Heinzelmann, N.C. (2024): Précis zu Weakness of Will and Delay Discounting*, Zeitschrift für philosophische Forschung* 78(2)

Die Monografie widmet sich einem philosophischen Dauerbrenner: den Phänomenen, die unterschiedliche Disziplinen wahlweise als „Willensschwäche“ („weakness of (the) will“), „Akrasia“, „Inkontinenz“ („incontinence“), „Versagen der Selbstkontrolle“, „Impulsivität“ oder „Zeitdiskontierung“ („delay discounting“) bezeichnen. Sie konzentriert sich auf Willensschwäche und Zeitdiskontierung, welche die Philosophie einerseits und die Verhaltenswissenschaften andererseits bisher weitgehend unabhängig voneinander erforscht haben. Dabei verfolgt sie zwei Ziele. Zum einen möchte sie Neulinge in die relevanten Grundlagen beider Fächer einführen. Dementsprechend widmet sich Teil I der Monografie der Einführung in die Philosophie der Willensschwäche für Nicht-Philosoph*innen und Teil II einer Einführung in die Verhaltenswissenschaft der Zeitdiskontierung. Zum anderen sucht die Monografie eine Theorie der willensschwachen Zeitdiskontierung zu entwickeln, wofür Teil III die beiden Ansätze zusammenführt.

Da Sie, liebe*r Leser*in dieser philosophischen Fachzeitschrift, vermutlich keiner Einführung in die Philosophie der Willensschwäche bedürfen, können wir an dieser Stelle direkt zu Teil II übergehen. Dieser stellt zunächst handlungstheoretische Annahmen der empirischen Verhaltensforschung vor, welche klassischen Diskontierungsmodellen zugrunde liegen (Kapitel 4, „Agency in descriptive research“). Zu ihnen gehört insbesondere das Modell eines *homo oeconomicus*, dessen Präferenzen, Wahlverhalten und Wertzuschreibungen axiomatisch zusammenfallen. Das heißt: Wenn ein *homo oeconomicus* A gegenüber B bevorzugt, schreibt er A einen größeren erwarteten Wert oder Nutzen zu und wählt A, wenn er sich zwischen A und B entscheiden soll. Vor diesem Hintergrund, so argumentiere ich, ist Willensschwäche am besten als Umkehrung eine Präferenz zu verstehen. So zieht es beispielsweise eine gesundheitsbewusste, aber willensschwache Person zunächst vor, auf einen der Donuts zu verzichten, die bei einem Philosophieworkshop herumgereicht werden, um lieber später in der Kaffeepause den Apfel zu essen, den sie sich mitgebracht hat. Als ihr jedoch ein Donut angeboten wird, greift sie zu. Sie scheint also zunächst den Apfel dem

Donut vorzuziehen, ihn mehr wertzuschätzen und ihn auch zu wählen. Dann aber kehrt sich die Rangfolge um.

Diese grobe Darstellung präzisieren klassische Diskontierungsmodelle der Verhaltenswissenschaften mathematisch, wie Kapitel 5 („Discounting“) ausführt. Zeitdiskontierung bedeutet ganz grob, dass ein Nutzen oder Wert abhängig von seiner zeitlichen Verzögerung („delay“) variiert. Typischerweise ist Zeitdiskontierung negativ: Je größer die Verzögerung ist, desto kleiner ist der Wert. So ist für uns in der Regel heute ein Wert, den wir morgen erhalten, größer als derselbe Wert, den wir erst in einem Jahr erhalten. Der mit seiner zeitlichen Verzögerung *diskontierte* Wert ist umso kleiner, je größer die zeitliche Verzögerung ist. Mathematisch lässt sich dieser diskontierte Wert als Produkt eines Diskontfaktors und des nicht-diskontierten Wertes beschreiben. Dieser Diskontfaktor wiederum lässt sich als Funktion von Zeit oder Verzögerung auffassen. Er hängt also von der zeitlichen Verzögerung ab.

In unserem Beispiel schwinden sowohl der Wert des Apfels als auch der des Donuts mit wachsender Verzögerung: Je länger wir auf den Genuss warten müssen, desto weniger wertvoll ist er für uns. Dies kann dazu führen, dass ein verzögerter Apfelkonsum wertvoller ist als ein verzögerter Donutkonsum und jemand dementsprechend ersteren präferieren und sich bei einer Wahl zwischen Apfel und Donut für die gesündere Option entscheiden würde. Umgekehrt kann aber auch der unmittelbar bevorstehende Donutkonsum wertvoller sein als ein verzögerter Apfelkonsum. Willensschwäche, verstanden als Umkehr von Präferenzen, liegt dann vor, wenn zwei diskontierte Werte über die Zeit hinweg oder mit sich verändernder zeitlicher Verzögerung ihre relative Rangfolge wechseln: in einer Konstellation ist ein Wert größer als der andere, in einer zweiten ist es umgekehrt. In unserem Beispiel diskontiert die Person zunächst den Wert des Apfels als auch den des Donuts. Der diskontierte Wert des Apfels liegt dabei über dem diskontierten Wert des Donuts. Als ihr der Donut angeboten wird, die Verzögerung sehr klein und der Wert des Donuts kaum noch diskontiert wird, ist der nach wie vor diskontierte Wert des Apfels kleiner. Dementsprechend dreht sich die Präferenz der Person um und ihr Wahlverhalten ändert sich.

Willensschwaches Verhalten auf diese Weise zu beschreiben, hat einige Nachteile, von denen hier nur zwei exemplarisch genannt seien. Erstens ist Willensschwäche als Umkehr einer Präferenz zu verstehen. Aber nicht alle Präferenzwechsel sind willensschwach und nicht jedes willensschwache Verhalten geht mit einem Präferenzwechsel einher. Deswegen konzentriere ich mich fortan auf die Schnittmenge der *willensschwachen Diskontierung*, die sowohl Präferenzwechsel entsprechend dem Modell der Zeitdiskontierung mit sich bringt als auch die in Teil I beschriebenen philosophischen Kriterien der Willensschwäche erfüllt.

Zweitens gibt es Fälle, die das Diskontierungsmodell nicht beschreiben kann, die aber als Paradebeispiele für willensschwaches Verhalten gelten und mit Präferenzwechsel einhergehen. Zu ihnen gehören die aus den 1970er-Jahren bekannten, sogenannten „Marshmallow"-Fälle: Einem Kind wird angeboten, entweder sofort eine kleinere Belohnung zu erhalten, etwa ein Marshmallow, oder auf eine größere Belohnung wie zwei Marshmallows zu warten. Das Kind scheint sich willensschwach zu verhalten, wenn es zunächst eine Weile auf die größere Belohnung wartet, sich dann aber doch für die kleinere entscheidet. Das klassische Diskontierungsmodell kann diesen Fall nicht beschreiben. Denn wenn der diskontierte Wert der größeren Belohnung schon anfangs größer war als der (nicht-diskontierte) Wert der kleineren, dann sollte er nach einer Weile, wenn die zeitliche Verzögerung etwas geschrumpft ist, immer noch größer sein. Aber das kann nicht sein; da das Kind die kleinere Belohnung wählt, muss deren Wert nun größer sein als der diskontierte Wert der größeren Belohnung.

Aus philosophischer Perspektive könnten diese Probleme ausreichen, um den Ansatz abzulehnen, willensschwaches Verhalten durch Diskontmodelle zu beschreiben. Dies wirft die Frage auf, welche Ziele wir in der Philosophie verfolgen. Suchen wir ausschließlich, der traditionellen Begriffsanalyse gemäß, notwendige und hinreichende Bedingungen für „Willensschwäche"? Wollen wir einen Begriff entwickeln, der an die Theorien und Befunde der Verhaltenswissenschaften anschlussfähig ist? Meines Erachtens sind beide Anliegen wichtig; die vorliegende Monografie konzentriert sich allerdings auf das zweite. Denn Diskontierungsmodelle sind, um nur wenige Beispiele zu nennen, in den Wirtschaftswissenschaften und dem Finanzwesen zur Ökonomie des Klimawandels, in der Psychologie zur Erforschung von Selbst- und Impulskontrolle, in der Psychiatrie zur

Beschreibung von Sucht, in den Neurowissenschaften für Modelle wertbasierter Entscheidungen sowie in der Biologie zum Verständnis tierischen Verhaltens und Lernens von zentraler Bedeutung. Die Philosophie kann mit einem überdisziplinären Verständnis willensschwachen Verhaltens zu all diesen Forschungsfeldern beitragen und sie untereinander vernetzen.

Außerdem legt das Modell der willensschwachen Diskontierung einen neuen philosophischen Ansatz nahe: ein Verständnis von Willensschwäche als kognitiver Verzerrung („bias“). Diesen Vorschlag entwickelt Teil III. Zunächst führt er in Kapitel 6 („Describing weakness of will“) in ein komplexeres Diskontierungsmodell aus der Ökonomie ein, das unter anderem auch Marshmallow-Fälle beschreiben kann. Dieses Modell versteht den Diskontfaktor als abhängig nicht nur von zeitlicher Verzögerung, sondern auch von Risiko und Unsicherheit. Es nimmt an, dass sich eine Belohnung früher oder später als erwartet oder gar nicht materialisieren kann. Diese Einflüsse sind der handelnden Person womöglich nicht bewusst und sie kann sie auch nicht direkt kontrollieren, ähnlich wie bei optischen Täuschungen und impliziten Vorurteilen. Die sich so manifestierende Willensschwäche ist also ein subtileres Phänomen als das viel diskutierte Unterliegen der Vernunft im Kampf mit dem Begehren. Historisch findet der Ansatz Vorbilder etwa in Platons *Protagoras* und bei Descartes, doch die zeitgenössische Verhaltenswissenschaft kann ihn auch empirisch unterstützen.

Dieser Ansatz kann auch neue Perspektiven in weitere philosophische Debatten einbringen, etwa in jene zur Rationalität (Kapitel 7, „Criticizing weakness of will“). Willensschwäche gilt traditionell als paradigmatischer Fall praktischer Irrationalität. Kognitive Verzerrungen werden dagegen unterschiedlich bewertet; während sie in Teilen der Verhaltensökonomie als Beleg für die irrationale Natur des Mängelwesens Mensch zu gelten scheinen, sehen Theorien der ökologischen („ecological“) oder begrenzten („bounded“) Rationalität sie als rational im Kontext nicht-idealer Gegebenheiten wie Zeitmangel oder Ressourcenknappheit. Die vorliegende Monografie argumentiert, dass willensschwache Diskontierung gängigen Standards von Rationalität gemäß – etwa verstanden als Kohärenz mentaler Einstellungen oder als angemessene Reaktion auf Gründe – irrational sein kann. Allerdings lässt sich die Frage, inwiefern und warum dies so ist, nur relativ zu der jeweils favorisierten

Rationalitätstheorie beantworten. Beispielsweise kann willensschwache Diskontierung irrational sein, weil sie mit Inkohärenz von Präferenzen oder der Verletzung prudentieller Gründe einhergeht. Theorien der ökologischen Rationalität gemäß kann es rational sein, in sehr unsicherheits- und risikobehafteten Umgebungen stark zu diskontieren, etwa in sozial instabilen Verhältnissen, in denen man sich besser auf das Marshmallow in der Hand als auf Versprechen einer besseren Zukunft verlassen sollte. In anderen Umständen ist eine solche Verzerrung jedoch nachteilig. Der Ansatz, als Gegenstand von Rationalität Akteure*innen in ihrer Umgebung zu bewerten, erlaubt eine differenzierte Bewertung von willensschwacher Diskontierung und verwandten Phänomenen.

Er zeigt auch Wege auf, wie sich den Problemen der willensschwachen Diskontierung auf alltäglicher und auf gesellschaftlicher Ebene praktisch begegnen lässt (Kapitel 7, „Practical takeaways“). Das mathematische Modell besagt, dass Zeitdiskontierung von vier Faktoren abhängt: den relativen erwarteten Werten der Entscheidungsoptionen, der zeitlichen Verzögerung, Unsicherheit und der individuellen Empfänglichkeit für sie. Diese Faktoren lassen sich direkt oder indirekt beeinflussen, etwa im Kleinen durch Änderungen des physischen oder temporalen Entscheidungsumfelds oder im Großen durch gesellschaftspolitische Maßnahmen zur Erhöhung der Stabilität und Sicherheit. Damit können sie uns bei individuellen wie kollektiven Schwierigkeiten von Lifestyle-Problemen bis hin zur Klimakatastrophe unterstützen.