Heiders Balancetheorie:

Eine strukturalistische Repräsentation und ein Computermodell

Klaus Manhart

Überarbeitete Fassung aus meiner Dissertation

„KI-Modelle in den Sozialwissenschaften“,
Oldenbourg-Verlag, München, 1995

www.klaus-manhart.de
mail@klaus-manhart.de

München, Januar 2009

1. Einleitung

Die Balancetheorie in der Fassung von Fritz Heider hat in den fünfziger bis siebziger Jahren eine entscheidende Initialisierungsrolle in den Sozialwissenschaften gespielt. Zum einen hat sie eine große Zahl von Experimenten angeregt und die Formulierung anderer - komplexerer - Gleichgewichtstheorien initiiert. Zum anderen hat sie die Anregung gegeben, genauer über Formalisierungsmöglichkeiten in den Sozialwissenschaften nachzudenken (Sukale 1971: 48).

In diesem Beitrag wird die Balancetheorie mit dem Apparat der strukturalistischen Wissenschaftstheorie rekonstruiert und interpretiert. Auf Basis dieser Rekonstruktion lässt sich ein einfaches Computermodell implementieren.

Wir gehen im folgenden so vor, dass wir das Ausgangsmodell von Heider zunächst informell vorstellen und anschließend seine logische Struktur in einer präzisierten, strukturalistischen Deutung angeben. An diesem sehr einfachen Beispiel lassen sich die allgemeinen Begriffe und Verfahren, wie sie in einführenden Artikeln in den wissenschaftstheoretischen Strukturalismus beschrieben sind, konkreter verdeutlichen.

Das Computermodell, eine PROLOG-Implementierung, stellen wir in zwei Varianten vor: einmal als Datengenerierungs-Modell und zum andern als Programm, in dem ausschließlich strukturalistische Prädikate definiert werden.

2. Informelle Darstellung der Balancetheorie[1]

Die Wurzeln der Theorie von Heider - und damit aller anderen Balancetheorien - liegen zum ersten in Spinozas Ethik, zum zweiten in dem feldtheoretischen Ansatz von Kurt Lewin und zum dritten in der Gestaltpsychologie, die in den dreißiger Jahren von Koffka, Köhler und Wertheimer begründet wurde. Von allen drei Einflussfaktoren wollen wir nur den gestaltpsychologischen kurz andeuten, da er uns als der zentralste erscheint.[2]

Gestaltpsychologen gehen davon aus, dass Menschen nach der „guten Gestalt“ oder „guten Form“ streben und gewisse Konfigurationen wegen ihrer Einfachheit und Kohärenz bevorzugen. Dieser Wunsch nach der guten Gestalt hat zur Folge, dass Personen versuchen, Inkonsistenzen zu vermeiden und sich um eine geordnete und kohärente Sicht der Welt bemühen.

Nach der - schon im Altertum erkannten - Konsistenzregel wird die Verbindung von Einzeleindrücken zu einem Gesamteindruck so vorgenommen, dass kein Widerspruch enthalten ist (Witte 1989: 324).[3] Dem Konsistenzprinzip zufolge trachten Personen im allgemeinen danach, ihre Einstellungen und Überzeugungen (kurz: Kognitionen) untereinander und mit ihrem Verhalten widerspruchsfrei zu organisieren. Beispielsweise sind die beiden Kognitionen „Ich treibe gern Sport“ und „Sport ist gesund“ konsistent. Hingegen sind die beiden Kognitionen „Ich rauche gern“ und „Rauchen ist gesundheitsschädlich“ inkonsistent. Inkonsistene Kognitionen erzeugen eine gewisse Spannung und einen psychischen Druck zur Herstellung von Konsistenz: „Nimmt eine Person z.B. an sich Verhaltensweisen wahr, die ihren Einstellungen widersprechen, so befindet sie sich in einem Zustand kognitiver Inkonsistenz. Einen solchen Zustand wird sie ... als unangenehm, da Spannung erzeugend, erleben. Sie wird deshalb motiviert sein, die beteiligten Kognitionen (wiederum) in ein konsistentes und damit spannungsfreies Verhältnis zu überführen, indem sie einzelne dieser Kognitionen oder auch ihr Verhalten ändert“ (Stahlberg/Frey 1987: 214).

---

[1] Für die informelle Darstellung der Balancetheorie vgl. Heider (1977), Stahlberg/Frey (1987), West/Wicklund (1985) und Irle (1975).

[2] Die wichtigsten Ideen der Gestaltpsychologie in Zusammenhang mit Sozialpsychologie finden sich in Deutsch/Krauss (1976).

[3] Witte (1989: 325) weist darauf hin, dass bereits Aristoteles das unten von Heider behandelte Prinzip formulierte, wonach die Freunde unserer Freunde auch unsere Freunde sein werden.

Auf der Basis des Konsistenzprinzips wurden in der Sozialpsychologie verschiedene theoretische Konzepte entwickelt, denen die grundlegende Annahme gemeinsam ist, dass Inkonsistenz Spannung erzeugt und eine Tendenz zur Herstellung von Konsistenz impliziert. Zu den bekanntesten sozialpsychologischen Konsistenztheorien gehört - neben der hier behandelten Balancetheorie - die Dissonanztheorie von Festinger (1978).

Die Gleichgewichtstheorie ist eine spezifische Konsistenztheorie. Sie beschäftigt sich mit der Widersprüchlichkeit in den Beziehungen, die eine Person zwischen sich und anderen Elementen ihrer Umwelt wahrnimmt. Ausgangspunkt ist Fritz Heiders Artikel „Attitudes and Cognitive Organization“ aus dem Jahr 1946. In diesem Aufsatz macht Heider die grundlegende Annahme, dass soziale Wahrnehmung den eben beschriebenen gestaltähnlichen Strukturprinzipien folgt und *ausgeglichene oder balancierte Zustände gegenüber unausgeglichenen oder unbalancierten Zuständen präferiert* werden. Gleichgewichtige Systeme werden als angenehm und kohärent im Sinne der Gestaltpsychologie empfunden: „Mit Gleichgewichtszustand ist eine Situation gemeint, in der die Relationen zwischen den Größen harmonisch zueinander passen; es gibt keinen Drang zu einer Veränderung“ (Heider 1977: 238).

Die Zustände „Ausgeglichenheit“ und „Unausgeglichenheit“ werden in der Heider-Theorie auf eine kognitive Konfiguration bestehend aus drei Elementen angewendet. Diese Konfiguration wird gebildet aus einer wahrnehmenden Person p, einer anderen Person o und einem impersonalen Objekt x. x kann dabei „eine Situation, ein Ereignis, eine Idee oder irgendein Ding etc.“ sein (Heider 1946: 107). Genau genommen unterscheidet Heider (1946) noch drei andere Systeme, nämlich dyadische Beziehungen mit zwei Elementen und triadische Konfigurationen mit drei Personen. In Anlehnung an Sukale (1971: 49) wollen wir diese einfacheren Systeme aber außer Betracht lassen.

Ein Beziehungssystem mit den drei Elementen p-o-x bildet eine Triade, die sich grafisch wie in Abb. 1 darstellen lässt. Wichtig ist, dass die Situation immer aus der Sicht von p analysiert wird und p, o und x dabei als Einheiten in der kognitiven Struktur von p repräsentiert sind.

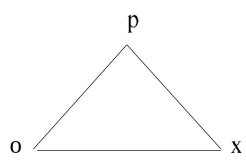

Abb. 1: Grafische Darstellung von Triaden

Um festzustellen, ob eine bestimmte Triade als balanciert oder unbalanciert wahrgenommen wird, unterteilt Heider die in der Triade bestehenden Relationen in positive und negative Einheits- oder Gefühlsbeziehungen („unit“ oder „sentiment relations“). Einheitsrelationen U bestimmen, ob je zwei Einheiten zusammengehören oder nicht, wie etwa p besitzt x oder p und o sind sich ähnlich.

Gefühlsrelationen L beziehen sich auf eine Bewertung von o oder x durch p und sind Sympathie- oder Antipathierelationen, z.B. p schätzt x oder o verabscheut x. Die Identifikation als Einheits- bzw. Gefühlsrelation ist für die Entscheidung, ob ein System als balanciert wahrgenommen wird oder nicht, aber nicht ausschlaggebend. Bestimmend ist nur die Anzahl der beteiligten positiven und negativen Beziehungen.

Die formale Definition von Gleichgewicht und Ungleichgewicht ist einfach: „Eine Triade ist im Gleichgewicht, wenn alle drei Relationen positiv sind oder wenn zwei Relationen negativ sind und eine positiv ist. Ungleichgewicht tritt dann auf, wenn zwei Relationen positiv sind und eine negativ ist“ (Heider 1977: 240). Der Fall mit drei negativen Relationen wird von Heider nicht eindeutig entschieden und als mehrdeutig und unbestimmt bezeichnet (Heider 1946: 110; 1977: 240).

Abb. 2 veranschaulicht alle möglichen Permutationen positiver und negativer Relationen in einer Triade. Von den so entstehenden acht Fällen sind die Triaden (a-d) im Gleichgewicht, die Triaden (e-g) im Ungleichgewicht und (h) ist die indefinite Triade.

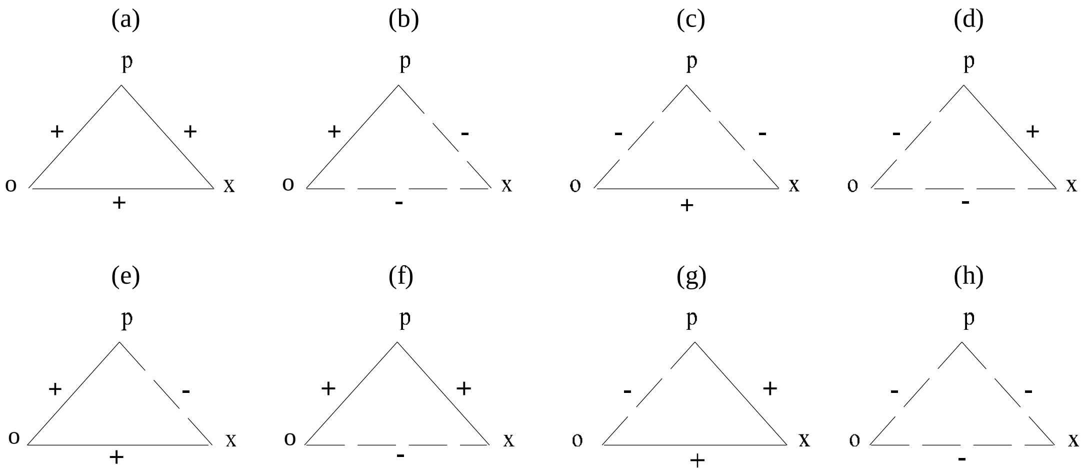

Abb. 2: Balancierte und unbalancierte Triaden nach Heider. Durchgezogene, mit '+' markierte Linien bedeuten positive, gestrichelte, mit '-' markierte Linien, negative Relationen (nach: Witte 1989: 327). Die Triaden (a-d) sind balanciert, die Triaden (e-g) sind unbalanciert und Triade (h) ist „unbestimmt“.

Die balancierten Triaden lassen sich metaphorisch aus der Sicht von p (und unter der Erweiterung, dass x ebenfalls eine Person ist) wie folgt interpretieren:
(a) Der „Freund“ meines „Freundes“ ist mein „Freund“.
(b) Der „Feind“ meines „Freundes“ ist mein „Feind“.
(c) Der „Freund“ meines „Feindes“ ist mein „Feind“.
(d) Der „Feind“ meines „Feindes“ ist mein „Freund“.

Die folgenden Situationen wären Beispiele für balancierte und unbalancierte Triaden (vgl. Heider 1977: 212 und 242):

(1) p ist mit o befreundet und p wählt eine Partei x, die o ablehnt.
(2) zwei Personen p und o sind befreundet und haben die gleiche Einstellung zur Kirche x.
(3) Sie (p) hasst immer die Personen (o), mit denen sie arbeiten (x) muß.
(4) John (p) billigt einen Zug (x) seines politischen Gegners (o).

Konfiguration (1) korrespondiert dabei der unbalancierten Triade (f), (2) der balancierten Triade (a) oder (b), je nachdem, ob p und o eine beidseitig positive oder beidseitig negative Einstellung zur Kirche haben und (3) und (4) der unbalancierten Triade (g). Heider legt eine umfassende Beispielsammlung balancierter und unbalancierter Situationen aus Literatur und Alltagsleben vor.

Das Balanceprinzip von Heider lässt sich nun leicht angeben. Da gleichgewichtige Systeme als angenehm und harmonisch empfunden werden, sind diese über die Zeit stabil und werden nicht verändert. Unbalancierte Systeme werden hingegen als disharmonisch und spannungsvoll empfunden, sind instabil und tendieren zu einer Änderung in gleichgewichtige Systeme: „If no balanced state exists, then forces towards this state will arise ... If a change is not possible, the state of imbalance will produce tension“ (Heider 1946: 107-108). ***Heiders fundamentale theoretische Aussage besteht also darin zu behaupten, dass ungleichgewichtige Zustände einen psychischen Druck erzeugen und dazu neigen, gleichgewichtig zu werden***. Dieses einfache Gesetz der Gleichgewichtstendenz bildet das grundlegende Prinzip, welches allen Versionen von Balancetheorien gemeinsam ist.

Die Herstellung von Gleichgewicht geschieht im einfachsten Fall durch Änderung der Relationen. In Beispiel (1) könnte ein Gleichgewichtszustand dadurch erzeugt werden, dass p versucht, o zur Wahl einer anderen Partei zu bewegen. Eine andere

Möglichkeit wäre, dass p seinerseits seine Parteienpräferenz überdenkt. Eine dritte Möglichkeit ist, dass p seine Relation zu o ändert. Die Balancetheorie macht aber keine Vorhersagen darüber, *welche* Veränderungen in einem unbalancierten System zu erwarten sind. Sie behauptet nur, *dass* unbalancierte Triaden dazu tendieren, gleichgewichtig zu werden; wie dies bewerkstelligt wird, bleibt unspezifiziert.

Eine Vielzahl von Experimenten wurde durchgeführt zur Überprüfung der balancetheoretischen Annahmen; wir wollen auf einige davon hinweisen.

Grundsätzlich können sich empirische Untersuchungen auf zwei Aspekte der Behauptung von Heider stützen. Man kann eine Triade nach dem Grad des gefühlsmäßigen Wohlbefindens beurteilen oder nach ihrer Stabilität. Beide Perspektiven sind eine Konsequenz der theoretischen Annahme, dass balancierte Zustände als angenehm empfunden werden und stabil bleiben.

In der Literatur stützt man sich in der Regel auf die erste Möglichkeit. Beispielsweise ließ Jordan (1953) 64 hypothetische p-o-x Triaden nach dem Grad ihres Wohlbefindens einstufen. Es zeigte sich wie erwartet, dass die vier balancierten Triaden signifikant positiver bewertet wurden als die unbalancierten. Allerdings ergaben sich auch Unterschiede *innerhalb* von balancierten Triaden: die Triaden (a) und (b) von Abb. 2 wurden als signifikant angenehmer erlebt als die anderen balancierten Triaden.

Viele Untersuchungen weisen auf unheilvolle Implikationen des Balanceprinzips für das Alltagsleben hin. Als Beispiel zitieren wir ein Experiment von Landy/Aronson (1969). In diesem Experiment mussten Studenten (p) die Rolle von Geschworenen übernehmen, die einen Angeklagten (o) wegen einer schweren Straftat (x) verurteilen sollten. Experimentell variiert wurde die Attraktivität des Angeklagten (attraktiv - neutral - unattraktiv). Nach der Balancetheorie müssten die Geschworenen bei einem sympathischen Angeklagten dahin tendieren, diesen in jeder Hinsicht positiv zu sehen und also seine Tat milde beurteilen. Bei einem unsympathischen Angeklagten würde Gleichgewicht hingegen hergestellt werden, indem der Angeklagte negativ gesehen wird und eine entsprechend schwere Strafe erhält. Die Ergebnisse entsprachen der Theorie: sympathische Angeklagte wurden zu durchschnittlich 5.5, unsympathische zu 7.1 Jahren Haft verurteilt.

In der Untersuchung von Landy/Aronson wurde ein unsympathischer Angeklagter infolge seiner mangelnden Attraktivität streng bestraft. In einer anderen Studie von

Lerner/Simmons (1966) wurde umgekehrt eine Frau, die eine strenge Strafe erhielt, infolge der Strafe als unsympathisch wahrgenommen. Die gleiche Frau wurde hingegen als attraktiv angesehen, wenn sie nicht weiter bestraft wurde. Die Autoren erklären ihr Ergebnis balancetheoretisch damit, dass eine attraktive Frau, die bestraft wird (Elektroschocks erhält), ein Ungleichgewicht erzeugt und der Gleichgewichtszustand durch Umpolung der Attraktivitätskognition wieder hergestellt wird.

Beide Untersuchungen lassen auf die Neigung von Menschen schließen, andere entweder durchgängig mit positiven oder durchgängig mit negativen Eigenschaften zu versehen: „Haben wir von jemandem im großen und ganzen eine gute Meinung, sind wir geneigt, Verhalten und Eigenschaften dieses Menschen insgesamt in einem positiven Licht zu sehen. Sind wir jemandem dagegen nicht wohlgesonnen, werden wir wahrscheinlich auch Eigenschaften, die wir neu an ihm entdecken, in diese Einstellung einbeziehen“ (West/Wicklund 1985: 157).

Weitere Untersuchungen werden angegeben z.B. in West/Wicklund (1985) und Witte (1989). Das Modell hat sich in einer Vielzahl von Experimenten weitgehend bestätigt und sich bei der Erklärung einer Reihe von Phänomenen wie interpersoneller Anziehung bewährt. Ein dem Gleichgewicht ähnlicher Begriff wurde von Newcomb (1953) mit der A-B-X-Theorie vorgestellt und auf die Kommunikation zwischen Menschen angewandt. Das Kongruenzprinzip von Osgood/Tannenbaum (1955) ist ebenfalls eine frühe Variante der Heider-Theorie, die auf die Richtung von Einstellungsänderungen abzielt.

3. Logische Struktur der Heider-Theorie

Wir halten uns bei der Darstellung der logischen Struktur der Theorie genau an die Version, wie sie Heider ursprünglich vorgelegt hat und wie sie eben zusammengefasst informell vorgestellt wurde. In der strukturalistischen Rekonstruktion wird die präsystematisch gegebene Theorie informell mengentheoretisch expliziert und axiomatisch dargestellt (für eine Einführung siehe Balzer/Moulines/Sneed, 1987 oder Manhart, 1995).

Wenn wir die Theorie in dieser Weise rekonstruieren, müssen wir zunächst überlegen, welche Begriffe und Relationen in der Theorie verwendet werden und von welchem Typ sie sind. In der Heider-Theorie sind die Grundbegriffe und Relationen besonders einfach, letztere aber - wie sich gezeigt hat - wenig präzise. Heider unterscheidet zwischen zwei Arten von Entitäten, die er als Grundbegriffe behandelt: Personen und Nicht-Personen. Nicht-Personen können dabei eine Vielzahl von Entitäten sein: Ereignisse, Objekte, Situationen etc. Zwischen diesen beiden Grundbegriffen herrschen zwei Typen von Relationen: die gefühlsgeladene L-Relation und die U-Relation. Auf die Problematik der L- und U-Relationen weisen Cartwright/Harary (1956) und Sukale (1971) hin.

Heider formalisiert seine Theorie nicht, sondern symbolisiert sie lediglich, d.h. er verwendet nur abkürzende Symbole zu deren Darstellung. Beispielsweise wird die Gefühlsrelation mit „L“ symbolisiert, wenn sie positiv ist und mit „~L“, wenn sie negativ ist. Analog wird die Einheitsrelation mit „U“ bzw. „~U“ symbolisiert.

Unklar bleibt, was das Symbol „~„ vor einem Relationssymbol aus der Relation macht: „Das Gegenteil der Relation (z.B. aus Liebe Hass) oder schlicht das Komplement (z.B. aus Liebe Nicht-Liebe, beziehungsweise: die Relation Liebe liegt nicht vor)? Diese irreführende Symbolisierung, die in der mangelhaften Definition des Negationszeichens bestand, machte sich sofort bei den ersten Experimenten bemerkbar, wo Unklarheiten über die Interpretation negierter Relationen auftauchten und zur willkürlichen Klassifikation von Daten führte“ (Sukale 1971: 49). Im allgemeinen scheint es so, als ob „~L“ in der Bedeutung von „Nichtmögen“ - die gegenteilige Relation - aufgefasst ist, während „~U“ als Komplement verwendet wird. In beiden Fällen wird jedoch das gleiche Symbol benutzt.

Diese Unklarheiten wären vermieden worden, wenn L- und U-Beziehungen mit Hilfe der Relationentheorie und allgemeiner Eigenschaften von Relationen beschrieben worden wären, wie es im strukturalistischen Theorienkonzept üblich ist. Da L- und U-Beziehungen im Hinblick auf Balance aber keine Rolle spielen, ersparen wir uns den Umweg über die Codierung in L-und U-Relationen und beziehen uns gleich auf positive und negative Relationen als Grundbeziehungen.

Die in der Heider-Theorie verwendeten Begriffe bestehen also zunächst aus einer dreielementigen Menge O sowie positiven und negativen Relationen. Die erste Definition führt diese Grundbegriffe ein und charakterisiert sie hinsichtlich ihrer mengentheoretischen Typen. Auf der Basis dieser Grundbegriffe können weitere Begriffe definiert werden: die Menge O zerfällt z.B. in die Klasse der Personen und Nicht-Personen und auf der Grundlage der Relationen kann die Menge aller Triaden und Triadenzustände definiert werden.

Die Klasse der Grundbegriffe und Relationen wird normalerweise im potentiellen Modell angegeben. Bevor wir dies tun, müssen wir uns noch einmal die zentrale Behauptung der Balancetheorie verdeutlichen. Das grundlegende inhaltliche Axiom von Heider ist, dass Ungleichgewichtszustände dazu ***neigen***, gleichgewichtig zu werden. Wenn wir „neigen“ dahingehend interpretieren, dass unbalancierte Strukturen ***über die Zeit*** hinweg in balancierte überführt werden, dann ist implizit in dieser Behauptung ein Zeitindex enthalten. Damit muss in die Modelldefinition ein Zeitindex eingeführt werden, so dass die Zeit und eine entsprechende Ordnungsrelation auf ihr auch als Grundbegriffe im potentiellen Modell benötigt werden. Andererseits können die oben genannten Grundmengen und Relationen unabhängig von der Zeit definiert werden. Wir führen die grundlegenden Mengen und Relationen deshalb zuerst ohne Bezugnahme auf die Zeit in einem Prädikat „...ist ein Heider-Graph“ ein und verwenden dieses Prädikat dann zusammen mit dem Zeitbegriff und der auf ihr definierten Ordnungsrelation im Definiens des potentiellen Modells.

Definition 1

x ist ein *Heider-Graph* ($x \in HG$) genau dann, wenn es O, P und N gibt, so dass gilt:

(1) $x = \langle O,P,N \rangle$

(2) O ist eine nicht-leere Menge der Kardinalität 3

(3) $P \subseteq O \times O$

(4) $N \subseteq O \times O$

(5) $P \cap N = \varnothing$

(6) Für alle $x \in O$: $\neg \langle x,x \rangle \in P \cup N$

(7) Wenn $x,y \in O$ dann $\langle x,y \rangle \in P \cup N$ oder $\langle y,x \rangle \in P \cup N$

Definition 1 legt den mengentheoretischen Typ der Grundbegriffe und Relationen fest. Die Axiome besagen, dass O eine dreielementige Menge ist (2) und P und N auf O definierte Relationen (3,4) sind, welche disjunkt (5) und irreflexiv sind (6). In Axiom (7) wird die von Heider implizit getroffene Annahme der Vollständigkeit der Struktur formuliert, d.h. dass zwischen allen Paarelementen aus O entweder eine P- oder N-Relation *existieren muss*. Die uninterpretierten Grundbegriffe sollen folgende intendierte Interpretation haben: O soll eine Menge von Personen und Nicht-Personen sein und P bzw. N die Menge der positiv bzw. negativ verknüpften Objektpaare. $\langle x,y \rangle \in P$ ist zu lesen als: zwischen x und y besteht eine positive Relation und entsprechend
$\langle x,y \rangle \in N$: zwischen x und y besteht eine negative Relation.

Auf der Basis dieser Grundbegriffe lassen sich nun verschiedene abgeleitete Beziehungen definieren, die zur weiteren Verwendung nützlich sind. Zunächst wird definiert, was allgemein eine Relation R ist und dann, was eine Nicht-Person (NP) und eine Person (PE) ist (vgl. hierzu auch Sukale 1971: 53).

Definition 2

Wenn $x = \langle O,P,N \rangle \in HG$ dann

(1) $R := P \cup N$

(2) $x \in NP$ genau dann wenn $x \in O$ und es existiert kein y, so dass $y \in O$ und xRy

(3) $x \in PE$ genau dann wenn $x \in O$ und $\neg\, x \in NP$

(4) $TR := R \times R \times R$

Definition 2-1 legt R als Vereinigung von P und N fest. Die Definitionen 2-2 und 2-3 bestimmen Personen und Nicht-Personen. Personen werden von Nicht-Personen rein formal dadurch unterschieden, dass alles, was kein Anfangspunkt einer Relation ist, eine Nicht-Person ist und Personen diejenigen Elemente, die keine Nicht-Person sind. Definition 2-4 legt die Triadenmenge TR als Menge von Tripeln, also als dreifaches Cartesisches Produkt über R fest.

Als nächstes müssen die balancierten, unbalancierten und indifferenten Triaden von Abb. 2 eingeführt werden. Diese werden einfach als Teilmengen I (Indefinit), G (Gleichgewicht) und U (Ungleichgewicht) von TR definiert.

Definition 3

Wenn $x = \langle O,P,N \rangle \in HG$ dann wird TR wie folgt in Teilmengen $G, U, I \subseteq TR$ zerlegt

(1) $I = \{\langle a,b,c \rangle \mid a,b,c \in N\}$

(2) $G = \{\langle a,b,c \rangle \mid a,b,c \in P \vee a \in P \wedge b,c \in N \vee b \in P \wedge a,c \in N \vee c \in P \wedge a,b \in N\}$

(3) $U = \{\langle a,b,c \rangle \mid a \in N \wedge b,c \in P \vee b \in N \wedge a,c \in P \vee c \in N \wedge a,b \in P\}$

Damit können wir nun das potentielle Modell definieren, in dem die Zeit T (genauer: T und die Kleiner-Relation als endliche, lineare Ordnung) als neuer Grundbegriff eingeführt wird und der in Definition 1 festgelegte Heider-Graph auf die Zeit bezogen wird.

Definition 4

x ist ein *potentielles Modell* der Heider-Theorie ($x \in M_p(HT)$) gdw es O, T <, P, N gibt, so dass gilt:

(1) $x = \langle O, T, <, P, N \rangle$

(2) O ist eine dreielementige Menge

(3) $\langle T,< \rangle$ ist eine endliche, lineare Ordnung

(4) $P: T \rightarrow Pot(O \times O)$ und $N: T \rightarrow Pot(O \times O)$

(5) Für alle $t \in T$: $\langle O, P(t),N(t) \rangle \in HG$

Axiom (4) besagt, dass P und N Funktionen sind, die jedem Zeitpunkt $t \in T$ genau ein Element aus der Potenzmenge, also der Menge aller Teilmengen $O \times O$, zuordnen. Axiom (5) fordert, dass das Tripel $\langle O, P(t),N(t) \rangle$ für alle betrachteten Zeitpunkte t ein Heider-Graph ist. In diesem Axiom ist die Forderung enthalten, dass alle Objekte über die verstrichenen Zeiteinheiten die gleichen bleiben müssen. Ansonsten wäre es möglich, zwei ganz verschiedene Grundmengen zu einem Balancesystem zu verbinden, die gar nichts miteinander zu tun haben.

Als nächste Struktur wäre das partielle potentielle Modell zu definieren. Damit ist zu klären, ob theoretische Begriffe in der Heider-Theorie (HT) vorkommen. Diese Frage ist leicht zu entscheiden: es gibt keine Begriffe, zu deren Messung man die

Gültigkeit der Heider-Theorie voraussetzen muss. Zwar sind P und N Einstellungen und die Bestimmung von Einstellungen ist wie jede Einstellungsmessung stark theoriegeleitet. Theoretische Begriffe im Sinn der Theorie der Einstellungsmessung gehen aber in HT als nichttheoretische Begriffe ein (Stephan 1990: 75). In der Heider-Theorie existieren somit keine HT-theoretischen Terme, und es besteht keine Notwendigkeit, zwischen partiellen potentiellen Modellen und potentiellen Modellen zu unterscheiden.

Mit Definition 5 ist die Festlegung der Grundbegriffe und abgeleiteten Begriffe abgeschlossen.

Definition 5
$x(t) := \langle O, P(t), N(t) \rangle$

Wir können nun das potentielle Modell durch Hinzufügen des eigentlichen Axioms zu einem Modell ergänzen.

Definition 6
x ist ein *Modell* von HT ($x \in M(HT)$) gdw es O, T, <, P und N gibt, so dass gilt:
(1) $x = \langle O,T,<,P,N \rangle$
(2) $x \in M_p(HT)$
(3) Für alle $t \in T$ und für alle a: Wenn $t < \max(T)$ und $a \in TR_{x(t)}$ und $a \in U_{x(t)}$, dann gibt es ein $t' \in T$ so dass gilt: $t < t'$ und $a \in G_{x(t')}$ und für alle $t'' > t'$: $a \in G_{x(t'')}$

Definition 6 repräsentiert mit Axiom (3) Heiders Kernaussage, nach der unbalancierte Triaden über eine Zeitperiode hinweg in balancierte überführt werden. Definition 6-3 bildet das eigentlich inhaltliche Axiom und „Fundamentalgesetz der Balancetheorie". U und G sind dabei die in Definition 3 eingeführten ungleichgewichtigen und gleichgewichtigen Triaden. In unserer Rekonstruktion besagt das Gesetz informell, dass für alle Zeitpunkte t und alle Triaden a gilt: Wenn t kleiner ist als das Maximum von T und a zu t unbalanciert ist, dann gibt es einen Zeitpunkt t' > t, bei dem a balanciert ist und für alle Zeitpunkte t'', die größer sind als t', bleibt a balanciert.

Die folgende Sequenz von Heider-Triaden über sechs Zeitpunkte würde danach die Modelldefinition erfüllen:

t_1	t_2	t_3	t_4	t_5	t_6
U	U	U	G	G	G

Das zweite Konjunktionsglied nach dem Existenzquantor (für alle t" > t': $a \in G_{x(t'')}$) ist in unserer Interpretation von Heider nötig. Ansonsten könnte nämlich nach dem erstmaligen Wechsel einer unbalancierten Triade in eine balancierte, diese zu einem späteren Zeitpunkt wieder unbalanciert werden. Das folgende Beispiel zeigt eine solche Situation:

t_1	t_2	t_3	t_4	t_5	t_6
U	U	U	G	U	U

Wir meinen, Heider so zu interpretieren, dass dies ausgeschlossen ist; andernfalls könnte einfach das zweite Konjunktionsglied weggelassen werden.

Das Fundamentalgesetz (6-3) ist ***bewusst vage gehalten***, dürfte aber die Intention Heiders genau ausdrücken. Zum einen wird nicht spezifiziert, *welche* Relationen geändert werden, zum andern wird die ***Zeitspanne***, in der diese Änderung erfolgen soll, *nicht festgelegt*. An diesem Axiom kann die „Liberalität“ des strukturalistischen Theorienkonzepts verdeutlicht werden. Obwohl das Axiom inhaltlich sehr vage ist, kann der Zusammenhang in strukturalistischer Deutung völlig exakt angegeben werden.

Damit ist der mathematische Strukturkern der Theorie vollständig festgelegt. Zu den bislang vorliegenden Strukturklassen M_p und M fehlt nun noch die Information, auf *was* diese Strukturen angewendet werden sollen. Da HT eine empirische Theorie sein soll, muss sie einen bestimmten Realitätsausschnitt erfassen. Zu den bereits eingeführten Strukturen kommt deshalb nun als dritte Strukturklasse die Menge der intendierten Anwendungen I hinzu.

Wie oben bereits erwähnt, liefern Heider und andere (Experimental-)Psychologen viele Beispiele für die Anwendung der Theorie, die man als triadische Einstellungssysteme bezeichnen kann. Die ursprünglich von Heider genannte Menge I_0 wurde im historischen Verlauf allerdings kaum vergrößert, vielmehr ergaben sich Hinweise, dass nicht alle Einstellungssysteme intendierte Anwendungen sind.

Berger et al (1962: 9-36) beschreiben folgende Situation (vgl. auch Opp 1984: 32): in bestimmten Gemeinschaften existiert eine Norm, dass Männer Alkohol trinken sollen, während Frauen diesen meiden sollten. Angenommen, ein Mann liebt ein Frau und er mag Alkohol, während seine Frau Alkohol ablehnt. Diese Triade ist unbalanciert und führt nach der Theorie zu Druck auf Veränderung und Gleichgewicht. Tatsächlich befindet sich diese Situation jedoch vollständig in Übereinstimmung mit existierenden Normen, und es besteht deshalb mit hoher Wahrscheinlichkeit kein Veränderungsdruck. An diesem Beispiel zeigen sich noch einmal konkret die unterschiedlichen Auffassungen von Kritischen Rationalisten und Strukturalisten. Kritische Rationalisten wie Opp interpretieren dies als Falsifikation der Heider-Theorie (Opp 1984: 32) bzw. als Aufforderung, den Theoriekern zu modifizieren. In Anbetracht vieler positiver experimenteller Befunde und erfolgreicher Anwendungen der Theorie würde dies jedoch eine völlig überzogene Reaktion darstellen, die jeglicher Intuition widerspricht, und die kein Naturwissenschaftler in einer analogen Situation machen würde.

Aus strukturalistischer Perspektive ist der Theoriekern infolge der Offenheit von I immun gegen Widerlegung. Die berichtete Anomalie ist ein Hinweis darauf, dass stark normbestimmte Situationen nicht zum Anwendungsbereich der Heider-Theorie gehören und ausgeschlossen werden sollten. Dies ist auch die Interpretation von Berger et al (1962: 13): „These exceptions are not so much a gap in Heider's theory as an indication of the way in which its scope is to be defined and limited“.[4] Heider selbst hat schon Balancesysteme aus dem Anwendungsbereich eliminiert, in denen Eifersucht, Neid und Wettbewerb eine Rolle spielen (Heider 1946: 110-111; 1977: 233). Eventuell müssen bei den intendierten Systemen noch weitere Einschränkungen gemacht werden, die mit einem vorsichtigen Gebrauch der Regel der Autodetermination ermittelt werden können. „Das sollte nicht ad hoc, sondern nach sorgfältigem Vergleich mit den bislang erfolgreichen Anwendungen geschehen, so dass ein systematischer Unterschied zwischen den erfolgreichen und den erfolglosen Anwendungen angegeben werden kann“ (Stephan 1990: 84).

Definition 7 ist ein *Versuch*, die Menge der intendierten Anwendungen von HT allgemein zu charakterisieren.

---

[4] Vgl. hingegen die Reaktion von Opp (1984: 32-33), der die Bereitschaft, Anomalien zu akzeptieren bzw. aus dem Anwendungsbereich auszuschließen, kritisiert.

Definition 7

x ist eine intendierte Anwendung von HT ($x \in I(HT)$) gdw es O, T, <, P und N gibt, so dass gilt:

(1) $x = \langle O,T,<,P,N \rangle$

(2) $x \in M_p(HT)$

(3) O, P und N sind kognitiv repräsentierte Mengen in einem menschlichen Individuum

(4) Zwischen je zwei $o_1,o_2 \in O$ bestehen (kognitiv repräsentierte) Relationen, die als positiv oder negativ klassifiziert werden können (Einstellungen)

(5) x referiert nicht auf Situationen, die bestimmt sind von Normen, Eifersucht, Neid, Wettbewerb ...

Intendierte Anwendungen der Heider-Theorie sind also alle Personen, deren kognitive Systeme sich in der Begrifflichkeit der Theorie beschreiben lassen und insbesondere die Axiome (3), (4) und (5) von Definition 7 erfüllen. Axiom (5) sollte - unter Anwendung der Autodeterminationsregel - näher charakterisiert werden.

Wichtig ist, dass die Menge der intendierten Anwendungen hier pragmatisch festgelegt ist und nicht etwa „formal definiert“ wird. Es sei anhand dieses konkreten Beispiels nochmals darauf hingewiesen, dass die Menge der intendierten Anwendungen aus prinzipiellen Gründen nicht formal festgelegt werden kann. Dies würde die Theorie zu einer mathematischen Theorie machen.

Die letzte Definition fasst schließlich alle Strukturelemente in einem Prädikat „... ist Heiders Balancetheorie“ zusammen.

Definition 8

$HT = \langle M, M_p, I \rangle$ ist *Heiders Balancetheorie* gdw

(1) M_p ist die Menge der potentiellen Modelle (Def. 4)

(2) M ist die Menge aller Modelle (Def. 6)

(3) I ist die Menge der intendierten Anwendungen (Def. 7)

Die drei Strukturelemente charakterisieren Heiders Balancetheorie vollständig: M_p definiert die von der Theorie verwendete Begrifflichkeit, M enthält das Fundamentalgesetz der Balancetheorie, und I legt die intendierten Anwendungen fest. Alle Personen mit Einstellungen, auf welche die Modelldefinition 6 angewendet werden kann, sind damit Modelle der Heider-Theorie.

4. Wissensbasierte Modellierung - Datengenerierungsmodell

Die Balancetheorie in der Fassung von Heider ist ein fruchtbares Ausgangsmodell, aber infolge seiner Einfachheit und der beschränkten Menge intendierter Anwendungen theoretisch nur bedingt interessant. Insbesondere bietet die Theorie auf dem Computer wenig. Die Datenbasis ist so klein und überschaubar, dass die einfachen Folgerungen sofort sichtbar sind.

Wir möchten das Computermodell aber hier benutzen, um die beiden unterschiedlichen Modellierungsarten - Datengenerierungs-Modell und strukturalistisches Modell - an diesem einfachen Beispiel zu erläutern. Für die folgenden Abschnitte sind elementare Kenntnisse der Logikprogrammiersprache PROLOG notwendig.

Die Datenbasis des Heider-Modells könnte aus Kognitionen in Form von einfachen Subjekt-Verb-Objekt-Sätzen bestehen, wobei das Verb eine L- oder U-Relation ausdrücken müsste und Subjekt bzw. Objekt Personen und Nichtpersonen sind. Um linguistische Satzanalysen zu vermeiden, muss gewährleistet sein, dass Kognitionen wie

```
(D1)  John is married to Berta. He prefers Democrates while Berta
      dislikes Democrates
```

oder

```
(D2)  John und Berta, beide miteinander verheiratet, sind
      entschiedene Rauchgegner
```

als PROLOG-aufbereitete Subjekt-Prädikat-Objekt-Sätze eingelesen werden:

```
(D1') john 'is married to' berta. john prefers democrates. berta
      dislikes democrates.
(D2') john 'is married to' berta. john hates smoking. berta hates
      smoking.
```

Die Voraussetzungen an die Datenbasis müssten durch entsprechende Prüfprädikate sichergestellt sein. Ein Klassifikationsprädikat könnte dann die Klassifikation der Verben in die beiden Grundrelationen P und N übernehmen. Hierzu kann ein einfaches Lexikon bereitgestellt werden, mit dem die wichtigsten Verben in einer Liste gesammelt und in „positiv“ oder „negativ“ klassifiziert werden.

```
lexikon(positiv,[likes, loves, positiv, 'is married to', ...]).
```

```
lexikon(negativ,[hates,dislikes,...]).
```

Ist das Verb nicht in der Liste enthalten, kann an den Benutzer bezüglich der Relationsbewertung eine Frage gestellt werden.

Den Relationen P und N entsprechen die beiden 2-stelligen PROLOG-Prädikate `positiv/2` und `negativ/2`, so dass aus (D1) die PROLOG-Fakten

```
positiv(john,berta).
positiv(john,democrates).
negativ(berta,democrates).
```

entstehen würden.

Damit kommen wir zu den interessanteren Prädikaten. Für die in Definition 3 festgelegten Triadenteilmengen G, U, I $\subseteq$ TR benutzen wir den Namen `triade/2` mit der Struktur an erster Argumentstelle und dem Triadenzustand G, U oder I an zweiter Stelle. Die Regeln nach Definition 3 könnten dann wie folgt repräsentiert werden:

```
triade((positiv(P,O), positiv(P,X), positiv(O,X)), gleichgewicht)
   :-
   positiv(P,O),
   positiv(P,X),
   positiv(O,X).

triade((positiv(P,O), negativ(P,X), negativ(O,X)), gleichgewicht)
   :-
   positiv(P,O),
   negativ(P,X),
   negativ(O,X).

triade((positiv(P,O), negativ(P,X), positiv(O,X)), ungleichgewicht)
   :-
   positiv(P,O),
   negativ(P,X),
   positiv(O,X).
...
```

Die erste Regel kann gelesen werden als: Die Struktur ... ist eine Triade im Gleichgewicht, wenn zwischen allen Objekten positive Relationen bestehen.

Wir können nun an die Wissensbasis bereits Fragen nach den Eigenschaften von Triaden stellen. Zur Verdeutlichung sind PROLOG-Fragen im weiteren Verlauf der Arbeit fett gedruckt.

Das Ergebnis einer Anfrage an die Datenbasis (D1) mit

```
?- triade(X,Y).
```

wäre dann

```
X = (positiv(john,berta), positiv(john,democrates),
    negativ(berta,democrates))
Y = ungleichgewicht
```

Ein Beweisversuch zur Herleitung einer gleichgewichtigen Triade in (D1) schlägt fehl:

```
?- triade(X,gleichgewicht).
no
```

Auf der Grundlage dieser Regeln lassen sich weitere Regeln bezüglich Stabilität und Instabilität von Triaden einführen.

```
triade((X,instabil) :-
   triade(X,ungleichgewicht).

triade(X,stabil) :-
   triade(X,gleichgewicht).
```

Die erste Regel besagt, dass eine Triade X instabil ist, wenn sie ungleichgewichtig ist. Die zweite Regel besagt, dass eine Triade X stabil ist, wenn sie im Gleichgewicht ist.

Schließlich lässt sich ein Prädikat `tendenz/2` definieren, das die Möglichkeiten der Entwicklung unbalancierter in balancierte Strukturen beinhaltet. `tendenz(X,Y)` ist zu lesen als: Die Struktur X hat eine Tendenz zu Y.

```
tendenz(X,X) :-
   triade(X,stabil).

tendenz(X,Y) :-
   triade(X,instabil),
   transformation(X,Y).
```

Das erste `tendenz`-Prädikat besagt, dass sich die Struktur tendenziell nicht verändert, wenn X eine stabile Triade ist. Das zweite Tendenz-Prädikat bedeutet, dass sich X tendenziell nach Y verändert, wenn X instabil ist und eine Transformation von X nach Y möglich ist.

Das Transformations-Prädikat überführt unbalancierte Strukturen in balancierte, wobei hier als Möglichkeit nur die Änderung der *Relationen* in Betracht gezogen wurde. Für jede Struktur gibt es drei Möglichkeiten, so dass mit

```
transformation(X,Y)
```

jedem X drei Y-Strukturen zugeordnet sind.

Die Transformations-Prädikate für Triade (e) von Abb. 2 lauten dann beispielsweise:

```
transformation((positiv(P,O),negativ(P,X),positiv(O,X)),
               (positiv(P,O),positiv(P,X),positiv(O,X))).
transformation((positiv(P,O),negativ(P,X),positiv(O,X)),
               (negativ(P,O),negativ(P,X),positiv(O,X))).
transformation((positiv(P,O),negativ(P,X),positiv(O,X)),
               (positiv(P,O),negativ(P,X),negativ(O,X))).
```

Mit (D2) als Datenbasis wird die Anfrage

```
?- tendenz(X,Y).
```

beantwortet mit

```
X = (positiv(john,berta), negativ(john,rauchen),
    negativ(berta,rauchen))
Y = (positiv(john,berta), negativ(john,rauchen),
    negativ(berta,rauchen)).
```

Dies bedeutet, die Relationen ändern sich nicht über die Zeit (da die Datenbasis schon balanciert ist). Mit der ungleichgewichtigen Triade (D1) als Datenbasis ergeben sich drei Alternativen für eine Relationenänderung:

```
X = (positiv(john,berta), positiv(john,democrates),
     negativ(berta,democrates))
Y = (positiv(john,berta), negativ(john,democrates),
     negativ(berta,democrates))
;
Y = (positiv(john,berta), positiv(john,democrates),
     positiv(berta,democrates))
;
Y = (negativ(john,berta), positiv(john,democrates),
     negativ(berta,democrates))
;
no
```

Die eben vorgestellte Implementierung ist ein Datengenerierungs-Modell, weil das Modell neue Daten erzeugt, die z.B. mit realen Systemen verglichen werden können. In der nächsten Version erzeugt das Modell keine neuen, empirisch direkt interpretierbaren Daten.

5. Wissensbasierte Modellierung – Strukturalistische PROLOG-Prädikate

In der strukturalistisch ausgerichteten Implementierung soll das Programm die „deklarative Rolle“ der Axiome übernehmen und Gültigkeitsüberprüfungen vornehmen. In diesem Fall repräsentieren die PROLOG-Regeln also programmiersprachlich die strukturalistischen Prädikate. Der Benutzer gibt die Beschreibung einer Struktur als Datenbasis vor und kann mit den PROLOG-Regeln feststellen, ob die Prädikate in der vorliegenden Datenbasis gelten oder nicht gelten.

Die Inputdaten werden hier also von einer Struktur gebildet, das Programm prüft, ob die Struktur die Axiome erfüllt und liefert als Outputdaten Wahrheitswerte. Im Gegensatz zur ersten Version werden bei dieser Variante keine neuen Modelldaten generiert. Man kann die so erzeugten Outputs wie „x ist ein potentielles Modell“ etc. allenfalls als Meta-Daten interpretieren.

Wenn wir uns bei der Implementierung an die eingeführten strukturalistischen Prädikate halten, müssen wir die Datenbasis etwas ändern, da potentielles Modell und Modell explizit auf die Zeit Bezug nehmen. Im potentiellen Modell und im Modell betrachten wir eine endliche Menge von Zeitpunkten und für jeden dieser Zeitpunkte die Relationen zwischen je zwei Elementen der Objektmenge.

Unsere vom Benutzer einzugebenden Grundprädikate bilden zunächst das Prädikat `zeit/1`, welches den betrachteten Zeitraum in einer Liste festlegt und das Prädikat `objekte/1`, in dem die Menge der betrachteten Objekte festgelegt wird.

```
zeit([1,4,5,8]).
objekte([john,democrates,berta]).
```

Der Benutzer muss als weitere Grundbegriffe die P- und N-Relationen definieren und gibt diese in Form einfacher Subjekt-Verb-Objekt-Sätze ein, zusammen mit dem Zeitpunkt, zu dem sie gelten.

Das Einlesen funktioniert im Prinzip nach dem oben genannten Verfahren, in dem ein Klassifikator für die Einordnung des Verbs in die Mengen P oder N sorgt. Die Relationen werden dann zusammen mit dem Zeitpunkt, zu dem sie bestehen, abgespeichert, so dass wir z.B. folgende Datenbasis erhalten:

```
positiv(1,john,berta).
positiv(1,john,democrates).
negativ(1,berta,democrates).
positiv(4,john,berta).
positiv(4,john,democrates).
negativ(4,berta,democrates).
positiv(5,john,berta).
positiv(5,john,democrates).
positiv(5,berta,democrates).
positiv(8,john,berta).
positiv(8,john,democrates).
positiv(8,berta,democrates).
```

Bei der Eingabe der Relationen können natürlich einerseits Zeitpunkte nicht in der Zeitliste enthalten sein und andererseits Objekte nicht in der Objektliste. Dies zu überprüfen ist u.a. Aufgabe der strukturalistischen Prädikate. Das Prädikat für den Heider-Graphen wird z.B. schon dann scheitern, wenn die Kardinalität der Objektliste ≠ 3 ist. Faktenbasen der eben angegebenen Art sind also nun Prüfkandidaten für die strukturalistischen Prädikate.

Wir versuchen uns im folgenden möglichst genau an die vorgegebenen strukturalistischen Prädikate zu halten unter Verzicht auf Effizienz und knappen Code in PROLOG. Das PROLOG-Prädikat `heider_graph/2` definiert den Heider-Graphen gemäß Definition 1, wobei wir gleich den Zeitindex hinzugenommen haben.

`heider_graph(T,X)` ist also zu lesen als: (das Tupel) `X` ist zu `T` ein Heider-Graph. `T` ist hierbei wegen PROLOG-Konvention eine Variable, die einem t ∈ T der obigen Definitionen entspricht. Der Menge T entspricht in den PROLOG-Prädikaten die Variable `T_liste` (Text hinter `%` ist Kommentar).

```
heider_graph(T,X) :-            % X ist Heider-Graph zu Zeit T, wenn
   X = [O,P,N],                 % X = [O,P,N] und
   objekte(O),                  % O ist die Menge der Objekte und
   not O=[],                    % O ist nicht leer und
   O = [O1,O2,O3],              % O hat die Kardinalität 3 und
   atom(O1),atom(O2),atom(O3),  % die Elemente von O sind Atome und
   zeit(T_liste),               % T_liste ist die Zeitmenge und
   member(T,T_liste),           % T ist Element von T_liste und
   findall([X1,X2],positiv(T,X1,X2),P), % P ist Menge pos. Rel. und
   findall([Y1,Y2],negativ(T,Y1,Y2),N), % N ist Menge neg. Rel. und
   kreuzprodukt(O,O,Kreuz),     % Kreuz ist Cartes. Prod O × O und
   teilmenge(P,Kreuz),          % P ⊆ Kreuz und
   teilmenge(N,Kreuz),          % N ⊆ Kreuz und
   durchschnitt(P,N,[]),        % P ∩ N = ∅ und
   vereinigung(P,N,Rel),        % P ∪ N = Rel und
   irreflexiv(O,Rel),           % Rel ist irreflexiv und
   vollstaendig(Kreuz,Rel),!.   % zwischen allen Paarelem exist Rel.
```

Das PROLOG-Prädikat heider_graph/2 entspricht ziemlich genau Definition 1. objekte(O) instantiiert die Objektmenge. Die nachfolgenden fünf Clausen (bis atom(O3)) überprüfen die Objektmenge O. Die Clause not O = [] ist eigentlich überflüssig, da sie von der nächsten Clause impliziert wird und wurde nur zur Verdeutlichung eingefügt. Die PROLOG-Built-in-Prädikate atom(O1),..., atom(O3) stellen sicher, dass die Elemente von O weder Listen noch Zahlen noch Variable sind (d.h. atom(X) scheitert, wenn X=[a,b], X=3 oder X=Y), sondern Atome wie z.B. a, hans, uni_muenchen.

Die Prädikate zeit(T_liste) und member(T,T_liste) sorgen dafür, dass der Heider-Graph nur für die vorgegebenen Zeitpunkte betrachtet wird. Das Built-in-Prädikat findall/3 sammelt alle Paare, zwischen denen eine positive bzw. negative Relation besteht in eine Liste P bzw. N, so dass P und N genau die in Axiom (3) $P \subseteq O \times O$ und (4) $N \subseteq O \times O$ festgelegten Teilmengen enthalten. Die vier Hilfsprädikate kreuzprodukt/3, teilmenge/2, durchschnitt/3 und vereinigung/3 definieren die entsprechenden mengensprachlichen Operationen. Deren Arbeitsweise findet sich in der PROLOG-Literatur.

Die Axiome (6) und (7) von Definition 1 beziehen sich auf die Vereinigung von P und N, welche zunächst mit dem Vereinigungsprädikat durchgeführt wird. Axiom (6) entspricht dem Hilfsprädikat irreflexiv/2, das erfolgreich ist, wenn die Relation $Rel = P \cup N$ irreflexiv ist:

```
irreflexiv([],_).
irreflexiv([X|R],Rel) :-
   not member([X,X],Rel),
   irreflexiv(R,Rel).
```

Die letzte Clause vollstaendig(Kreuz,Rel) verifiziert mit Hilfe der Kreuzproduktliste schließlich, dass zwischen allen Paarelementen [X,Y] - mit Ausnahme von X=Y - eine P- oder N-Relation besteht. Da das heider_graph-Prädikat für jeden Heider-Graphen nur einmal erfüllbar sein soll, ist ein Cut zur Verhinderung von Backtracking eingefügt.

heider_graph/2 kann nun benutzt werden, um für eine beliebige Datenbasis für einen beliebigen Zeitpunkt festzustellen, ob diese ein Heider-Graph ist. Hierzu versucht der PROLOG-Interpreter alle Teilziele des Prädikats heider_graph/2 zu beweisen. Bei instantiiertem Zeitpunkt T und uninstantiiertem zweiten Argument liefert das Prädikat im negativen Fall `no` - d.h. das Datenbeispiel ist kein Heider-Graph - im positiven Fall gibt es das Tupel `[O,P,N]` zu `T` ∈ `T_liste` aus.

Beispielsweise liefert

```
?- heider_graph(2,X).
no
```

weil der Zeitpunkt 2 nicht in der Zeitliste enthalten ist und die Clause

`member(T,T_liste)`

damit scheitert.

Für

```
?- heider_graph(4,X).
```

gelingt jedoch der Beweis mit folgender Ausgabe:

```
X=[[john,democrates,berta],
  [[john,berta],[john,democrates]],[[berta,democrates]]].
```

Die erste Teilliste ist hierbei die Objektmenge, und die zweite Teilliste ist die Liste der Relationen mit den positiven Relationen als erster und den negativen Relationen als zweiter Teilliste.

Werden die drei Fakten

```
positiv(9,john,berta).
positiv(9,a,b).
positiv(9,berta,john).
```

hinzugefügt, so scheitert `heider_graph(9,X)` weil `[a,b]` nicht Element des Kreuzprodukts aus O × O ist (genauer: die Clause `teilmenge(P,Kreuz)` scheitert).

Auf einen grundlegenden Unterschied zwischen den mengensprachlichen Definitionen und den PROLOG-Regeln müssen wir an dieser Stelle noch hinweisen. Während in der mengensprachlichen Definition aufgrund der logischen Äquivalenz („gdw“) sowohl der Schluss von links nach rechts als auch umgekehrt möglich ist, können wir in PROLOG nur von rechts nach links ableiten. Am Beispiel: PROLOG schließt aus der Existenz des Tupels X mit der Objektmenge O usw. auf einen Heider-Graphen; es ist aber in PROLOG nicht wie in der mengensprachlichen Definition möglich, vom Heider-Graphen auf die Existenz von X, O etc. zu schließen. Logische Äquivalenzen sind in PROLOG zumindest nicht innerhalb *einer* Regel darstellbar. Dieser Unterschied betrifft grundsätzlich alle folgenden Definitionen.

Die nächsten Prädikate entsprechen Definition 2:

```
r(T,R) :-                    % R ist Menge der Relat. z. Zeitp. T wenn
   heider_graph(T,[_,P,N]),% [_,P,N] Heider-Graph zu T ist und
   vereinigung(P,N,R).     % R = P ∪ N.

np(T,X) :-                   % X ist Nicht-Person z. Zeitpunkt T wenn
   heider_graph(T,[O,P,N]),% [O,P,N] Heider-Graph zu T ist und
   member(X,O),             % X Element von O ist und
   not positiv(T,X,_),      % X nicht Anfangsp.d.pos.Rel.zu T ist und
   not negativ(T,X,_).      % X nicht Anfangsp. d.neg. Rel. zu T ist.

pe(T,X) :-                   % X ist eine Person zum Zeitpunkt T wenn
   heider_graph(T,[O,_,_]), % [O,_,_] Heider-Graph zu T ist und
   np(T,Y),                  % Y Nicht-Person zu T ist und
   differenz(O,[Y],X_liste),% X_liste Diff.menge O und [Y] ist und
   member(X,X_liste).        % X Element aus X_liste ist.
```

Beispielsweise sind damit folgende Anfragen möglich:

Welche Personen sind zum Zeitpunkt 1 Element des Heider-Graphen?

```
?- pe(1,X).
X = john;
X = berta;
no
```

Alle Nichtpersonen zum Zeitpunkt 1:

```
?- np(1,X).
X = democrates;
no
```

Zu welchem Zeitpunkt sind John und a Personen?

```
?- pe(X,john).
X = 1;
X = 4
?- pe(X,a).
no
```

Definition 3 legt balancierte, unbalancierte und indefinite Triaden fest, die in acht Regeln definiert werden können. Wir geben hier nur zwei wieder.

```
triade(T,O,gleichgewicht) :-   % O ist zu T Triade im Gleichgew wenn
   heider_graph(T,[O,P,N]),    % [O,P,N] Heider-Graph zu T ist und
   member(A,O),                % A Element von O ist und
   pe(T,A),                    % A zu T eine Person ist und
   member(B,O),                % B Element von O ist und
   not B = A,                  % A und B verschieden sind und
   pe(T,B),                    % B zu T eine Person ist und
   member(C,O),                % C Element von O ist und
   np(T,C),                    % C zu T eine Nichtperson ist und
   member([A,B],P),            % [A,B] Elem der P-Relation ist und
   member([A,C],P),            % [A,C] Elem der P-Relation ist und
   member([B,C],P),!.          % [B,C] Element der P-Relation ist.

triade(T,O,ungleichgewicht) :-
   heider_graph(T,[O,P,N]),
   member(A,O),
   pe(T,A),
   member(B,O),
   not B = A,
   pe(T,B),
   member(C,O),
   np(T,C),
   member([A,B],N),
   member([A,C],P),
   member([B,C],P),!.
```

Das erste Teilziel prüft jeweils, ob zu `T` ein Heider-Graph vorliegt. Die nächsten sieben Prädikate sorgen dafür, dass die richtigen Elemente aus der Objektmenge `O` instantiiert werden: `A` und `B` müssen Personen sein, die verschieden sind und `C` muss eine Nicht-Person sein. Erst mit den letzten drei `member`-Prädikaten wird die Zuordnung zu Gleichgewichts-, Ungleichgewichts- und Indifferenzzuständen getroffen. Da jede Triade genau einer der acht Regeln zugeordnet werden kann, braucht im Erfolgsfall nicht noch eine andere Regel geprüft zu werden, so dass am Ende jeder Regel ein Cut zur Verhinderung von Backtracking eingefügt wird.

Die Anfrage nach dem Zustand der Triade zum Zeitpunkt 4 ist beispielweise:

```
?- triade(4,X,Zustand).
X       = [john,democrates,berta]
Zustand = ungleichgewicht
```

Bei der Regel für das potentielle Modell können wir uns wieder ziemlich eng an Definition 4 anlehnen. Allerdings müssen wir hier für einzelne Axiome wieder Hilfsprädikate einführen.

```
pot_modell(X) :-              % X ist ein potentielles Modell wenn
   X=[O,T,'<',Pt,Nt],         % X das Tupel [O,T,'<',Pt,Nt] ist und
   objekte(O),                % O die Objektmenge ist und
   O=[O1,O2,O3],              % O dreielementig ist und
   atom(O1),atom(O2),atom(O3),   % die Elemente von O Atome sind und
   zeit(T_liste),                % T_liste Menge d. Zeitp. ist und
   lineare_ordnung(T_liste,'<'), % (T_liste,'<') e. lin Ordn ist und
   kreuzprodukt(O,O,Kreuz),    % Kreuz Kreuzprod v. O und O ist und
   potenzmenge(Kreuz,Potmenge),% Potmenge die Potmen v.Kreuz ist und
   funktion(positiv,T_liste,Pt,Potmenge), % Pt: T_liste → Potm. und
   funktion(negativ,T_liste,Nt,Potmenge), % Nt: T_liste → Potm. und
   sind_alle_heider_graphen(T_liste,[O,Pt,Nt]). % [O,Pt,Nt] alle
                                                % Heider-Graph. zu
                                                % T_liste sind.
```

Zunächst wird wieder die Objektmenge abgeprüft. Das Prädikat `lineare_ordnung/2` kontrolliert, ob die Liste der Zeitpunkte sortiert nach der Kleiner-Relation vorliegt. Es ist so allgemein geschrieben, dass jede beliebige Ordnungsrelation einer numerischen Liste geprüft werden kann:

```
lineare_ordnung([X,Y], Rel) :-
   Z =.. [Rel,X,Y],
   call(Z).

lineare_ordnung([X,Y|R],Rel) :-
   Z =..[Rel,X,Y],
   call(Z),
   lineare_ordnung([Y|R],Rel).
```

Die Relation - z.B. Kleiner, Größer, Kleiner/Gleich etc. - wird an zweiter Argumentstelle übergeben. Der Univ-Funktor `=..` macht aus der mit der Relation und beiden Zahlen konstruierten Liste `[Rel,X,Y]` ein Prädikat: Beispielweise wird so aus `[<,3,5]` das Prädikat `<(3,5)`. Bei einem Ziel, in dem der Funktor eine instantiierte Variable ist, muss das Ziel anschließend mit `call/1` aufgerufen werden. Das Prädikat `lineare_ordnung/2` ruft sich rekursiv so lange auf, bis die Ordnungsrelation entweder nicht erfüllt ist - in diesem Fall scheitert das Prädikat - oder die abgearbeitete Liste nur mehr aus zwei Elementen besteht, die der Ordnungsrelation genügen - in diesem Fall ist das Prädikat erfolgreich.

`kreuzprodukt/3` und `potenzmenge/2` sind wieder die entsprechenden mengensprachlichen Prädikate. Das nächste Prädikat `funktion/4` wird mit `positiv` oder `negativ` an erster Argumentstelle aufgerufen und berechnet für jeden Zeitpunkt die zugeordneten positiven und negativen Relationen.

```
funktion(Rel,[],[],_).
funktion(Rel,[T|T_rest],[(T,L)|R1], Potmenge) :-
   Z =.. [Rel,T,X1,X2],
   findall([X1,X2],Z,L),
   teilmenge(L,Potmenge),
   funktion(Rel,T_rest,R1,Potmenge).
```

Die Listen `Pt` und `Nt` enthalten nach Abarbeitung der `funktion`-Prädikate Mengen von Paaren `(T,L)`, wobei jedem Zeitpunkt `T` (erstes Element) die Menge der zu `T` bestehenden Relationen `L` zugeordnet ist (zweites Element).

```
Pt: [(1,[[john,berta], [john,democrates]]), (4,[[john,berta], ...]
Nt: [(1,[[berta,democrates]]), (4,[[berta,democrates]],...]
```

Schließlich fordert das zentrale fünfte Axiom von Definition 5, dass für alle betrachteten Zeitpunkte ein Heider-Graph vorliegen muss. Auch hier müssen wir ein Hilfsprädikat definieren, das die Zeitpunkte rekursiv abarbeitet und für jeden Zeitpunkt verifiziert, dass ein Heider-Graph gegeben ist:

```
sind_alle_heider_graphen([],[_,[],[]]).
sind_alle_heider_graphen([T|T_rest],[O,[(T,P)|P_rest],
                         [(T,N)|N_rest]]) :-
   heider_graph(T,[O,P,N]),!,
   sind_alle_heider_graphen(T_rest,[O,P_rest,N_rest]).
```

`sind_alle_heider_graphen/2` arbeitet die Zeitliste rekursiv zusammen mit `Pt` und `Nt` ab und verifiziert für jeden Zeitpunkt, dass ein Heider-Graph vorliegt. Das Prädikat ist nur erfolgreich, wenn für *alle* Zeitpunkte der vorgegebenen Menge Heider-Graphen existieren.

In unserem Beispiel liefert

```
?- pot_modell(X).
```

die folgende Struktur und damit den Beweis, dass die Datenbasis ein potentielles Heider-Modell ist:

```
X  = [
       [john,democrates,berta],              % Objektliste
       [1,4,5,8,9],                          % Zeitpunkte
       '<',                                  % Ordnungsrelation
       [
         (1,[[john,berta],[john,democrates]]), % P-Relation zu T=1
         (4,[[john,berta],[john,democrates]]), %            zu T=4
         (5,[[john,berta],[john,democrates],[berta,democrates]]),
         (8,[[john,berta],[john,democrates],[berta,democrates]])
       ],
       [
         (1,[[berta,democrates]]),             % N-Relation zu T=1
         (4,[[berta,democrates]]),             %            zu T=4
         (5,[]),                               %            zu T=5
         (8,[])                                %            zu T=8
       ]
     ]

yes
```

Schließlich lässt sich in Anlehnung an Definition 6 das PROLOG-Modell-Prädikat einführen, das in PROLOG etwas umgeschrieben werden muss.

```
modell(X) :-
   X=[O,T_liste,_,_,_],
   pot_modell(X),!,
   axiom_3_test(T_liste, T_liste,O).


axiom_3_test([],_,_).

axiom_3_test([T1|T_rest],T_liste,O) :-
   triade(T1,O,ungleichgewicht),
   member(T2,T_liste),
   T1 < T2,
   triade(T2,O,gleichgewicht),!,
   axiom_3_test(T_rest,T_liste,O).

axiom_3_test([T1|T_rest],T_liste,O) :-
   triade(T1,O,gleichgewicht),
   member(T2,T_liste),
   T1 >= T2,
   triade(T2,O,ungleichgewicht),!,
   axiom_3_test(T_rest,T_liste,O).
```

Zunächst verifiziert `modell(X)`, dass `X` ein potentielles Modell ist. Für den Test von Axiom (3) wird ein eigenes Prädikat eingeführt, das alle Elemente der Zeitliste abarbeitet (wegen Allquantor über T). Das Prädikat ist erfüllt, wenn es für jeden Zeitpunkt `T1`, zu dem eine ungleichgewichtige Triade existiert, einen Zeitpunkt `T2` > `T1` gibt, zu dem die Triade gleichgewichtig ist.

Man beachte, dass `T2` automatisch durch PROLOG-Backtracking gesucht und bestimmt wird. Ist die Triade zu `T1` im Gleichgewicht, so ist die Bedingung von Axiom 3 nicht erfüllt. Da wir aber die Zeitliste vollständig abprüfen, müssen wir auch diesen Fall noch berücksichtigen, der der dritten Clause von `axiom_3_test` entspricht. Sie besagt folgendes: Ist eine Triade zu `T1` im Gleichgewicht, dann gab es einen Zeitpunkt `T2` ≤ `T1`, zu dem sie ungleichgewichtig war.

In unserem Beispiel liefert das Modell-Prädikat den Beweis, dass mit der Datenbasis ein Heider-Modell vorliegt, und es wird - bei Aufruf mit Variable - das Tupel wie oben ausgegeben. Scheitert das Modell-Prädikat nach Beweis von `pot_modell(X)`, so liegt zwar ein potentielles Heider-Modell vor, aber das grundlegende inhaltliche Axiom gilt nicht in der Datenbasis. Wären alle Relationen in der kleinen Datenbasis über alle Zeitpunkte z.B. immer positiv, ist `pot_modell(X)` erfüllt, aber `modell(X)` scheitert.

Der Theoriekern der strukturalistisch interpretierten Heider-Theorie liegt damit als PROLOG-Programm vor. Die Prädikate könnten noch etwas ausgebaut und informativer aufbereitet werden, die Ausgaben leserlicher gestaltet und eine Erklärungskomponente hinzugefügt werden. Die Erklärungskomponente sollte insbesondere Rechtfertigungen darüber liefern, warum ein Prädikat zutrifft bzw. nicht zutrifft, also z.B. warum eine Faktenbasis ein potentielles Modell, aber kein Modell ist. In Anbetracht der Einfachheit des Modells erscheint dies aber nicht sehr lohnenswert.

Literatur

Balzer, W./Moulines, C.U./Sneed, J.D. (1987). *An Architectonic for Science. The Structuralist Program*. Dordrecht: Reidel.

Berger, J./Cohen, B.P./Snell, J.L./Zelditch, M. (1962). *Types of Formalization in Small-Group Research*. Boston.

Cartwright, D./Harary, F. (1956). Structural Balance: A Generalisation of Heider's Theory. *Psychological Review*, 63, S. 277-293 (wieder abgedruckt in: S.Leinhardt (ed.) 1977, S.9-25).

Deutsch, M./Krauss, R.M. (1976). *Theorien der Sozialpsychologie*. Frankfurt a.M.: Fachbuchhandlung Psychologie.

Festinger, L. (1978). *Theorie der kognitiven Dissonanz*. Bern: Huber (zuerst 1957: A Theory of Cognitive Dissonance. Evanstone, Ill.: Row Peterson).

Heider, F. (1946). Attitudes and Cognitive Organization. *Journal of Psychology*, 21, S.107-112 (wieder abgedruckt in: S.Leinhardt (ed.) 1977, S.3-8).

Heider, F. (1977). *Psychologie der interpersonalen Beziehungen*. Stuttgart: Klett (zuerst 1958: The Psychology of Interpersonal Relations. New York: Wiley).

Heider, F. (1979). On Balance and Attribution. In P.W.Holland/S.Leinhardt, S.11-23.

Holland, P.W./Leinhardt, S. (eds.) (1979). *Perspectives on Social Network Research*. New York: Academic Press.

Irle, M. (1975). *Lehrbuch der Sozialpsychologie*. Göttingen: Hogrefe.

Jordan, N. (1953). Behavioral Forces that are a Function of Attitude and Cognitive Organization. *Human Relations*, 6, S.273-287.

Landy, D./Aronson, E. (1969). The Influence of the Character of the Criminal and his Victim on the Decisions of Simulated Jurors. *Journal of Experimental Social Psychology*, 5, S.141-152.

Lerner, M.J./Simmons, C.H. (1966). Observer's Reaction to the "Innocent Victim": Compasion of Rejection? *Journal of Personality and Social Psychology*, 4, S.203-210.

Manhartt, K. (1995) KI-Modelle in den Sozialwissenschaften, München: Oldenbourg

Methodology. *Journal of Theoretical Biology*, 113, S.311-330.

Newcomb, T.M. (1953). An Approach to the Study of Communicative Acts. *Psychological Review*, 60, S.393-404.

Opp, K.D. (1984). Balance Theory: Progress and Stagnation of a Social Psychological Theory. *Philosophy of the Sociological Sciences*, 14, 1, S.27-49.

Osgood, C.E./Tannenbaum, P.H. (1955). The Principle of Congruity in the Prediction of Attitude Change. *Psychological Review*, 62, S.42-55.

Stahlberg, D./Frey, D. (1987). Konsistenztheorien. In D.Frey/S.Greif (eds.), *Sozialpsychologie* (2.Aufl.). München: Psychologie Verlags Union, S.214-221.

Stephan, E. (1990). *Zur logischen Struktur psychologischer Theorien*, Berlin: Springer.

Sukale, M. (1971). Zur Axiomatisierung der Balancetheorie. Eine wissenschaftstheoretische Fallstudie. *Zeitschrift für Sozialpsychologie,* 2, S.40-57.

West, S.G./Wicklund, R.D. (1985). *Einführung in sozialpsychologisches Denken*. Weinheim: Beltz.

Witte, E.H. (1989). *Sozialpsychologie. Ein Lehrbuch*. München: Psychologie Verlags Union.