

Diese Auflistung verschiedener Umstände macht deutlich, dass sich weitergehende Forschung als Ziel setzen sollte, den Moderator des finanziellen Anreizsystems eindeutig auszulegen, den Untersuchungsgegenstand klar zu definieren und die Experiment-Bedingungen so zu gestalten, dass die Ergebnisse generalisierbar oder mindestens vergleichbar sind.

3. Empirisches Forschungsdesign – Die Wirkung finanzieller Anreize in Abhängigkeit des Aufgabenkontexts

Die vorangegangenen Abhandlungen haben dargelegt, dass für beide zentralen Moderatoren weiterer Forschungsbedarf besteht. Dass sich diese Arbeit nicht direkt eignet, um das Human-Framing tiefgehender zu untersuchen, wurde bereits erklärt. Stattdessen soll das folgende Experiment Klarheit zur Wirkungsrichtung finanzieller Anreizsysteme in unterschiedlichen Ausprägungen in Entscheidungssituationen im Management bringen.

Um diese Aussagefähigkeit zu erlangen, sollen die Anreize auf die Teilnehmenden in unterschiedlichen Aufgaben-Szenarien angewendet werden. Wie in Kapitel 2.1 angeführt wurde, haben sich bereits verschiedene Studien mit dem Effekt unterschiedlicher Aufgabencharakteristika auseinandergesetzt. Castelo et al. sowie Wang et al. haben dabei detaillierte Aufgabenstellungen zu spezifischen Domänen verwendet. Diese bilden allerdings nicht den Management-Bereich ab und lassen sich schlecht generalisieren. Daher sollen analog zum Vorgehen von Hertz und Wiese übergeordnete Aufgabenschemata verwendet werden, die soweit möglich reale Entscheidungsszenarien nachstellen und sich anhand grober Charakteristika unterscheiden.

3.1 Forschungsfrage

Aus den Überlegungen ergibt sich die folgende **Forschungsfrage**:

Welchen Einfluss haben unterschiedliche finanzielle Anreize in verschiedenen Aufgabenkontexten auf zentrale abhängige Variablen algorithmisch unterstützter Entscheidungsprozesse?

3.2 Variablen und Hypothesen

Zunächst werden die Ausprägungen der **unabhängigen Variablen** bestimmt. Die Typen finanzieller Anreize werden analog Greiner et al. als kein finanzieller Anreiz (gleichbleibende Bezahlung), Entlohnung abhängig von der Leistung sowie turnierbasierter Lohn festgelegt. Da die Aufgabenvarianten nicht aus der Literatur übernommen werden können, müssen eigene Schemata abgeleitet werden. Dazu werden die Charakteristika einer Zahl von Experimenten verglichen und anschließend gruppiert.¹ Es ergeben sich die folgenden Aufgabenschemata:

¹ Das genaue Vorgehen zur Ableitung der Aufgabenschemata ist in Anhang A nachzulesen.

- i. Standard-Aufgabe (strukturiert, simpel, low-stakes)
- ii. Risikoreiche Entscheidung (strukturiert, komplex, high-stakes)
- iii. Unstrukturierte Schätzaufgabe (unstrukturiert, komplex, low-stakes)²

Als **abhängige Variablen** werden die Nutzungsintensität, der Aufwand und die Leistung erhoben.³ Der Aufwand, gemessen in der Bearbeitungszeit, wird zwar in der Literatur wenig thematisiert, kann jedoch ggf. zur Erklärung des resultierenden Effekts dienen: Ashton begründet in seiner Studie das Backfiring finanzieller Anreize damit, dass diese verändern würden, wie Teilnehmende eine Aufgabe wahrnehmen. Eine simple Aufgabe erscheine durch den Leistungsdruck der Anreize als herausfordernd, sodass sich Teilnehmende eigene, aufwändige Lösungsstrategien überlegen würden, statt den Vorschlag der Entscheidungshilfe zu nutzen, der aufgrund der Überlegenheit algorithmischer Instrumente die bessere Leistung gebracht hätte (vgl. Ashton, 1990, S. 150, 173). Mit der Erhebung der Bearbeitungsdauer kann Ashtons Erklärungsversuch getestet werden. Nach seinem Ansatz müsste das Backfiring außerdem vor allem bei einfachen Standard-Aufgaben auftreten, da diese aufwändiger gelöst würden als nötig.

H1a – Finanzielle Anreize in Standard-Aufgaben bewirken ein Absenken von Nutzungsintensität und Leistung bei einem Anstieg des Aufwands.

H1b – Finanzielle Anreize in risikoreichen Entscheidungen und unstrukturiert Schätzaufgaben steigern Nutzungsintensität und Leistung bei gleichbleibendem Aufwand.

Da aus den Studien keine statistisch signifikanten Ergebnisse zu Unterschieden in den Wirkungen von Leistungs- und Turnier-Anreizen hervorgegangen sind, wird jetzt im Gegenteil angenommen, dass beide Varianten denselben Effekt haben.

H2 – Leistungs- und turnierbasierte Anreize unterscheiden sich nicht in ihrer Wirkung.

3.3 Gestalt des Experiments

Wie die meisten Studien im Forschungsfeld zur Algorithmus-Aversion wird auch die vorliegende in einer Labor-Umgebung durchgeführt, da diese Form die meiste Kontrolle bietet und die Umsetzung als Feldexperiment in realen Organisationen voraussichtlich an Geheimhaltungsvorschriften und einer geringen Anzahl bereitwilliger Teilnehmenden scheitern würde. Die Studie wird als **3x3 faktorielles, Between-subjects Experiment** durchgeführt. Dabei stellen die beiden bereits besprochenen unabhängigen Variablen in ihren drei Ausprägungen die Dimensionen dar. Dementsprechend wird es neun Versuchsgruppen geben, die jeweils isoliert von den anderen Gruppen einer Faktorkombination ausgesetzt werden.

² Beispielhafte Aufgabenstellungen sind in Anhang B nachzulesen.

³ Das genaue Vorgehen zur Bestimmung der abhängigen Variablen ist in Anhang C nachzulesen.

3.4 Studienablauf

Nach den Vorüberlegungen und Vorbereitungen beginnt die Studie mit der **Registrierung** der Teilnehmenden vor Ort im Labor sowie einer **Einverständniserklärung** zur Verarbeitung der erhobenen Daten. Die angemeldeten Personen werden dann zufällig zu einer der neun Versuchsgruppen zugeteilt und hinsichtlich ihrer Aufgaben im Rahmen des Experimentes **instruiert**. Anzumerken ist dabei, dass der eigentliche Zweck der Studie nicht Teil der Vorbemerkungen sein sollte. Eine thematische Einführung zum Thema Algorithmus-Aversion könnte bei den Teilnehmenden die Haltung gegenüber der Entscheidungshilfe verfälschen. Trotzdem darf die Einweisung keine Täuschung enthalten. Die Teilnehmenden könnten z.B. neutral darüber informiert werden, dass sie verschiedene Instanzen eines Aufgabentyps zu lösen haben, für die ihnen eine Algorithmus-basierte Unterstützung bereitgestellt wird.

Nach dem Bestehen eines **Aufmerksamkeits-Tests** beginnt die Phase des **Hauptexperiments** und damit auch die Zeitmessung für die Variable des Aufwands. Jede Aufgabeninstanz wird analog zu Greiner et al. in zwei Stufen durchgeführt. Den initialen Wert überlegen sich die Teilnehmenden selbst. Erst im zweiten Schritt erfolgt die Beratung durch den Algorithmus. Das zwei-phasige Design ermöglicht eine saubere Berechnung des WOA. Nach der Übermittlung des finalen Werts für die letzte Aufgaben-Instanz stoppt die Zeitmessung. Es folgt eine **Post-Befragung**, die zum einen die Wahrnehmung der Aufgabencharakteristika Strukturiertheit, Komplexität und Schwere der Konsequenz bei den Teilnehmenden abfragt. Damit kann sichergestellt werden, dass beobachtete Effekte tatsächlich mit spezifischen Aufgabenmerkmalen zusammenhängen. Zum anderen sollen die Teilnehmenden selbst angeben, wie viel sie annehmen, die Entscheidungshilfe genutzt zu haben. Die Studie endet mit der **Auszahlung** des Teilnahme-Entgelts sowie des ggf. erzielten Bonus.

3.5 Datenauswertung

Die Datenauswertung erfolgt in mehreren Schritten mit zunehmendem Erkenntnisgewinn. Zunächst werden die Mittelwerte und Standardfehler der abhängigen Variablen für einen deskriptiven Überblick über alle Versuchsgruppen dargestellt. Anschließend wird eine Regression durchgeführt, um die Haupt- und Interaktionseffekte der Moderatoren sowie deren Signifikanz zu analysieren. Da dadurch noch keine Aussage zur statistischen Bedeutung von Unterschieden zwischen einzelnen Faktorstufen getroffen werden kann, werden t-Tests für die identifizierten relevanten Differenzen innerhalb einer Gruppe durchgeführt. Mittels Korrelations-Untersuchung wird festgestellt, ob die implizite These, eine höheres WOA führe direkt zu einer höheren Leistung, zutrifft. Zuletzt wird der Zusammenhang zwischen der gemessenen Intensität der Algorithmus-Nutzung und der Selbsteinschätzung ebenfalls mittels Korrelation geprüft.