohärenten Präferenzen einhergeht. In diesem Sinne wäre etwa die Akratikerin irrational, die zuerst einen Apfel einem Donut vorzieht, diese Präferenz dann aber umkehrt. Gemäß einer Theorie, die Rationalität als Kohärenz versteht, ist die Diskontierung wegen der mit ihr einhergehenden Inkohärenz der Präferenzen irrational. Die Willensschwäche ist aus genau demselben Grund irrational. Wie das Beispiel veranschaulicht, sind irrationale Diskontierung und irrationale Willensschwäche also im gleichen Sinne irrational.

Bleiben wir beim Thema Irrationalität, auf das auch Patricia Rich einen Schwerpunkt ihres Kommentars legt. Dabei konzentriert sich Rich auf Willensschwäche als kognitive Verzerrung. Kapitel 6 meiner Monografie argumentiert auf Basis ökonomischer Modelle und empirischer Befunde, dass ein auf Unsicherheit reagierender kognitiver Prozess willensschwacher Diskontierung zugrunde liegt und willensschwache Handlungen dementsprechend als verzerrt betrachtet werden können. Kognitive Verzerrungen sind in der Philosophie vor allem in Bezug auf (moralische) Verantwortung und ihre Rolle bei Diskriminierung diskutiert worden; ich ziehe sie für mein Verständnis von Willensschwäche heran. Generell liegt der Einfluss kognitiver Verzerrungen auf unsere Handlungen außerhalb unseres Bewusstseins und unserer Kontrolle; zudem gibt es situative und individuelle Unterschiede (vgl. Rich S.2). So dürften viele die diffuse Gefahr einer Demenzerkrankung beim Donutessen in Richs Beispiel diskontieren, sich aber bereits über den Zusammenhang zwischen beidem unsicher sein (der allerdings unabhängig ist von zeitlicher Diskontierung). Gerade weil „Menschen viele entscheidungsrelevante Ereignisse (wie Demenz) als tief ungewiss betrachten“ (Rich S.2), dürfte Ungewissheit ihre Entscheidungen fast durchweg beeinflussen, wie das Diskontierungsmodell es besagt.

Unter welchen Bedingungen sind solche Entscheidungen also irrational? Die Beantwortung dieser Frage ist davon abhängig, was wir unter Irrationalität verstehen. Rich wirbt für ein ökologisches Verständnis von (Ir)rationalität, das Entscheidungen mitsamt ihrer Umgebung in den Blick nimmt und in meiner Monografie nur kurz Erwähnung findet. Dieses Verständnis scheint mir aber in der Tat vielversprechend, unter anderem, weil es in der philosophischen und der empirischen Forschung gleichermaßen Anwendung finden kann. Neben dem ökonomischen, biologischen und evolutionären Umfeld könnte auch das kulturelle, soziale und biografische Berücksichtigung finden. Der in der Monografie vorgeschlagene Ansatz

besagt, dass zeitliche Diskontierung unter anderem von individueller Unsicherheitstoleranz abhängt, die wiederum von unserer Umgebung geprägt wird. Einem ökologischen Verständnis von Rationalität zufolge könnte eine geringere Toleranz in stark unsicherheitsbehafteten Umgebungen rational sein, in Umgebungen mit geringer Unsicherheit aber irrational. Ein umfassenderes Verständnis von willensschwacher Diskontierung mithilfe ökologischer Rationalitätstheorien zu entwickeln wäre ein spannendes und fruchtbares Forschungsprojekt. Es scheint auch Parallelen mit dem Ansatz in der Moralpsychologie aufzuweisen, moralische Intuitionen und Handlungen im Kontext ihres evolutionären und soziokulturellen Umfelds zu erforschen und zu bewerten. Allerdings geht es über den Rahmen von WW weit hinaus – und sollte idealerweise in einem darauf aufbauenden Forschungsvorhaben weiterverfolgt werden.

Abschließend lässt sich aus Sicht der Autorin resümieren, dass unser Symposium vor allem die grundlegende Natur und Ausbaufähigkeit der in *Weakness of Will and Delay Discounting* vorgelegten Konzeption willensschwacher Diskontierung aufgezeigt hat, etwa hinsichtlich synchron erscheinender Willensschwäche (vgl. Brüllmann) und ihrer Bewertung aus Perspektive ökologischer Rationalitätstheorien (vgl. Rich). Es wäre mir eine Freude und Ehre, das Projekt im Austausch mit den am Symposium teilnehmenden und anderen Kolleg*innen weiterzuführen.