

### Prognosen als Methode zur Reduktion der Offensichtlichkeit von Umfragebefunden: ein Pilotexperiment

Borg, Ingwer; Hillenbrand, Carola

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Borg, I., & Hillenbrand, C. (2003). Prognosen als Methode zur Reduktion der Offensichtlichkeit von Umfragebefunden: ein Pilotexperiment. *ZUMA Nachrichten*, 27(52), 7-19. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-207778>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

# PROGNOSEN ALS METHODE ZUR REDUKTION DER OFFENSICHTLICHKEIT VON UMFRAGEBEFUNDEN: EIN PILOTEXPERIMENT

INGWER BORG & CAROLA HILLENBRAND<sup>1</sup>

Untersucht wird hier eine Methode, die den Eindruck der Offensichtlichkeit von Umfragebefunden reduzieren soll. Versuchspersonen (Vpn) wurden gebeten, die Ergebnisse für verschiedene Items aus angenommenen Umfragen vorherzusagen bzw. sie bekamen die Vorhersagen anderer Personen gezeigt. Es zeigte sich, dass diese Maßnahmen dazu führten, dass die Vpn Umfragebefunde als überraschender und als interessanter bewerten als Personen der Kontrollgruppe. Keinen systematischen Einfluss haben die Prognosen auf die Bewertung der praktischen Nützlichkeit der Befunde.

This paper investigates whether prognoses on the results of surveys help to reduce the common impression of obviousness when encountering survey findings. An experiment was conducted where subjects were asked to predict the results of a number of fictitious survey items and/or they were shown predictions made by other people on these items. It was found that making and/or seeing such prognoses indeed makes survey findings more unexpected and more interesting. On the other hand, prognoses have no systematic effect on the subjects' ratings of the practical usefulness of the findings.

## 1. Einleitung

Den Sozialwissenschaften wird häufig unterstellt, dass das, was sie mit ihrer Forschung herausfinden, offensichtlich ist. Das ist insbesondere für die Umfrageforschung ein Problem, weil ihre Ergebnisse unter dem Eindruck des „Habe ich sowieso schon gewusst“, des

---

<sup>1</sup> Carola Hillenbrand ist Research Fellow am Henley Management College Greenlands, Henley-on-Thames.

„Das hätte ich Ihnen alles vorher schon sagen können“ kaum noch die nötige kritische Aufmerksamkeit bekommen.

Die scheinbare Offensichtlichkeit sozialwissenschaftlicher Befunde hängt eng zusammen mit dem Hindsight Bias [auch: I-knew-it-all-along-Effekt (Stahlberg u.a. 1993) oder Rückschaufehler (Pohl 1992)]. Hierzu gibt es zahlreiche Studien. So berichtet Baratz (1983) z.B. ein Experiment mit Studenten, denen sie vermeintliche Forschungsbefunde (z.B. „Personen, die regelmäßig zur Kirche gehen haben statistisch mehr Kinder als Personen, die selten zur Kirche gehen“) bzw. deren Gegenteil („Personen, die ... haben statistisch weniger ...“) präsentierte. Die Vpn gaben jeweils an, dass der Befund sie nicht überrascht, sondern dem entspricht, was sie vorausgesagt hätten. Das Gefühl, schon alles gewusst zu haben, geht einher mit der Rückschautäuschung. Es tritt interessanterweise verstärkt auf bei Experten: „Oft zeigt sich Expertentum mehr in übersteigerter Selbstsicherheit („overconfidence“) als in einer objektiv ansteigenden Genauigkeit und Leistung bei Entscheidungen und Prognosen“ (Fiedler 1993: 7).

Die Rückschautäuschung und die damit einhergehende Overconfidence haben zu zahlreichen Erklärungen geführt. Die älteste Theorie ist die von Fischhoff (1983), der das Phänomen als Folge einer sofortigen Assimilation neuer Informationen in vorhandene Wissensstrukturen, als Echtzeit-Updating der Wissensbasis durch neue Informationen interpretiert. Demnach würde die Rückschau auf diesen bereits veränderten Strukturen basieren. Stahlberg u.a. (1995) argumentieren dagegen, dass Personen, die sich nicht an ihr ursprüngliches Urteil erinnern können (und nur bei diesen soll der Hindsight Bias auftreten), die nachträgliche Information als Anker für die Rekonstruktion des ursprünglichen Urteils verwenden. Campell/Tesser (1983) verweisen zudem auf motivationale Faktoren (Selbstwertgefühl, soziale Erwünschtheit, Bedürfnis nach Kontrolle), denen zumindest die öffentliche Darstellung des Urteils unterliegen könnte.

Das „Problem der Offensichtlichkeit“ (Lazarsfeld 1949) ist in der Literatur wiederholt bemerkt worden. In einem genialen Review des monumentalen „The American Soldier“ hat Lazarsfeld (1949) es wie folgt genutzt. Er berichtet zunächst einige Befunde dieser Studie, so z.B.:

- Soldaten mit besserer Bildung hatten größere Anpassungsprobleme im Krieg als Soldaten mit geringer Bildung.
- Soldaten aus den Südstaaten kamen mit dem heißen Südseeklima besser zurecht als Nordstaatler.
- So lange der Krieg andauerte waren die Soldaten motivierter nach Hause zurückzukehren als nach dem Krieg.

Alle Befunde erscheinen offensichtlich und leicht erklärbar. Nachdem der Leser diese abgenickt hat, klärt ihn Lazarsfeld über die *wirklichen* Befunde auf: Soldaten mit besserer Bildung hatten geringere Anpassungsprobleme als weniger Gebildete; Südstaatler kamen mit dem Südpazifiklima nicht besser zurecht als Nordstaatler; und die Rückkehrmotivation war nach dem Krieg größer als während des Krieges. Lazarsfeld merkt dazu an, dass wenn er diese Befunde berichtet hätte, der Leser dann eben diese als offensichtlich bezeichnet hätte.

Den Trick, den Rezipienten des Forschungsergebnisses auf eine Aussage festzulegen, bevor er das tatsächliche Ergebnis sieht, kann man als Versuch zum Debiasing, zur Enttäuschung bezeichnen. Lazarsfeld hat damit ein Forschungsgebiet antizipiert, das fragt, wie man kognitive Täuschungen verringern oder eliminieren kann. Fischhoff (1977, 1983) hat sich hiermit als erster beschäftigt, wenn auch nur mit der Absicht, auf diese Weise die Robustheit des Hindsight Bias zu testen. Von ihm und anderen wurden verschiedene Debiasing-Methoden vorgeschlagen und getestet, wie z.B. das Zurücknehmen von berichteten Ergebnissen als „falsch“ (Hasher u.a. 1981); die Präsentation von „sehr überraschenden Feedback-Informationen“ (Kahneman 1992); Manipulationen der Glaubwürdigkeit der Feedbackinformation (Pohl 1998); Manipulationen der Kosten suboptimaler Urteile (Arkes 1991) bzw. der Incentives für optimale Urteile (Hell u.a., 1988); oder eine vollständige Aufklärung über den Hindsight Bias (Fischhoff 1977; Pohl/Hell 1996).

Alle diese Techniken dienen dazu, die Genauigkeit der rekonstruierten Urteile zu verbessern. Für die Reduktion der scheinbaren Offensichtlichkeit von Umfragebefunden in der Umfragepraxis erscheinen sie wenig geeignet. Folgt man Fischhoffs (1983) Theorie, dann müsste man verhindern, dass diese Befunde sofort und unbemerkt in die Wissensbasis der Zielpersonen eingehen. Dafür ließe sich evtl. die „Consider the Opposite“-Technik anwenden, in der die Zielperson einen Hinweis zu einer eher „bedächtigen“ Rezeption der Befunden bekommt, der wie folgt lauten könnte: „Ask yourself at each step whether you would have made the same evaluation had exactly the same study produced results on the other side of the issue“ (Lord u.a., 1984: 1233). Ebenso könnte man sie auch auffordern, für jeden Befund möglichst viele und gute Argumente zu suchen, die apriorisch ein anderes Ergebnis nahe legen („counterfactual reasoning“). Diese Strategien reduzieren i. Allg. erfolgreich den Hindsight Bias.

Die Technik von Lazarsfeld erscheint von allen Ansätzen am besten geeignet für den praktischen Einsatz. Sie lässt sich so verwenden, dass man für einen Teil der Items schriftlich Prognosen erhebt und diese dann später mit den tatsächlichen Ergebnissen vergleicht. Borg (2002, 2003) hat dieses Verfahren im Rahmen von Mitarbeiterbefragungen verwendet. Dort wurden Führungskräfte gebeten, vorherzusagen,

zu welchen Prozentsatz ihre Mitarbeiter einigen ausgewählten Items jeweils zustimmen. Die mittleren Prognosen wurden später mit den empirischen Zustimmungsteilen verglichen. Eine Variante dabei war, auch den Range der Prognosen oder sogar ihre gesamte Verteilung zu zeigen. Dieses Feedback führte dazu, dass erhebliche Diskussionen entstanden, weil einige der Prognosen immer ganz falsch waren und weil der Range der Prognosen immer beträchtlich war. Die Führungskräfte lernten so schnell, dass sie zumindest uneins waren in der Wahrnehmung der Einstellungen und Meinungen der Mitarbeiter. Zudem zeigten sich einige inhaltliche Auffälligkeiten. So sagten die Führungskräfte z.B. die Zufriedenheit der Mitarbeiter bei den Themen, mit denen sie selbst relativ zufrieden waren (z.B. Bezahlung), viel zu hoch voraus. Derartige Befunde führten zu zusätzlichen Diskussionen und zu einer scheinbar offeneren und interessierteren Einstellung den Umfrageergebnissen gegenüber.

Diese Beobachtungen aus der Praxis sollen im folgenden experimentell überprüft werden. Wir fragen also, ob die Abgabe von Prognosen und verschiedene Formen der Rückmeldung von Prognosen tatsächlich dazu führen, dass Vpn Umfragebefunde als interessanter und informativer beurteilen. Wir fragen weiter, ob auch die Prognosen von anderen Personen zu ähnlichen Effekten führen oder ob es erforderlich ist, dass man selbst Prognosen abgibt. Zudem wollen wir wissen, ob verschiedene Formen der Rückmeldung – mittlere Prognosewerte, Range der Prognosen – zu unterschiedlichen Effekten führen.

## **2. Methode**

### **2.1 Versuchspersonen**

Als Versuchspersonen (Vpn) dienten 125 Personen aus dem Raum Augsburg. Unter den Vpn waren keine Studenten. Das Alter der Vpn lag zwischen 17 und 65 Jahre, mit einem Mittel von 35. Die Verteilung der Geschlechter war 65 Frauen und 60 Männer. Die Stichprobe war eine Gelegenheitsstichprobe, die über Aushänge und Handzettel rekrutiert wurde. Als Incentive wurden den Vpn die Teilnahme an einem bayrischen Weißwurstfrühstück bzw. an Kaffee und Kuchen angeboten. Die Vpn wurden per Zufall einer von fünf Gruppen (siehe unten) zugewiesen.

### **2.2 Unabhängige Variablen**

Vor den Treatments wurden zwei Fragen gestellt: „Für wie sinnvoll halten Sie psychologische Forschungsarbeit?“ (mit einer Ratingskala von 1=„nicht sinnvoll“ bis 6=„sehr sinnvoll“) und „Halten Sie es für schwierig, die Ergebnisse sozialwissenschaftlicher Umfragen vorherzusagen?“ (mit einer Ratingskala von 1=„sehr einfach“ bis 6=„sehr schwierig“).

Das Treatment wurde eingeleitet mit der Frage: „Auf den folgen Seiten werden Ihnen zwölf interessante Fragen aus aktuellen Umfragen vorgestellt.“ (Diese Fragen waren in Pretests überprüft worden in Bezug auf ihre Interessanztheit und Realitätsnähe.) Die Fragen lauteten z.B.: „In einer repräsentativen Umfrage wurden ADAC-Mitglieder befragt, ob sie die Einführung einer Autobahngebühr befürworten“ und „In einer repräsentativen Umfrage wurden deutsche Steuerzahler gefragt, ob sie bei der Einkommenssteuererklärung schummeln.“

Die Vpn hatten dann, je nach Versuchsgruppe, zunächst folgende Aufgabe. (a) Ohne Prognosen: „Zu Beginn sehen Sie die Fragestellung der jeweiligen Untersuchung. Daran schließt sich das Ergebnis und die graphische Darstellung des Ergebnisses an. Bitte lesen Sie sich die Ergebnisse aufmerksam durch!“ Oder (b) mit Prognosen: „Wir wollen von Ihnen wissen, welches statistische Ergebnis Sie für jede dieser Fragen erwarten. Hierbei sollen Sie angeben, wie viel Prozent der in der repräsentativen Umfrage befragten Personen den jeweiligen Aussagen zugestimmt haben.“ Diese Urteile wurden für Bedingung (b) sequentiell Item für Item erhoben.

Hieran anschließend bekamen die Vpn entweder keine Rückmeldung über evtl. Prognosewerte oder eines der beiden Treatments. (1) „Vor Ihnen hat bereits eine ähnliche Gruppe von Personen Vorhersagen abgegeben, zu welchen Ergebnissen die Umfragen ihrer Meinung nach führen. ... Im Folgenden zeigen wir Ihnen die durchschnittlich vorhergesagten Werte dieser Vergleichsgruppe. Im Anschluss daran stehen die tatsächlichen Umfrageergebnisse und eine graphische Darstellung der Ergebnisse“. Oder (2): „Vor Ihnen hat bereits... Dabei kam es teilweise zu recht unterschiedlichen Vorhersagen, Im Folgenden zeigen wir Ihnen zu jeder Fragestellung den kleinsten und den größten Wert, der von den Personen der Vergleichsgruppe vorhergesagt wurde...“.

Das Design weist also fünf Experimentalgruppen mit jeweils 25 Vpn auf:

| Gruppe | Prognosen | Rückmeldung von Prognosen     |
|--------|-----------|-------------------------------|
| 1      | keine     | keine                         |
| 2      | eigene    | keine                         |
| 3      | keine     | mittlere Prognosen            |
| 4      | eigene    | mittlere Prognosen            |
| 5      | eigene    | Range der Prognosen (Min-Max) |

### 2.3 Abhängige Variablen

Die Vpn sollten danach die Umfragebefunde bewerten auf den folgenden sechs-stufigen, bipolaren Skalen: „spannend vs. langweilig“, „nutzlos vs. hilfreich“, „offensichtlich (war ja klar) vs. latent (hätte ich nicht vermutet)“, „unsinnig vs. sinnvoll“, „altbekannt vs. neuartig“, „interessant vs. uninteressant“, „informativ vs. irrelevant“, „abgehoben vs. bodenständig“, „unauffällig vs. beachtenswert“, „lebensnah/praktisch vs. theoretisch“ und „überraschend vs. erwartet“.

Danach wurden die Vpn gebeten, folgende Fragen zu beantworten: „Was schätzen Sie, in wie viel Prozent der vorgestellten Umfrageergebnisse hätten/haben Sie in etwa eine richtige Prognose abgegeben?“; „Für wie sinnvoll halten Sie psychologische Forschungsarbeit?“; „Halten Sie es für schwierig, die Ergebnisse sozialwissenschaftlicher Umfragen vorherzusagen?“ (Die beiden letzten Fragen wiederholen die Fragen von oben.)

In einer Unterbrechungsphase sollten die Vpn die Grandiositätsskala ausfüllen (siehe unten). Schließlich wurde ihnen noch folgende Aufgabe unterbreitet: „Abschließend notieren Sie bitte kurz, welche von den dargestellten Umfragethemen Ihnen noch in Erinnerung sind. Geben Sie dabei auch die tatsächlichen (nicht die von Ihnen vorhergesagten) Umfrageergebnisse an, soweit Sie sich daran erinnern können.“

### 2.4 Moderatorvariablen

Eine naheliegende Vermutung ist die, dass der Eindruck der Offensichtlichkeit auch von bestimmten persönlichkeitspsychologischen Dispositionen abhängt, insbesondere von Narzissmus und Grandiosität. Um dies zu prüfen, wurde die von Meyer (1999, 2000) beschriebene Grandiositätsskala verwendet, ein kurzes 11-Item-Instrument. Sie enthält Items wie „Ich würde mich als außergewöhnlichen Menschen beschreiben“, „Ich habe eine Menge Qualitäten, die mich von anderen Menschen abheben“ und „Ich spüre häufig die Bewunderung anderer Menschen“. Dazu schreibt Meyer (1999: 8): „Die Items entsprechen damit konzeptuell der Facette ‚Grandiosität‘ der narzisstischen Persönlichkeitsstörung im DSM-IV. Die Skala kovariiert substantiell positiv mit Bewunderung, narzisstisch sozialen Kognitionen und Exhibit-Eitelkeit und ist unabhängig von der Skala Sensitivität. Ein starker positiver Zusammenhang besteht zur Extraversion ( $r=.42$ ) und der Rosenberg Selbstwertskala ( $r=.37$ ). Die Skala ist negativ korreliert mit Impression-Management“.

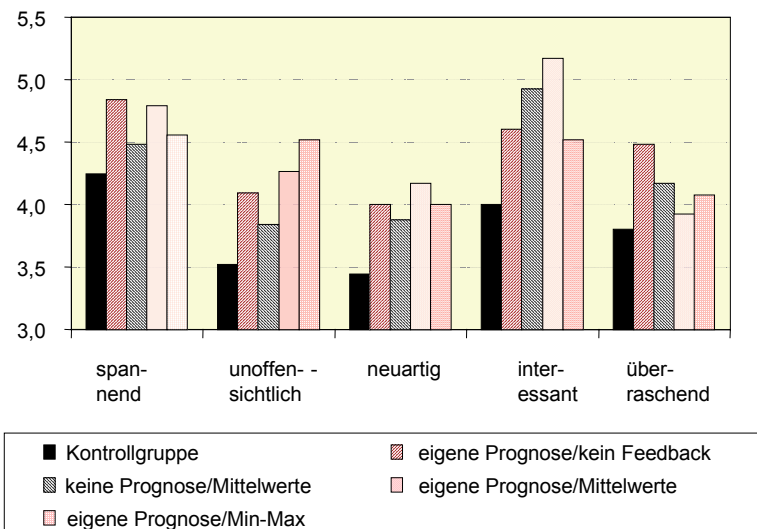
## 3. Ergebnisse

Eine MDS-Analyse (Borg/Groenen 1997) zeigt, dass die 125 Vpn die Beurteilungskriterien in zwei Gruppen strukturieren. Diese umfassen zum einen die Kriterien Latent (i.S. von „hätte ich nicht vermutet“), Interessant, Überraschend, Spannend und Neuartig zum

anderen die Skalen Sinnvoll, Praktisch, Hilfreich, Informativ-Relevant und Bodenständig. Das Attribut Beachtenswert hat keine klare Zuordnung und liegt in der MDS-Ebene zwischen den Clustern. Die beiden Eigenschaftscluster kann man als Unerwartet-Interessant bzw. als Nützlich bezeichnen.

Die Bewertung der präsentierten Umfragebefunde bei den Kriterien des Clusters Unerwartet-Interessant zeigt Abbildung 1. Wie man sieht, beurteilen alle Gruppen, die irgendein Treatment bekamen (eigene Prognosen, Rückmeldung der Prognosen anderer oder beides) die Umfragebefunde in jedem der 20 Fälle im Mittel positiver als die Vpn der Kontrollgruppe ( $p=0,000$  nach dem Vorzeichentest). Insbesondere erkennt man, dass die Umfragebefunde nach den Treatments als vergleichsweise weniger offensichtlich beurteilt werden. Die Art der Treatments spielt offenbar keine systematische Rolle. Insofern zeigt Abbildung 1, dass die Treatments sich dahingehend ausgewirkt haben, dass die Umfrageergebnisse als unerwarteter, interessanter und überraschender erlebt werden.

**Abbildung 1: Bewertung von Umfragebefunden auf den Kriterien des Clusters Unerwartet-Interessant**





Bei den Kriterien des Clusters Nützlich zeigt sich dagegen kein systematischer Effekt der Treatments (Abbildung 2). Hier sind die Skalenmittelwerte der Treatmentgruppen in neun Fällen größer, in zehn Fällen kleiner und einem Fall gleich groß wie der entsprechende Skalenmittelwert der Kontrollgruppe. Bei keiner der Skalen sind alle Mittelwerte der Treatmentgruppen größer als der Mittelwert der Kontrollgruppe. Zudem ist der Range der Mittelwerte kleiner als beim Cluster Unerwartet-Interessant.

**Abbildung 2: Bewertung der Umfragebefunde auf den Kriterien des Clusters „nützlich“**

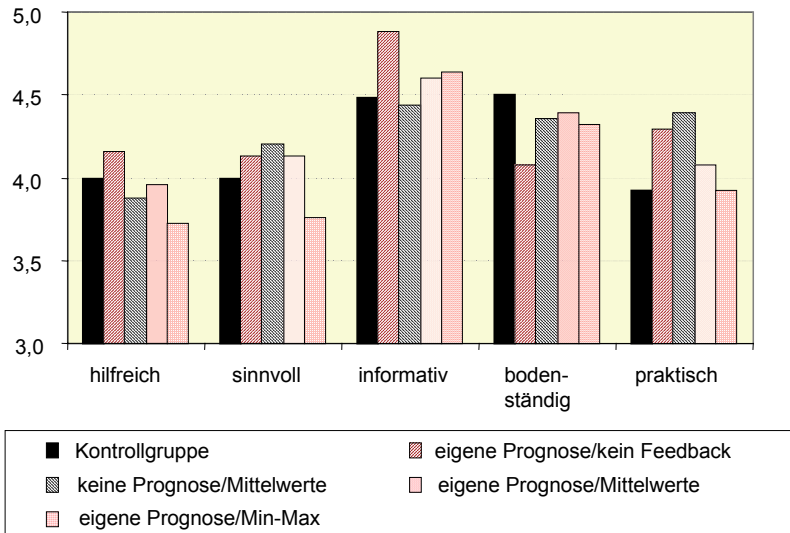
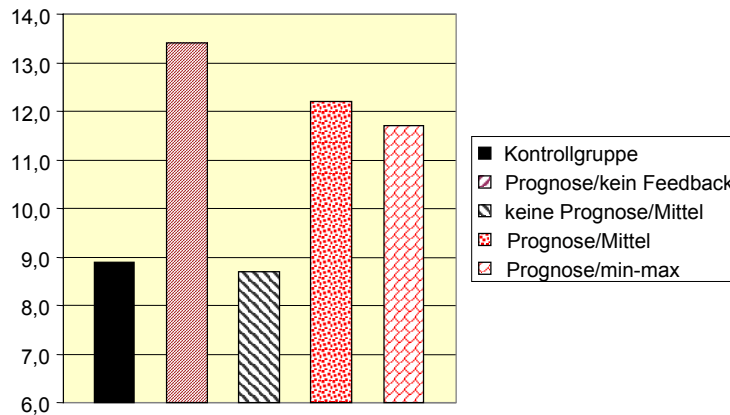


Abbildung 3 zeigt, dass das Treatment „eigene Prognosen“ auf eine andere abhängige Variable einen einfachen Effekt hat: Wie man sieht werden die Umfragebefunde besser erinnert, wenn die Vpn eigene Prognosen abgegeben haben. Der Erinnerungsskore wurde dabei so konstruiert, dass die Vpn einen Punkt dafür bekamen, wenn sie den Inhalt des Items erinnern konnten und einen zusätzlichen Punkt, wenn sie den Umfragewert innerhalb von plus/minus 5 Prozentpunkten richtig wiedergeben konnten. Wie man sieht, sind die Erinnerungswerte deutlich (und signifikant i.S. von t-Tests) besser für die Gruppen, die eigene Prognosen abgegeben haben.

**Abbildung 3: Mittlere Erinnerungswerte für die Umfrageitems unter fünf Bedingungen**



Systematische Effekte der Treatments zeigen sich auch beim Vergleich der Antworten auf die allgemeinen Fragen „Für wie sinnvoll halten Sie psychologische Forschungsarbeit?“ und „Halten Sie es für schwierig, die Ergebnisse sozialwissenschaftlicher Umfragen vorherzusagen?“, die ganz am Anfang und nochmals ganz zum Schluss des Experiments gestellt wurden. Während bei der Kontrollgruppe die mittleren Ratings nur um 0,06 bzw. um 0,03 Skaleneinheiten ansteigen, sind die Veränderungen bei den vier Treatmentgruppen deutlich größer: Beim Kriterium „sinnvoll“ wachsen bei den Treatmentgruppen die Ratings um 0,44\*, 0,48\*, 0,64\*\* bzw. 0,64\* Einheiten; beim Kriterium „schwierig“ wachsen sie um 0,68\*, 0,20, 0,40\*\* bzw. 0,32 Einheiten (paarweise t-Tests). Die Vpn in den Treatmentgruppen meinen also am Ende des Experiments verstärkt, dass es schwierig ist, Umfrageergebnisse vorherzusagen. Ebenso führen die Treatments dazu, dass psychologische Forschung ganz allgemein als sinnvoller bewertet wird.

Keine systematische Beziehung zu den Bewertungen oder zu den Erinnerungswerten haben dagegen die Grandiositätsscores der Vpn. Sie korrelieren lediglich mit dem Rating, wie schwer es ist, Umfragedaten i. Allg. vorherzusagen, signifikant. Die Korrelationen sind  $r=-0,22^*$  vor den Treatments und  $r=-0,27^*$  nach den Treatments. Vpn mit hohen Grandiositätswerten sind also allgemein der Meinung, dass es relativ leicht ist, Umfragedaten vorherzusagen. Diese Überzeugung steigt nach dem Experiment noch etwas an, obwohl sie die tatsächlichen Umfragebefunde auch nicht anders – insbesondere auch nicht als weniger unoffensichtlich – bewerten als Personen mit geringen

Grandiositätswerten. Allerdings erweist sich die Grandiositätsskala für die 125 Vpn als wenig homogen ( $\text{Alpha}=0,71$ ). Subskalen bzw. die Einzelitems korrelieren allerdings alle ähnlich wie die Gesamtskala mit den anderen Variablen. Sie korrelieren insbesondere alle negativ mit der Einschätzung, wie schwierig die Vorhersage von Umfragedaten ist.

#### **4. Diskussion**

Wir haben hier gezeigt, dass es möglich ist, die Offensichtlichkeit von Umfragebefunden durch die Verwendung von Prognosen zu reduzieren. Genauer gesagt geben Personen, die eigene Prognosen abgegeben haben, aber auch solche, die nur die Prognosen anderer gesehen haben, in stärkerem Ausmaß an, dass sie die Umfragebefunde überraschend, interessant und nicht offensichtlich finden. Die praktische Relevanz der Umfragebefunde wird dagegen von den verschiedenen Gruppen nicht systematisch verschieden bewertet. Prognosen führen zudem dazu, dass Umfragebefunde ganz allgemein als schwieriger vorhersagbar und psychologische Forschung allgemein als sinnvoller eingeschätzt wird.

Ein nicht erwarteter Befund ist, dass die Abgabe eigener Prognosen zwar die Erinnerung an die Items und die Umfragebefunde verbessert, aber nicht dazu führt, dass die Umfragebefunde als interessanter bewertet werden als in dem Fall, in dem die Vpn keine eigenen Prognosen abgeben, sondern nur die Prognosen anderer sehen. Überlegenswert erscheint aber, ob hier nicht doch ein Effekt dann entstehen könnte, wenn man den Personen ihre jeweils eigenen, individuellen Prognosen und nicht irgendwelche Mittelwerte zurückspegelt. Allerdings erscheint es möglich, dass dieser Ansatz unter gewissen Bedingungen auch kontraproduktiv sein kann. So zeigen Sanna u.a. (2002), dass auch die Methode des „consider the opposite“ dann nicht zu einem Debiasingeffekt führt, wenn die Personen es schwer finden, Argumente dafür zu finden, dass die berichteten Befunde wenig wahrscheinlich sind. Die Prognosetechnik basiert in gewissem Umfang darauf, dass einige der Prognosen falsch sind oder dass, im Fall von Fremdprognosen, zumindest der Range der Prognosen groß ist. Ist das nicht so, besteht kein theoretischer Grund, warum Prognosen die Offensichtlichkeit der Befunde reduzieren sollten. Nur im Fall eigener Prognosen könnte eine Offensichtlichkeitsreduktion auch dadurch entstehen, dass den Personen das Abgeben der Prognosen erlebbar schwer fällt, dass sie also nicht so leicht „verfügbar“ sind.

Das Persönlichkeitsmerkmal Grandiosität führt zu keinen besonderen Einsichten. Das kann allerdings daran liegen, dass diese Skala mit Studenten als Probanden entwickelt wurde. Die Vpn in dieser Studie waren aber keine Studenten. Für weitere Experimente wäre es daher von Interesse, eher bewährte Instrumente zur Messung der Persönlichkeit zu verwenden, weil unserer Meinung nach die Hypothese, dass die Bewertung von Um-

fragebefunden von einer allgemeinen Disposition der Besserwisserei o.ä. abhängt, nicht ohne weiteres aufgeben werden sollte.

Grundsätzlich lässt sich auch diskutieren, ob Lazarsfeld mit seinem Problem der Offensichtlichkeit überhaupt den Nagel auf den Kopf getroffen hat und ob es nicht vielmehr um die Frage von Trivialität geht. Ob die Personen das Umfrageergebnis „interessant“ finden, ob sie „das so vermutet hätten“, erscheint letztlich nicht entscheidend: Die Personen können ja durchaus ein Ergebnis richtig vorhersagen, es dann aber trotzdem nicht trivial finden, weil das Ergebnis den Status einer Meinung von dem einer Hypothese zu dem einer bestätigten Hypothese, also salopp gesprochen von Vermutung zu Wissen, verändert hat. Aus wissenschaftlicher Sicht jedenfalls ist es nicht entscheidend, dass ein Ergebnis überraschend ist, sondern dass es ein Datum – und keine noch so überzeugende Vermutung – ist, das auch anders hätte ausfallen können. Armstrong (2002: 5) bemerkt hierzu richtig, dass „findings need not be surprising to be important“. In der Tat sind die meisten wissenschaftlichen Befunde eher nicht überraschend. So zeigen Gordon u.a. (1978) z.B., dass Nichtpsychologen drei Viertel der Ergebnisse von 61 publizierten Forschungsstudien auf Grund der Beschreibung der Studie richtig voraussagen. Es bleibt also weiteren Studien überlassen, wie man durch eine geeignete Technik die scheinbare Trivialität von Umfragebefunden direkt reduzieren kann. Möglicherweise ist der Eindruck der Nicht-Offensichtlichkeit beim Nichtwissenschaftler aber leichter zu erzeugen als der von Nicht-Trivialität.

Problematisch bleibt bei dieser Untersuchung, dass die präsentierten Umfragebefunde nur fiktiv waren und somit die Frage danach, ob man das alles hätte vorhersagen können, keine objektive Bewertungsgrundlage hat. Ebenso problematisch ist es, dass die meisten Umfragethemen für die Vpn trotz aller Bemühungen im Pretest natürlich keine besondere Relevanz hatten. Trotzdem meinten die Vpn, die eigene Prognosen abgegeben haben und/oder die die Prognosen anderer gesehen haben, nach dem Experiment verstärkt, dass psychologische Forschung ganz allgemein sinnvoll ist – vielleicht auch deshalb, weil ihre Ergebnisse doch nicht so leicht vorhersagbar sind.

Bedenkenswert ist für das obige Experiment auch noch, dass Fehlbewertungen für die Vpn ohne offensichtliche negative Folgen blieben. Das ist im oben beschriebenen Kontext von Mitarbeiterbefragungen ganz anders. Dort sind die Daten für die Führungskräfte potentiell sehr bedeutsam: Fehlbewertungen können hier dazu führen, dass Geld verloren wird oder dass zumindest Chancen vertan werden. Wir erwarten daher, dass sich die Prognosen in einem relevanten Umfragekontext mit schwerwiegenden Folgen für die Rezipienten deutlicher auswirken sollten. Weitere Untersuchungen dazu werden wir noch anstellen.

## Korrespondenzadresse

*Prof. Dr. Ingwer Borg*  
 ZUMA  
 Postfach 12 21 55  
 68 072 Mannheim  
 email: borg@zuma-mannheim.de

*Carola Hillenbrand*  
 Henley Management College, Greenlands  
 Henley-on-Thames  
 Oxfordshire, RG9 3AU, England

## Literaturhinweise

- Armstrong, J.S., 2002: Discovery and communication of important marketing findings: Evidence and proposals. *Journal of Business Research*, 5775: 1-16.
- Baratz, D., 1983: How justified is the „obvious“ reaction? *Dissertation Abstracts International* 44/02B, 644B. Zitiert nach D.G. Myers 1995: *Social Psychology*. New York: MacGraw-Hill.
- Borg, I., 2002: What makes follow-up processes successful? Vortragspapier für die IT Survey Group. Boston, USA. April.
- Borg, I., 2003: *Führungsinstrument Mitarbeiterbefragung*. Göttingen: Verlag für angewandte Psychologie. 3. Auflage.
- Borg, I./Groenen, P.J.F., 1997: *Modern multidimensional scaling*. New York: Springer.
- Fiedler, K., 1993: Kognitive Täuschungen: Faszination eines modernen Forschungsprogramms S. 7-12 in: W. Hell (Hrsg.), *Kognitive Täuschungen: Fehlleistungen und Mechanismen des Urteilens, Denkens und Erinnerns*. Heidelberg: Spektrum.
- Fischhoff, B., 1977: Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance* 3: 349-358.
- Gordon, M.E./Kleinman, L.S./Hanie, C.A., 1978: Industrial-organizational psychology: open they ears o house of Israel. *American Psychologist* 33: 893-905.
- Lazarsfeld, P.F., 1949: The American soldier – an expository review. *Public Opinion Quarterly* 13: 377-404.
- Meyer, U., 1999: Entwicklung eines Fragebogens zum Konstrukt Narzismus. Forschungsbericht. Fachbereich 06 Psychologie, Uni Gießen.
- Meyer, U., 2000: Quantitative und qualitative Differenzierung narzistischer Persönlichkeitsmerkmale. In: C. Tarnai/J. Reinecke (Hrsg.), *Angewandte Klassifikationsanalyse*. Münster: Waxmann.
- Meyers, D.G., 1994: *Exploring social psychology*. New York: MacGraw-Hill.

Pohl, R., 1992: Der Rückschau-Fehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen. *Kognitionswissenschaft* 3: 38-44.

Sanna, L.J./Schwarz, N./Stocker, S.L., 2002: When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28: 497-502.

Stahlberg, D./Eller, F./Rohmann, A./Frey, D., 1995: Der Knew-it-all-along-Effekt in Urteilsituationen von hoher und geringer Selbstwertrelevanz. *Zeitschrift für Sozialpsychologie* 24: 94-102.