

Der Einfluss von Häufigkeitsformaten auf die Messung von subjektiven Wahrscheinlichkeiten

The Impact of Frequency Formats on the Measurement of Subjective Probability

Mandy Beuer-Krüssel und Ivar Krumpal

Zusammenfassung

Die Messung subjektiver Wahrscheinlichkeiten ist ein zentrales Anliegen vieler Bevölkerungssurveys zu selbstberichteter Delinquenz (z. B. ALLBUS 1990 und 2000). Ein bekanntes Problem ist hierbei die subjektive Überschätzung von Risiken im Zusammenhang mit seltenen Ereignissen. Fehler in der Risikoeinschätzung (z. B. ‚Nicht-Berücksichtigung von Basisraten‘ oder ‚Überschätzung‘) können sowohl auf kognitive Heuristiken der Befragten als auch auf Eigenschaften der Items zurückgeführt werden. Der erste Teil des Beitrags diskutiert und vergleicht Strategien und Formate der Messung von Wahrscheinlichkeiten, insbesondere Häufigkeiten versus Prozente. Hierbei zeigt sich, dass die Abfrage von Basisraten in Form von Häufigkeiten einen biasreduzierenden Effekt auf die Wahrscheinlichkeitseinschätzung seltener Ereignisse ausübt. Im zweiten Teil des Beitrags werden die theoretischen Vorteile von Häufigkeitsskalen in einen zweistufigen Messansatz subjektiver Wahrscheinlichkeiten überführt und durch ein Methodenexperiment empirisch belegt. Am Beispiel von subjektiven Entdeckungswahrscheinlichkeiten beim Schwarzfahren wird ein Kontexteffekt des Häufigkeitsformats demonstriert, der die Risikoüberschätzung seltener Ereignisse verringert. Ein solches Format könnte in der Praxis der empirischen Sozialforschung vermehrt Anwendung finden.

Abstract

In many surveys on deviant behavior the measurement of subjective probability is an important goal (e. g. German General Social Surveys (ALLBUS) 1990 and 2000). A well-known problem is the subjective overestimation of risks in connection with rare events. Errors in risk assessment (e. g. ‚base rate neglect‘ or ‚overestimation‘) can result from cognitive heuristics of the respondents as well as item characteristics. The first part of the article discusses strategies and formats of probability measurement and, in particular, compares frequencies versus percentages. It can be shown that the availability of base rates in the frequency format reduces bias in subsequent probability assessments of rare events. In the second part, the theoretical advantages of frequency scales are transformed into a two-step measurement procedure of subjective probability, and a bias-reducing effect of the frequency format is empirically demonstrated. A methodical experiment on fare dodging shows that subjective probabilities of being caught for dodging the fare (the rare event) are more accurate if base rates are activated via frequencies. This kind of format could be applied more frequently in empirical social research.

1 Einleitung¹

„As scientists and as technologists we should discard the idea of a 'true' or 'objective' probability. Instead, we should think of probability judgements as the result of an individual's feelings of uncertainty, translated into a numerical response by internal decision processes.“ (Phillips 1970: 254)

Die Relevanz der Messung von subjektiven Wahrscheinlichkeiten ist – trotz bekannten Problemen der Validität – unbestritten: In den Sozial- und Wirtschaftswissenschaften sowie in medizinischen Surveys bilden subjektive Wahrscheinlichkeiten die Basis für Entscheidungsfindungen, Verhaltensprognosen und Risikoeinschätzungen. So werden Befragte² im ALLBUS 1990 und 2000 nach ihrer Wahrscheinlichkeit gefragt, bei verschiedenen Delikten, wie Schwarzfahren oder Steuerhinterziehung, entdeckt zu werden. Die Messung dieser subjektiven Risiken erfolgt mit Kategoriialskalen, die über sogenannte *Vague Quantifier* verbalisiert sind.³ Andere Surveys messen subjektive Risiken, so etwa die subjektive Wahrscheinlichkeit Opfer einer Straftat zu werden, ebenso über *Vague Quantifier* (vgl. Wohlfahrtssurveys 1993, 1998; British Crime Surveys 2004, 2005). Studien der Umfragemethodologie und Kognitionspsychologie zeigen, dass die Verwendung von *Vague Quantifiers* häufig zu einer starken inter- und intrapersonellen Variabilität bei der Interpretation der verbalen Wahrscheinlichkeitsausdrücke und folglich zu Mehrdeutigkeiten und Widersprüchlichkeiten bei der Übertragung der *Vague Quantifiers* in numerische Äquivalente führt (vgl. Bradburn/Miles 1979; Krumpal et al. 2008; Wright et al. 1994).⁴ Die vorliegende Arbeit untersucht deshalb alternative Messansätze der Erfassung subjektiver Risiken.

Solche Alternativen sind numerische Skalen (Häufigkeiten oder Prozente), die sich durch eine geringere Anfälligkeit für Fehlrteile (Biases) sowie eine konsistentere Verwendung auszeichnen. Allerdings ist auch die Risikoeinschätzung

1 Wir danken Veronika Andorfer, Dorothea Böhr, Heiko Rauhut und Thomas Voss für anregende Gespräche, konstruktive Hinweise sowie ihre Unterstützung bei diesem Forschungsprojekt. Wir danken zudem den beiden anonymen Gutachtern für ihre wertvollen Verbesserungsvorschläge.

2 Im Folgenden wird die männliche Form verwendet. Dies dient ausschließlich der sprachlichen Vereinfachung.

3 Item zu Schwarzfahren im ALLBUS 2000: „Stellen Sie sich vor, Sie würden ein öffentliches Verkehrsmittel benutzen, ohne einen gültigen Fahrausweis zu besitzen. Wie wahrscheinlich wäre es Ihrer Ansicht nach, daß ein Kontrolleur Sie dabei entdecken würde? Benutzen Sie für Ihre Antwort bitte die Liste.“ Antwortvorgaben als Kategoriialskala mit *Vague Quantifiers*: ‚Sehr unwahrscheinlich‘, ‚Eher unwahrscheinlich‘, ‚Ungefähr 50 zu 50‘, ‚Eher wahrscheinlich‘, ‚Sehr wahrscheinlich‘.

4 Je nach Personen-, Item-, und Kontexteigenschaften variieren die Zuweisungen der numerischen zu den verbalen Risikoeinschätzungen, sodass eine Vergleichbarkeit zwischen Personen, Items oder Surveys in vielen Fällen problematisch ist. In einem Experiment von Reuband (2002) ordnen Befragte den *Vague Quantifiers* numerische Wahrscheinlichkeiten zu (Prozente oder Häufigkeiten). Hierbei zeigen sich Inkonsistenzen in den Zuordnungen. Eine Anwendung numerischer Skalen (insbesondere Häufigkeiten) wird deshalb empfohlen.

auf numerischen Skalen, insbesondere auf Prozentskalen, nicht unproblematisch (Hoffrage et al. 2000). So stehen subjektive Einschätzungen der Befragten häufig im Widerspruch zu mathematischen Wahrscheinlichkeitsgesetzen.⁵ Subjektive Wahrscheinlichkeiten korrespondieren oftmals nur unzureichend mit tatsächlichen Risiken (Johnson/Bruce 2001). Befragte unterschätzen systematisch die Wahrscheinlichkeit häufiger (bzw. alltäglicher) und überschätzen die Wahrscheinlichkeit seltener Ereignisse (*Overestimation*; vgl. Fischhoff et al. 2000; Pinkerton et al. 2000; Warr 1980; Yamagishi 1997). Zudem ist bekannt, dass Befragte die Wahrscheinlichkeit eines Ereignisses überwiegend danach beurteilen, wie repräsentativ es für ein jeweiliges Vergleichsobjekt ist. Entsprechend ignorieren sie wichtige Basisrateninformationen und sind unempfindlich gegenüber der Stichprobengröße (*Base Rate Neglect*; vgl. Tversky/Kahneman 1974).

Zunächst werden nachfolgend die Vorzüge von Häufigkeitsskalen gegenüber Prozentskalen theoretisch hergeleitet und diskutiert (vgl. Brase 2002; Gigerenzer/Hoffrage 1995; Gigerenzer 1998). Erstere können das wohl größte Problem der Wahrscheinlichkeitsbeurteilung eindämmen: Die subjektive Überschätzung des Risikos seltener Ereignisse. Auch andere Fehler, wie zum Beispiel die *Nicht-Berücksichtigung von Basisraten* oder die Fehlbeurteilung logischer Verknüpfungen, sind über die Häufigkeitsskala weniger stark ausgeprägt als über die Prozentskala. Die theoretischen Vorteile von Häufigkeitsskalen werden in einen zweistufigen Messansatz subjektiver Wahrscheinlichkeiten überführt und durch ein Methodenexperiment empirisch belegt. Am Beispiel von subjektiven Entdeckungswahrscheinlichkeiten beim Schwarzfahren wird ein Kontexteffekt des Häufigkeitsformats demonstriert, der die Risikoüberschätzung seltener Ereignisse verringert. Ein solches Format könnte in der Praxis der empirischen Sozialforschung vermehrt Anwendung finden.

5 Als Beispiel kann hier die Verletzung der Konjunktionsregel angeführt werden. Die Konjunktionsregel besagt, dass die Wahrscheinlichkeit einer logischen UND-Verknüpfung zweier Ereignisse nicht größer sein kann als die einer ihrer Komponenten: $P(A \cap B) \leq P(A)$ und $P(A \cap B) \leq P(B)$. Trotz der *Schlüssigkeit* dieses Wahrscheinlichkeitsgesetzes schätzen Befragte Ereigniswahrscheinlichkeiten sowie deren Verknüpfungen falsch ein. Veranschaulicht wird diese Fehleinschätzung durch das sogenannte ‚Linda Problem‘. Befragte werden zunächst mit Beschreibungen einer fiktiven Person konfrontiert und anschließend gebeten, die Wahrscheinlichkeit einzuschätzen, mit der ‚Linda‘ unterschiedliche Berufe und Hobbys ausübt. Es zeigt sich, dass die Verknüpfungen von Aktivitäten von der Mehrzahl der Befragten als wahrscheinlicher eingestuft werden als die einzelnen Aktivitäten selbst. Dies ist dann der Fall, wenn die Konjunktion als *repräsentativer* für die beschriebene Person eingestuft wird als ihre Komponenten (Tversky/Kahneman 1983). Beispielsweise wird ‚Linda‘ als selbstbewusste, ältere Single-Frau beschrieben, die Philosophie studierte sowie als Studentin an Themen wie Diskriminierung und sozialer Gleichberechtigung großes Interesse zeigte. Befragte wurden gefragt, ob es wahrscheinlicher sei, dass ‚Linda‘ Bankangestellte [P(B)] ist oder dass sie Bankangestellte [P(B)] UND innerhalb der Frauenbewegung aktiv [P(A)] ist [P(A ∩ B)]. Da die Konjunktion als wesentlich *repräsentativer* für die Persönlichkeitsbeschreibung von ‚Linda‘ eingestuft wird, überschätzt die Mehrzahl der Befragten deren Wahrscheinlichkeit. Das empirisch beobachtete Muster war hierbei wie folgt: $P(A) > P(A \cap B) > P(B)$ (vgl. Tversky/Kahneman 1983: 297f.).

2 Fehleinschätzungen von Risiken

2.1 Kognitiver Ansatz

Welche Mechanismen verursachen die zu hohen bzw. zu niedrigen Einschätzungen von Risiken? Hierzu wird zunächst eine terminologische Festlegung vorgenommen: (1) *Allgemeine Wahrscheinlichkeiten* beziehen sich auf Risiken der oder des ‚typischen Anderen‘. Befragte beurteilen demnach das Risiko, dass ein Ereignis X für eine Menge von Personen oder eine Einzelperson (nicht sie selbst) auftritt. (2) Der Terminus *subjektive Wahrscheinlichkeit* bezeichnet dagegen das Risiko des Befragten, dass ein Ereignis X für ihn persönlich eintritt. (3) Von beiden unabhängig bezeichnet die *tatsächliche* oder *objektive Wahrscheinlichkeit* das Risiko, mit dem ein Ereignis X in der Realität stattfindet.

Kognitive Ansätze erklären Fehleinschätzungen von Risiken über eine fehlerhafte Wahrnehmung und Informationsverarbeitung der Befragten. Sie nehmen an, dass menschliche Wahrnehmung von Rationalitätsannahmen abweicht, und dass subjektive Wahrscheinlichkeiten systematische Unterschiede zu den normativen Standards der Wahrscheinlichkeitstheorie aufzeigen (Kahneman/Tversky 1973, 1982a). Das Heuristik- und Bias-Konzept von Kahneman und Tversky kann als Grundlage des Kognitiven Ansatzes angesehen werden:

„In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors.“ (Kahneman/Tversky 1973: 237)

Die *Repräsentativitätsheuristik* beschreibt die Bewertung des Ausmaßes der Übereinstimmung zwischen einem Sample und einer Population, einem Beispiel und einer Kategorie, d. h. inwieweit die Objekt- oder Ereigniseigenschaften den Kategorieneigenschaften ähneln.⁶ Sie beruht nicht ausschließlich auf dem Prinzip der Gleichheit, sondern kann ebenso kausale Annahmen und Wechselwirkungsbeziehungen widerspiegeln. Ihre Verwendung führt oftmals zu Insensibilität gegenüber vorhergehenden Wahrscheinlichkeiten (Basisraten) und Stichprobengrößen sowie

6 Die Gleichheit von Sample und Population wird z. B. über den Vergleich von Verhältnissen ermittelt. Dabei wird das Mengenverhältnis von Objekten in Beziehung gesetzt. Geht man bei Geburten von einer natürlichen Geschlechterverteilung von Mädchen zu Jungen von 50:50 aus, so ist es nicht verwunderlich, dass Befragte eine Geschlechterverteilung von z. B. 3:3 als wahrscheinlicher einschätzen als eine von 1:5. Auch das Prinzip der Zufälligkeit scheint bei letzterem Beispiel verletzt (vgl. Kahneman/Tversky 1972: 432ff.).

zu Unempfindlichkeit hinsichtlich der Vorhersagbarkeit von Ereignissen⁷ (Tversky/Kahneman 1974: 1124ff.). Dagegen generiert die *Verfügbarkeitsheuristik* Wahrscheinlichkeitseinschätzungen in Abhängigkeit von der Leichtigkeit, mit der relevante Beispiele erinnert oder vorgestellt werden können.⁸ Die Anzahl relevanter Fälle, die schnell und ohne Aufwand kognitiv verfügbar sind, bilden demnach die Grundlage zur Einschätzung von Wahrscheinlichkeiten. Die Leichtigkeit des Erinnerns bzw. der Konstruktion von ähnlichen Szenarien entspricht jedoch häufig nicht der tatsächlichen Auftrittswahrscheinlichkeit der zu beurteilenden Ereignisse (Tversky/Kahneman 1974: 1127f.). Als Folge der Anwendung dieser zwei Heuristiken kann es zu verschiedenen Fehleinschätzungen kommen.⁹

Fehleinschätzungen resultieren u. a. aus der *Nicht-Berücksichtigung von Basisraten*, d. h. Informationen zu vorangegangenen Wahrscheinlichkeiten, Populationsverhältnissen oder zentralen Tendenzen. Mit der *Nicht-Berücksichtigung von Basisraten* (Kahneman/Tversky 1973) einher geht die Überbewertung von spezifischen Informationen über das Zielereignis, über welches Vorhersagen getroffen werden. Jede subjektive Wahrscheinlichkeitseinschätzung basiert auf (1) Einzelfallinformationen sowie (2) Basisraten. Die Gewichtung der beiden Faktoren ist abhängig von der Diagnostizität, die ihnen der Befragte zuschreibt. In vielen Fällen erfolgt eine Einschätzung lediglich basierend auf der *Repräsentativität* der Einzel-

- 7 Ein Beispiel wird bei Tversky und Kahneman (1974: 1126) beschrieben: Zwei Gruppen von Befragten werden Beschreibungen der Leistung von Lehramtsstudenten in deren praktischer Übungsstunde gegeben. Die eine Gruppe soll die Qualität der beschriebenen Unterrichtsstunde evaluieren, während die andere den Ruf/die Kompetenz der Lehramtsstudenten fünf Jahre nach dieser Übungsstunde vorhersagen soll. Die abgegebenen Urteile unter diesen beiden Bedingungen waren identisch. Die Vorhersage des entfernt liegenden Ereignisses war identisch mit den Evaluationen zur Qualität der Übungsstunde. Die Prognosen richteten sich demnach ausschließlich an der begrenzten Vorhersagekraft dieses Einzelereignisses aus.
- 8 Ihr empirischer Nachweis kann beispielsweise in Worthäufigkeits-Experimenten erbracht werden. Es werden die Wörter als häufiger in einem Text auftretend eingestuft, die leichter vorzustellen sind. Ist es wahrscheinlicher, dass ein zufällig aus einem englischen Text gezo- genes Wort mit einem ‚k‘ beginnt oder ein ‚k‘ an dritter Stelle hat? Da Wörter mit dem An- fangsbuchstaben ‚k‘ einfacher abrufbar/vorstellbar sind, werden diese als häufiger vorkom- mend eingestuft. Die Realität zeigt jedoch den umgekehrten Fall (Tversky/Kahneman 1973: 211f.). Zentral ist hierbei die subjektiv erfahrene Leichtigkeit der Erinnerung: „Presumably, they monitor their cognitive processes and infer that a given class of events is frequent when relevant exemplars are easy to bring to mind but rare when exemplars are difficult to bring to mind“ (Schwarz 1998: 88).
- 9 Neben der Erklärung des Auftretens von systematischen Fehleinschätzungen über die ge- nannten kognitiven Heuristiken kann als weiterer Erklärungsansatz das Argument Essers (1986) angebracht werden, wonach Antwortverzerrungen stärker im Zusammenhang mit dem Frageinhalt betrachtet werden sollten. Laut Esser treten Antwortverzerrungen vor allem dann auf, wenn die Einstellungsintensität der Befragten schwach ausgeprägt ist, und somit die Relevanz der Frage für die Befragten fehlt bzw. gering ist. So könnten die beobachteten Fehleinschätzungen in den Worthäufigkeits-Experimenten (siehe Fußnote 8) ebenso vor dem Hintergrund einer möglicherweise fehlenden Relevanz der Fragestellung für die Befragten beurteilt werden.

fallinformationen, wohingegen Basisraten kaum Einfluss zeigen. In Experimenten wurden Befragte mit Persönlichkeitsbeschreibungen und Listen mit Studienfächern sowie der Frage konfrontiert, wie groß die Wahrscheinlichkeit sei, dass die fiktive Person ihr Studium im jeweiligen Fach abgeschlossen hat. Es zeigte sich, dass die Abschlussfächer in dem Maße als wahrscheinlich eingestuft wurden, in dem das Persönlichkeitsprofil mit den stereotypen Mitgliedern des jeweiligen Faches korrespondierte, also repräsentativ für dieses war. Dass die Fächer Unterschiede in der realen Studentenzahl aufweisen (also verschiedene Basisraten), hatte keinen Einfluss auf die Wahrscheinlichkeits einschätzungen.¹⁰

Hier liefert der Kognitive Ansatz jedoch bereits Lösungsstrategien für das Problem der *Nicht-Berücksichtigung von Basisraten*. Dieses könne folglich vermindert werden, indem Basisraten nicht nur als willkürliche Angaben über die Verteilung in der Gesamtpopulation vermittelt werden, sondern ihre Diagnostizität, Relevanz, Spezifität und Kausalität hervorgehoben wird (Bar-Hillel 1980: 216). Ziel der Messung subjektiver Wahrscheinlichkeiten sollte es demnach sein, die individualisierten Einzelfallinformationen *und die Basisraten* als gleichwertige, relevante Informationen zu kommunizieren, damit insbesondere letztere bei der Wahrscheinlichkeitseinschätzung berücksichtigt werden (Ajzen 1977; Bar-Hillel 1980; Gigerenzer et al. 1988; Ginosar/Trope 1980; Nisbett/Ross 1980). Befragte greifen eher dann auf gegebene Basisraten-Informationen zurück, wenn inkonsistente und nicht kausale Einzelfallinformationen, die nur schwer in Bezug auf existierende stereotype Repräsentationen zu interpretieren sind, zu Mehrdeutigkeiten führen (Ginosar/Trope 1980; Hendrickx et al. 1989; Schwarz et al. 1991). Koehler (1996) kann zeigen, dass von den Befragten direkt erfahrene oder selbst generierte Basisraten eher genutzt werden als statistisch vermittelte, und dass Basisraten in Aufgabenkontexten mit

10 Die fiktive Person ‚Tim‘ wird als intelligent, ordnungsliebend und systematisch sowie in ihrem Schreibstil mechanisch beschrieben. ‚Tim‘ ist wenig kreativ, hat wenig Sympathie für andere Menschen und generell wenig Spaß an der Interaktion mit anderen. Die normativen Regeln der Vorhersage (welches ‚Tims‘ Abschlussfach sei) wurden durch die Anwendung der *Repräsentativitätsheuristik* von der Mehrheit der Befragten verletzt. Über 95 % schätzten die Wahrscheinlichkeit, dass ‚Tim‘ Informatikabsolvent ist, als größer ein als die, dass er seinen Abschluss in den Geistes- oder Erziehungswissenschaften hat. Das obwohl ihnen klar war, dass es sehr viel mehr Absolventen in den letzteren beiden Fächern gibt (wie es die Basisraten vorgeben). Siehe hierzu die Experimentalbeschreibungen bei Kahneman und Tversky (1973: 238f.). Ob die Basisrate tatsächlich ignoriert oder ihre Diagnostizität von den Befragten lediglich unterschätzt wird, müsste darüber hinaus dadurch geprüft werden, ob sich die Einschätzung bei zusätzlich angegebener Information über Basisraten im Vergleich zum Fehlen dieser Information verändert oder nicht.

einer natürlichen und alltäglichen Problematik stärker verwendet werden als die im Kontext künstlicher oder konstruierter Probleme.¹¹

Das Überschätzen der Auftrittswahrscheinlichkeit seltener sowie das Unterschätzen der Wahrscheinlichkeit alltäglicher bzw. häufiger Ereignisse gehört ebenfalls zu den systematischen Fehlern, welche sich aufgrund der Nutzung von Heuristiken ergeben können. Beispielsweise werden bei der Frage nach der Wahrscheinlichkeit verschiedener Todesursachen die seltenen (Unfälle, Selbstmord, Feuer) überschätzt, während die alltäglichen (Diabetes, Schlaganfall) unterschätzt werden (Slovic et al. 2004). *Über- und Unterschätzung* sind außerdem stark affektiv bestimmt. Insbesondere Ereignisse, die mit schwerwiegenden Folgen verbunden sind oder in den Medien überbetont werden, erfahren die stärkste Überschätzung (Warr 1980). Begründet wird diese Verzerrung über kognitive Mechanismen des Erinnerns und der Betroffenheit: Die angenommene Wahrscheinlichkeit des Eintretens eines Ereignisses ist abhängig davon, wie *leicht* dieses Ereignis *erinnert* oder vorgestellt werden kann. Je eher ein Ereignis *kognitiv verfügbar* ist und je folgenreichtiger es ist, desto eher wird seine Auftrittswahrscheinlichkeit überhöht beurteilt (Pinkerton et al. 2000).¹²

2.2 Inputorientierter Ansatz

Die Ursachen der Beurteilungsverzerrungen sieht der Inputorientierte Ansatz nicht primär bei den Individuen, sondern bei den Eigenschaften der Items. Die funktionalen Beurteilungsanlagen der Befragten sehen sich mit einem suboptimalen Input konfrontiert, der zu den fehlerhaften Risikoeinschätzungen führt. Dabei spielen vorangehende Fragen, die Itemformulierung und das Skalenformat eine wichtige Rolle (Cosmides/Tooby 1996; Gigerenzer et al. 1991; Koehler 1996).

11 Persönlich erfahrene Informationen (Basisraten) sind anschaulicher und hervorstechender, damit also schneller verfügbar als ‚gelernte‘ Informationen. Zudem haben Menschen mehr Vertrauen in selbst generierte Basisraten, vor allem wenn diese über Erfahrungen aus erster Hand erworben wurden. Ausführliche Beschreibungen sowie methodische Anleitungen finden sich bei Koehler (1996).

12 In Experimenten mit US-Studenten zum Thema HIV-Infektion zeigte sich, dass diese die Wahrscheinlichkeit einer Ansteckung mit HIV stark überschätzten (im Vergleich zu den tatsächlichen Risiken). Begründet wurden dieser Effekt mit der extensiven Berichterstattung zur HIV-Epidemie in den Medien und den schweren Konsequenzen einer HIV-Infektion. Das Thema HIV wurde durch eine starke Medienpräsenz kognitiv verfügbar gemacht, was in vielen Fällen zu einer Überschätzung der tatsächlichen (geringen) Risiken der allgemeinen Bevölkerung geführt hat (Pinkerton et al. 2000: 16f.).

Prozentskalen

Im Gegensatz zu kategorialen Skalen mit einer geringen Anzahl verbaler Antwortkategorien erlauben Prozentskalen (üblicherweise Skalen von 0 bis 100 %, also 101 wählbaren Alternativen) eine präzisere Kommunikation von subjektiven Wahrscheinlichkeiten (Schnell/Kreuter 2000; Coutts-Heller 2002: 8).¹³ Die Kommunikation von Wahrscheinlichkeiten im Prozentformat ist dennoch fehleranfällig, da die von den Befragten angegebenen allgemeinen oder subjektiven Wahrscheinlichkeiten häufig von den tatsächlichen Risiken abweichen. Oftmals werden Ereigniswahrscheinlichkeiten im Prozentformat überschätzt (Black et al. 1995; Dominitz/Manski 1997). Fischhoff et al. (2000) konnten in ihrer Studie über die Erwartungen von Jugendlichen bezüglich wichtiger Lebensereignisse zeigen, dass die Befragten sowohl die Wahrscheinlichkeit positiver als auch negativer Ereignisse¹⁴ überschätzen. *Überschätzung* kann unter anderem mit der übermäßigen Nutzung von 50 %-Antworten erklärt werden. Befragte verwenden diesen Wert weniger in seiner natürlichen numerischen Form, sondern interpretieren ihn vielmehr als Kategorie ‚keine Ahnung‘. In einer Studie zu Schwangerschaftsrisiken deuteten die Befragten die Wahrscheinlichkeit von 50 % für einen genetischen Defekt beim Kind als ‚alles ist möglich‘ (Lippman-Hand/Fraser 1979: 118f.). Die vermehrte ‚Flucht in die Mitte‘, welche vor allem durch die Unsicherheit der Eltern begründet ist, führte so im prozentualen Wahrscheinlichkeitsformat zu einer Überschätzung der Risiken.¹⁵ Ein weiterer Grund für die überproportionale Verwendung der Mittelposition der Skala ist im sozialen Vergleich zu finden. Befragte orientieren sich bei der Kommunikation ihrer subjektiven Wahrscheinlichkeit häufig an der Referenzpopulation. Dabei interpretieren sie die Skalenmitte als ‚normalen‘ oder Durchschnittswert für diese Population und ordnen sich entsprechend ein (Schwarz et al. 1985; Wright et al. 1994).

Die Überschätzung von kleinen Ereigniswahrscheinlichkeiten kann ebenso durch die Präferenz der Befragten für runde Zahlen erklärt werden. Befragte neigen dazu ihre Antworten auf Werte wie 0, 10, ..., 50, ... und 100 % zu runden, statt die verfeinerten Wahrscheinlichkeitsangaben auf der Prozentskala zwischen

13 „For example, when subjects say it is 'likely' that someone will get a headache as a result of using a certain medicine and say it is 'likely' that someone else will develop a more severe symptom as a result of using another medicine, they are not referring to the same numeric probability. The 'likely' probability of the less severe symptom will be higher (...).“ (Coutts-Heller 2002: 7). Eine weiterführende Diskussion der Probleme von *Vague Quantifiers* findet sich bei Krumpal et al. (2008).

14 Zum Beispiel: Positiv – Abschluss des College bis zum 30. Lebensjahr. Negativ – Inhaftierung oder Tod innerhalb des nächsten Jahres.

15 Eine Möglichkeit zur Reduzierung von 50 %-Antworten ist die Einführung einer zusätzlichen Antwortkategorie ‚keine Ahnung‘ oder ‚neutral‘ (Coutts-Heller 2002: 9).

0 und 100 % im vollen Umfang auszunutzen (Dominitz/Manski 1997: 270). Befragte benutzen die Prozentskala häufig wie eine Kategorialskala, die von 0 bis 10 reicht, und geben Antworten oftmals als Vielfache von 10. Da in den Studien aber meist Ereignisse untersucht werden, deren Wahrscheinlichkeiten (weit) unter 10 % liegen, könnte ein solches Befragtenverhalten zu den beobachteten Überschätzungen führen (Slovic/Monahan 1995). Subjektive Wahrscheinlichkeitseinschätzungen spezifischer Ereignisse mit binärem Ausgang (wie die Entdeckungswahrscheinlichkeit eines Deliktes im kommenden Jahr) sind im Prozentformat fehleranfällig. Diese Probleme können auf evolutionäre Prozesse in der Entwicklung der menschlichen kognitiven Fähigkeiten zurückgeführt werden. Menschen besitzen demnach keinen angeborenen kognitiven Mechanismus, der ihnen die Verarbeitung von Einzelergebniswahrscheinlichkeiten im Einklang mit den normativen Standards der Wahrscheinlichkeitstheorie ermöglicht (Brase et al. 1998; Cosmides/Tooby 1996; Gigerenzer 1996b, 1998, 2000).

Häufigkeitsskalen

Laut Gigerenzer (2000) liege der Hauptschwachpunkt der kognitiven und motivationalen Erklärungen für die Fehler bei der Risikoeinschätzung darin begründet, dass diese die Struktur der Aufgabe und deren Beziehung zur Struktur der natürlichen Umwelt des Befragten ausblenden würden. Die Fehleinschätzungen der Befragten seien keine Fehler im statistischen Denken, sondern lediglich bedingt durch eine suboptimale Konstruktion von Informations- und Antwortformaten innerhalb der Studien:

„The mind acts as if it were a frequentist; it distinguishes between single events and frequencies in the long run – just as probabilists and statisticians do. Despite the fact that researchers in the 'heuristics and biases' program routinely ignore this distinction fundamental to probability theory when they claim to have identified 'errors', it would be foolish to label these judgments 'fallacies.' (Gigerenzer 2000: 253f.)

Die Verwendung von Häufigkeitsformaten würde demnach, verglichen mit Einzelergebniswahrscheinlichkeiten im Prozentformat, zu einer klareren Einschätzung von Risiken führen. Häufigkeiten stellen eine intuitivere Metrik bei der Risikobeurteilung dar als Prozente (Gigerenzer/Hoffrage 1995; Schapira et al. 2001). Der Mensch verfügt über induktive Denkmechanismen, die rationale Prinzipien zum Ausdruck bringen. Diese werden aber nur dann korrekt angewendet, wenn Informationen in Häufigkeiten präsentiert und abgerufen werden. Die Präferenz für natürliche Häufigkeiten als Datenbasis und Skalenformat ist dabei evolutionär bedingt. Eine Datenbank, auf die der Mensch schon in Urzeiten zurückgreifen konnte, waren

seine eigenen, zählbaren Beobachtungen. Die beobachteten und aufsummierten Ereignishäufigkeiten dienen der Verbesserung von Entscheidungsprozessen, wie dem Finden erfolgversprechender Jagdgebiete. Beim *natürlichen Sampling* werden keine Basisraten benötigt um konditionale Wahrscheinlichkeiten zu ermitteln (vgl. Brase 2002; Brase et al. 1998; Gigerenzer 1996b; Gigerenzer/Hoffrage 1995).¹⁶ In konstruierten Experimenten braucht es jedoch Basisraten um die fehlenden Ausgangsinformationen zu liefern. Die Kommunikation von Risiken über Prozente, Chancen oder Odds-Ratios enthält genau diese Informationen nicht. Im Gegensatz zu normalisierten (relativen) Häufigkeiten und Wahrscheinlichkeiten beinhalten natürliche Häufigkeiten Informationen über Basisraten, weshalb sie als Informationsbasis für die Wahrscheinlichkeitsbeurteilung adäquater sind; sie stimmen mit den natürlichen Enkodierungsmechanismen¹⁷ überein (Cosmides/Tooby 1996; Gigerenzer 2000; Gigerenzer/Hoffrage 1999; Hoffrage et al. 2000, 2002; Jones et al. 1995). Natürliche Häufigkeiten sind (verglichen mit Prozenten) weniger abstrakt, einfacher zu verstehen sowie zu visualisieren und benötigen weniger kognitiven Aufwand bei der Einschätzung von Risiken (Brase 2002; Gigerenzer 1996b, 2000; Griffin/Buehler 1999). Zudem erhöhen sie die Anwendung der *Bayes Regeln*.¹⁸

Die Informationspräsentation in Häufigkeiten statt Prozenten steigert die Beurteilungsleistung der Befragten. Studien zur Qualität von Antwortskalen im Hinblick auf Wahrscheinlichkeitseinschätzungen zeigen, dass das Ersetzen von Prozent- durch Häufigkeitsskalen Fehleinschätzungen deutlich verringert und so zu Leistungssteigerungen bei der Genauigkeit der Wahrscheinlichkeitseinschätzungen führt (Brase et al. 1998; Evans et al. 2000; Gigerenzer et al. 1991; Hoffrage et al. 2000). Während Basisraten bei Prozentinformationen und bei Abfrage auf einer Prozentskala kaum von den Befragten genutzt werden, kann die *Nicht-Berücksichtigung von Basisraten* durch die Präsentation von natürlichen Häufigkeiten und die Nutzung von Häufigkeitsskalen als Antwortmöglichkeit reduziert werden (Gigerenzer 1998,

16 *Natürliches Sampling* bezeichnet die aufeinanderfolgende und fortlaufende Erfassung von Informationen über das Aktualisieren von Ereignishäufigkeiten, also das Zählen von erlebten oder überlieferten Objekten und Ereignissen (Gigerenzer/Hoffrage 1995: 686).

17 Fehler, welche sich bei der Einschätzung von Wahrscheinlichkeiten ergeben, seien demnach vorwiegend auf das Informationsformat zurückführbar (vgl. Gigerenzer/Hoffrage 1995: 685ff.). Die von Menschen erworbenen mathematischen Algorithmen seien lediglich für bestimmte Repräsentationen angelegt – die Eingangsinformationen müssten demnach im richtigen Format vorliegen, damit die Algorithmen optimal arbeiten können (wie auch ein Taschenrechner **nicht** mit binären Zahlen arbeiten kann).

18 Die Berechnung von Ereigniswahrscheinlichkeiten über Bayesianische Algorithmen fällt im Häufigkeitsformat leichter, da weniger und einfachere Rechenschritte notwendig sind. Detaillierte Informationen und Beispielstudien zu den *Bayes Regeln* und ihrem Zusammenhang zum Informationsformat geben Gigerenzer (1996a, 1996b, 1998, 2000), Gigerenzer/Hoffrage (1995, 1999), Hoffrage et al. (2002) sowie Mellers/McGraw (1999).

2000; Gigerenzer/Hoffrage 1999; Price 1998). Das Überschätzen seltener und Unterschätzen häufiger Ereignisse kann durch die Präsentation von Informationen im Häufigkeitsformat nicht vollständig unterbunden werden. Jedoch können Häufigkeiten das Ausmaß dieser Fehleinschätzungen verringern und zu realistischeren Werten führen (Teigen 1974: 62).

3 Möglichkeiten der Biasreduktion

Häufigkeitsskalen können vor allem bei der Ermittlung von Basisraten, also beispielsweise der *allgemeinen* Auftretswahrscheinlichkeit eines Ereignisses für eine Population, viele Vorteile bieten. Basisraten stehen jedoch nicht immer im Mittelpunkt des Interesses. Vielmehr sollen individuelle Entscheidungsprozesse und Handlungen erklärt werden, weshalb die Abfrage *subjektiver* Wahrscheinlichkeiten notwendig ist (z. B. Viktimisierungs- und Entdeckungswahrscheinlichkeiten). Da deren Messung mit Häufigkeitsskalen nur schwer realisierbar ist (Coutts-Heller 2002: 12), kann lediglich auf Prozentskalen zurückgegriffen werden. Hierbei wäre es von großem Vorteil den Beurteilungsprozess dergestalt zu beeinflussen, dass die Nutzung der Prozentskala mehr im Einklang mit der Nutzung der Häufigkeitsskala steht. Studien zu Kontexteffekten bei der Einstellungsmessung zeigen, dass formal gleiche Fragen in Abhängigkeit des Kontextes zu unterschiedlichen Antworten führen (Schwarz/Sudman 1992). Über die Variation des Kontextes wird also die Beantwortung von Folgefragen beeinflusst.

3.1 Antwortprozess und Kontexteffekte

In Surveys kann der Prozess der Beantwortung von Fragen in vier Phasen unterteilt werden: (1) Informationsaufnahme und Verstehen des Frageinhalts, (2) Abrufen relevanter Informationen aus dem Gedächtnis, (3) Benutzung der erinnerten Informationen zur Beantwortung der Frage sowie (4) Selektion und Wiedergabe einer Antwort (Tourangeau et al. 2000: 7). Informationsverarbeitung und Handeln sind Prozesse, die häufig unbewusst erfolgen. Gewissermaßen wird automatisch ein Konstrukt gewählt. Welches Konstrukt verfügbar ist, ist abhängig von aktiven Erinnerungshinweisen – unabhängig davon, ob sich der Befragte dieser Operation bewusst ist. Verfügbarkeit beschreibt dabei die Leichtigkeit, mit der eine Episode, emotionale Reaktion, vorangegangene Beurteilung, Wissensstruktur oder ein kognitives Konstrukt ins Bewusstsein gerufen wird (Feldman 1992: 51ff.). Die Erinnerung relevanter Informationen wird bestimmt von der Frageformulierung, den Frageinstruktionen, der Verfügbarkeit früherer Überlegungen und Urteile sowie dem Inhalt vorangegangener Items (Tourangeau 1992: 36).

Kontexteffekte sind Antworteffekte, die durch eine oder mehrere vorangegangene Fragen (und Antworten) oder durch die *Skalen* vorhergehender Fragen hervorgerufen werden (Billiet et al. 1992: 131). Diese vorangegangenen Fragen haben zwei Funktionen: (1) Informationen zu aktivieren, sodass diese zu einem späteren Zeitpunkt leichter (automatisch und unbewusst) abgerufen werden können, und (2) eine Informationsbasis zu liefern, über die sich die Fragebedeutung erschließt. Über einen *Priming Effekt* sind Informationen, die entweder direkt in der vorangegangenen Frage enthalten sind oder vom Befragten bei deren Beantwortung aktiviert wurden, zugänglicher und strahlen auf die Folgefrage aus (Strack 1992: 25). Eine Beurteilung, die einmal abgegeben wurde, dient entsprechend als *Anker*, dem nachfolgende Beurteilungen angeglichen oder kontrastiert werden (Knowles et al. 1992: 223).

3.2 Gesteuerte Kontexteffekte bei der Messung subjektiver Wahrscheinlichkeiten

Die Generierung subjektiver Wahrscheinlichkeiten hinsichtlich des Eintretens seltener Ereignisse ist häufig geprägt von *Überschätzung*, d. h. Befragte schätzen diese Wahrscheinlichkeiten im Vergleich zu objektiven Standards zu hoch ein (Pinkerton et al. 2000; Teigen 1974; Warr 1980).¹⁹ Eine Erklärung für diesen Bias ist, dass Befragte oftmals Basisraten, die die allgemeine Wahrscheinlichkeit repräsentieren, nicht in ihre Beurteilungen einschließen²⁰, obwohl sie einen adäquaten Anker darstellen. Durch gezielte Manipulation des Befragungskontextes kann die Verfügbarkeit solch ‚erwünschter‘ Informationen jedoch erhöht und so die Beurteilung des Befragten beeinflusst werden.²¹

Kreuter (2002) stellt fest, dass die Wahrnehmung von Wahrscheinlichkeiten und Risiken ein mehrdimensionales Konstrukt ist. Bei den Überlegungen zu ihren eigenen Risiken nannten Befragte „zunächst die allgemeinen Risiken, besannen sich auf ihre üblichen Alltagsroutinen und leiteten daraus ihre Antwort auf das eigene Risiko ab“ (Kreuter 2002: 228). Da jedoch nicht davon ausgegangen werden kann, dass alle Befragten diese kognitive Kalkulation ausführen, somit unterschiedliche

19 Aus den theoretischen Überlegungen von Esser (1986) könnte zudem argumentiert werden, dass das Ausmaß der Fehleinschätzungen vom Frageinhalt moderiert wird. Demnach wären bei Befragten mit einer stark ausgeprägten Einstellungsintensität geringere Fehleinschätzungen zu erwarten.

20 Viel wichtiger noch: Basisraten werden in vielen Studien als Referenzanker weder abgefragt noch vorgegeben.

21 Beispielstudien zu Kontexteffekten bei der Wahrscheinlichkeitsmessung finden sich u. a. bei Coutts-Heller (2002), Hoorens/Buunk (1993), Rottenstreich/Tversky (1997) sowie Windschitl (2002).

Interpretationsgrundlagen bestehen würden und folglich die Ergebnisse schwer vergleichbar wären, ist es sinnvoll, die Abfrage allgemeiner Wahrscheinlichkeiten (Basisraten) generell vor der subjektiver Wahrscheinlichkeiten vorzunehmen. Aufgrund der erhöhten kognitiven Verfügbarkeit würde dies die Nutzung der Basisraten bei der individuellen Wahrscheinlichkeitseinschätzung erhöhen und damit Fehleinschätzungen verringern. Dabei ist es besonders hilfreich, die Basisraten nicht lediglich vorzugeben, sondern von den Befragten zu erfragen. So werden nicht (scheinbar) willkürlich Werte festgelegt, sondern die subjektiven Annahmen und Erfahrungen der Befragten berücksichtigt, was die Glaubhaftigkeit, Relevanz und Eindeutigkeit der Basisraten unterstreicht und ihre Weiterverwendung in Folgefragen zusätzlich verstärken kann (vgl. Koehler 1996).

3.3 Allgemeine Hypothesen

In einem ersten Schritt führt die Abfrage allgemeiner Wahrscheinlichkeiten (Ankerfrage) zur Aktivierung von Informationen über die Verteilung bestimmter Eigenschaften oder Ereignisse in der Referenzpopulation. In einem zweiten Schritt nutzen die Befragten diese selbst generierten Basisrateninformationen zur Beurteilung ihrer subjektiven Wahrscheinlichkeit (Zielfrage). Somit liegt ein Kontexteffekt der vorangegangenen Ankerfrage auf die nachfolgende Zielfrage vor. Zur Abfrage von Basisraten sollten zudem Häufigkeitsskalen benutzt werden. Denn ihre Verwendung führt zu einer weiteren Reduzierung vieler systematischer Einschätzungsfehler, wie *Überschätzung*.

Durch eine experimentelle Variation des Skalenformats (Häufigkeitsskala versus Prozentskala) in der Ankerfrage kann somit überprüft werden: (1) Ob über die Häufigkeitsskala die allgemeine Wahrscheinlichkeit tatsächlich adäquater²² gemessen werden kann als mit Hilfe der Prozentskala. (2) Ob sich ein Kontexteffekt des Skalenformats der Ankerfrage auf die Beurteilung der subjektiven Wahrscheinlichkeit insofern ergibt, als die Zielfrage realistischer eingeschätzt wird, wenn die allgemeine Wahrscheinlichkeit zuvor über eine Häufigkeits- statt Prozentskala abgefragt wurde.

Die Nutzung von Häufigkeitsskalen bei der Wahrscheinlichkeitsbeurteilung anstelle von Prozentskalen erhöht die Eichung und Kohärenz des Urteils bezogen auf objektive Werte (vgl. Coutts-Heller 2002: 10). Außerdem stellt die Verschiebung –

22 D. h. weniger überschätzt bei seltenen Ereignissen (bzw. weniger unterschätzt bei häufigen Ereignissen) und somit näher an objektiven Wahrscheinlichkeiten. Die Häufigkeitsskala ruft bei der Abfrage allgemeiner Wahrscheinlichkeiten einen regelgeleiteten Denkmodus hervor, während die Prozentskala oftmals lediglich Heuristiken (wie die *Repräsentativitätsheuristik*) aktiviert.

weg von der Beurteilung von Einzelereigniswahrscheinlichkeiten (Prozenten) hin zu Einschätzungen der Ereigniswahrscheinlichkeiten ganzer Klassen von Menschen oder Objekten (Häufigkeiten) – einen Perspektivenwechsel im Wahrscheinlichkeitsdenken dar: Während Häufigkeitsskalen einen distributionalen Modus aktivieren, lösen Prozentskalen einen singulären Modus aus. Die Vorhersagen im distributionalen Modus basieren auf Wissen über Basisraten in Referenzpopulationen. Im singulären Modus hingegen beruhen sie auf Einstellungen zum und Erfahrungen über die Eigenschaften des spezifischen, zu beurteilenden Objekts/Ereignisses und sind damit anfälliger für systematische Fehleinschätzungen. Wahrscheinlichkeitsbeurteilungen im distributionalen Modus (Häufigkeitsformat) führen zu akkurateren Einschätzungen:²³

Hypothese 1

Die Ankerfrage im Häufigkeitsformat liefert realistischere Einschätzungen der allgemeinen Wahrscheinlichkeit als die im Prozentformat. Erstere führt zu geringeren Wahrscheinlichkeitseinschätzungen seltener Ereignisse.

Zudem wird vermutet, dass die Messung der allgemeinen Wahrscheinlichkeiten im Häufigkeitsformat (verglichen mit dem Prozentformat) einen biasverringenden Effekt auf die darauffolgende Abfrage der subjektiven Wahrscheinlichkeiten ausübt (vgl. Billiet et al. 1992: 131). Die abhängige Variable ‚subjektive Wahrscheinlichkeit‘ wird hierbei in beiden Fällen über eine Prozentskala gemessen. Es wird vermutet, dass der durch die vorgeschaltete Häufigkeitsskala aktivierte distributionale Denkprozess und die damit einhergehende stärkere Berücksichtigung von Basisraten, sich auf die nachfolgende Zielfrage übertragen und dort ebenso zu adäquateren Wahrscheinlichkeitseinschätzungen führen. Bei der Ankerfrage mit Prozentskala werden dagegen weniger günstige kontextuelle Auswirkungen auf die Zielfrage erwartet. Auch Reeves und Lockhart (1993) konnten in ihrer Studie zeigen, dass die Abfrage von Wahrscheinlichkeiten über das Häufigkeitsformat vor der Abfrage von Einzelereigniswahrscheinlichkeiten (diese schließen subjektive Wahrscheinlichkeiten ein) zu deutlichen Leistungsverbesserungen führt. Entsprechend kann eine Hypothese zum vermuteten Kontexteffekt abgeleitet werden:

23 Beispielstudien zur Unterscheidung des distributionalen und singulären Modus finden sich bei Bruine de Bruin et al. (2000), Kahneman/Tversky (1982b), Klar et al. (1996), Koehler (2001) sowie Reeves/Lockhart (1993): „Frequency problems evoke a distributional approach of probability and are mentally represented in such a way that the relevance of extensional rules is more compelling. Case-specific problems evoke a singular approach and are modeled in ways that support the use of nonextensional heuristics such as representativeness. [...] Presumably, solving frequency problems first not only cued extensional rules but also provided a model for how those rules could be applied. [...] Evidently, simply having subjects solve problems that naturally evoke extensional rules will lead them to apply extensional rules to problems that would otherwise evoke nonextensional strategies.“ (Reeves/Lockhart 1993: 212)

Hypothese 2

Befragte, welche die allgemeine Wahrscheinlichkeit zunächst auf einer Häufigkeitsskala beurteilen, schätzen ihre eigene Wahrscheinlichkeit in einer darauffolgenden Frage subjektiv geringer ein als Befragte, die zuvor die Ankerfrage im Prozentformat beantwortet haben.

4 Analyse einer experimentellen Schwarzfahrerstudie

4.1 Design

Unsere Analysen basieren auf einem Methodenexperiment, welches an der Universität Leipzig durchgeführt wurde. Im Zentrum der Studie steht neben allgemeinen Fragen zur Nutzung öffentlicher Nahverkehrsmittel die subjektive Entdeckungswahrscheinlichkeit beim Schwarzfahren. Dabei wurden Fragebögen an insgesamt 405 Studenten innerhalb zweier Vorlesungen zu zwei verschiedenen Messzeitpunkten verteilt: 242 Befragte nahmen an der ersten Welle im Juni sowie 163 an der zweiten im Juli 2007 teil. Im Abstand von zwei Wochen wurden in den Vorlesungen zwei unterschiedliche Fragebogenversionen im Wechsel an die Gruppen ausgegeben, vor Ort von den Studenten ausgefüllt und wieder abgegeben. Innerhalb dieser Fragebogenversionen wurde das Format zur Messung der allgemeinen Entdeckungswahrscheinlichkeit experimentell variiert (Häufigkeits- versus Prozentskala). Durch die Erhebung zu zwei Zeitpunkten mit zwei Gruppen wird es möglich, beide Befragtengruppen mit beiden Fragebogenversionen zu konfrontieren (faktorielles ‚between-within-Design‘²⁴). Das Wirken beider Anker auf die subjektive Entdeckungswahrscheinlichkeit kann also sowohl zwischen als auch innerhalb der Gruppen verglichen werden.

Die Fragebögen²⁵ enthalten neben dieser experimentellen Variation identische Fragen zu demografischen Aspekten (Alter, Geschlecht, Wohndauer in Leipzig, Hauptfach, Hochschulsemester), zu Themen der spezifischen Nutzung von öffentlichen Nahverkehrsmitteln (Transportmittel, Jahreszeit, Uhrzeit, Fahrscheinart, durchschnittliche Fahrtdauer), zur Beurteilung der Ticketpreise und des Schwarzfahrens sowie zur Einschätzung des aktiven Schwarzfahrens von Freunden/Bekanntem und des Befragten selbst. Der Einschätzung der subjektiven Entdeckungswahrscheinlichkeit beim Schwarzfahren geht eine Ankerfrage voraus, die die allgemeine Entdeckungs-

24 Detaillierte Informationen zum between- und within-Design sowie deren Anwendungen finden sich bei Birnbaum (1999), Kahneman/Tversky (1996), Price (1998) sowie Shadish et al. (2002).

25 Der vollständige Fragebogen in seinen zwei Versionen, der Datensatz und alle Analyseroutinen sind auf Anfrage erhältlich.

wahrscheinlichkeit thematisiert und deren Format sich zwischen den Versionen unterscheidet. Mittels dieser wird der Kontext vor der Zielfrage variiert, indem Basisraten über zwei unterschiedliche Skalen (Häufigkeits- und Prozentskala) abgefragt werden. Die Zielfrage befindet sich abgetrennt auf der nächsten Seite des Fragebogens, sodass die Ankerfrage bei deren Beantwortung nicht direkt sichtbar ist. Der Einfluss der vorangegangenen Frage wirkt also vorwiegend über die Aktivierung und kognitive Bereitstellung von Informationen, statt unmittelbar visuell zugänglich zu sein; das Rückblättern der Befragten kann dennoch nicht ausgeschlossen werden.

Abbildung 1 Die zwei Ankerfragen und die Zielfrage des Fragebogens

Ankerfrage im Häufigkeitsformat (experimenteller Split 1)

Angenommen, 100 beliebige Personen würden tagsüber (zwischen 7:00 und 17:00 Uhr) ohne einen gültigen Fahrschein mit öffentlichen Nahverkehrsmitteln (z. B. Straßenbahn oder Bus) in Leipzig jeweils eine Strecke von ca. 25 Minuten Dauer zurücklegen. Wie viele von diesen 100 Personen würden Ihrer Ansicht nach dabei von einem Kontrolleur entdeckt werden?

Geben Sie hierzu ihre Einschätzung auf einer Skala von 0 = „es wird keiner entdeckt“ bis 100 = „es werden alle entdeckt“ an. Sie können bei Ihrer Antwort jeden beliebigen Wert dazwischen angeben.

_____ Personen von 100 Personen

Ankerfrage im Prozentformat (experimenteller Split 2)

Angenommen, eine beliebige Person (nicht Sie selbst) würde tagsüber (zwischen 7:00 und 17:00 Uhr) ohne einen gültigen Fahrschein mit öffentlichen Nahverkehrsmitteln (z. B. Straßenbahn oder Bus) in Leipzig eine Strecke von ca. 25 Minuten Dauer zurücklegen. Für wie wahrscheinlich halten Sie es, dass diese Person dabei von einem Kontrolleur entdeckt wird?

Geben Sie hierzu ihre Einschätzung auf einer Skala von 0 % = „sie wird auf keinen Fall entdeckt“ bis 100 % = „sie wird auf jeden Fall entdeckt“ an. Sie können bei Ihrer Antwort jeden beliebigen Wert dazwischen angeben.

_____ Prozent

Zielfrage (beide Splits)

Angenommen, Sie würden tagsüber (zwischen 7:00 und 17:00 Uhr) ohne einen gültigen Fahrschein mit öffentlichen Nahverkehrsmitteln (z. B. Straßenbahn oder Bus) in Leipzig eine Strecke von ca. 25 Minuten Dauer zurücklegen. Für wie wahrscheinlich halten Sie es, dass Sie dabei von einem Kontrolleur entdeckt werden?

Geben Sie hierzu ihre Einschätzung auf einer Skala von 0 % = „ich werde auf keinen Fall entdeckt“ bis 100 % = „ich werde auf jeden Fall entdeckt“ an! Sie können bei Ihrer Antwort jeden beliebigen Wert dazwischen angeben.

_____ Prozent

In Anwendung der allgemeinen Hypothesen auf die Entdeckungswahrscheinlichkeiten beim Schwarzfahren resultieren die folgenden spezifischen Vorhersagen: (1) Die Ankerfrage im Häufigkeitsformat liefert realistischere Einschätzungen der allgemeinen Entdeckungswahrscheinlichkeit als die im Prozentformat. Erstere führt zu geringeren Wahrscheinlichkeitseinschätzungen. (2) Befragte, welche die allgemeine Entdeckungswahrscheinlichkeit zunächst auf einer Häufigkeitsskala beurteilen, schätzen in der darauffolgenden Frage ihre subjektive Entdeckungswahrscheinlichkeit geringer ein als Befragte, die zuvor die Ankerfrage im Prozentformat beantwortet haben.²⁶

4.2 Empirische Befunde

In der Fragebogenversion mit Häufigkeitsformat werden generell signifikant niedrigere Werte angegeben als in der mit Prozentformat, womit beide Vorhersagen zunächst bestätigt werden können.²⁷ Tabelle 1 liefert eine Übersicht der Ergebnisse des Mittelwertvergleichs:

Tabelle 1 Differenzen der Mittelwerte der allgemeinen und subjektiven Entdeckungswahrscheinlichkeit in Abhängigkeit des Skalensformats der Ankerfrage

	Allgemeine Entdeckungswahrscheinlichkeit	Subjektive Entdeckungswahrscheinlichkeit
Häufigkeitsformat (H)	23,10 (217)	33,57 (217)
Prozentformat (P)	35,09 (180)	39,87 (180)
Differenz (H-P)	-11,99	-6,30
p-Wert (H<P)	0,00	0,02

Anmerkung: Dargestellt sind die Mittelwerte (Fallzahlen in Klammern) der allgemeinen und subjektiven Entdeckungswahrscheinlichkeit, unterteilt nach Prozent- versus Häufigkeitsformat der Ankerfrage. Die empirischen Signifikanzniveaus (p-Werte) der Mittelwertdifferenzen wurden auf Grundlage von einseitigen T-Tests (Hypothese: $H < P$) berechnet.

- 26 Im Rahmen unseres Experiments lautet das seltene Ereignis ‚beim Schwarzfahren entdeckt werden‘. Unseren beiden Hypothesen liegt die Annahme zugrunde, dass die Einschätzung der allgemeinen und subjektiven Entdeckungswahrscheinlichkeit dann realistischer ist, wenn geringere Werte angegeben werden. Diese ‚weniger ist besser – Annahme‘ kann nicht explizit getestet werden, da die tatsächliche Entdeckungswahrscheinlichkeit nicht bekannt ist. Wir nehmen jedoch an, dass die tatsächliche Entdeckungswahrscheinlichkeit hinreichend klein ist und in beiden Formaten subjektiv überschätzt wird.
- 27 Informationen über die empirischen Häufigkeitsverteilungen der allgemeinen und subjektiven Entdeckungswahrscheinlichkeiten sind im Anhang zu finden.

Zunächst ist auffällig, dass sich die Angaben zwischen den allgemeinen und subjektiven Entdeckungswahrscheinlichkeiten unterscheiden, wobei letzteren im Durchschnitt höhere Werte zugeordnet werden.²⁸ Die Einschätzung der allgemeinen Entdeckungswahrscheinlichkeit im Häufigkeitsformat fällt verglichen mit der Abfrage im Prozentformat signifikant geringer aus (Differenz -11,99). Zudem verringert die Nutzung der Häufigkeitsskala als Anker, im Vergleich zur Nutzung der Prozentkala als Anker, die darauffolgende subjektive Entdeckungswahrscheinlichkeit beim Schwarzfahren (Differenz -6,30).

Die Befunde zu den allgemeinen Entdeckungswahrscheinlichkeiten können unter Kontrolle der Variablen (1) Befragtengruppe und (2) Welle bestätigt werden: Befragte, die die Ankerfrage im Prozentformat beantwortet haben, zeigen signifikant höhere Werte als die Vergleichsgruppe mit Häufigkeitsformat (Inter-Gruppenvergleich, Differenzen -16,55 und -5,97). Der Effekt des Skalenformats kann demnach für beide Gruppen unabhängig nachgewiesen werden (vgl. Tabelle 2).

Tabelle 2 Differenzen der Mittelwerte der allgemeinen Entdeckungswahrscheinlichkeit in Abhängigkeit des Skalenformats, der Welle und Befragtengruppe

		Gruppe 1 (G1)	Gruppe 2 (G2)	Differenz (H-P)	p-Wert (H<P)
Welle 1	Prozentformat (P)	39,02 (100)			
	Häufigkeitsformat (H)		22,47 (138)	-16,55	0,00
Welle 2	Prozentformat (P)		30,17 (80)		
	Häufigkeitsformat (H)	24,20 (79)		-5,97	0,05
Differenz (H-P)		-14,82	-7,70		
p-Wert (H<P)		0,00	0,00		

Anmerkung: Dargestellt sind die Mittelwerte (Fallzahlen in Klammern) der allgemeinen Entdeckungswahrscheinlichkeit, unterteilt nach Prozent- versus Häufigkeitsformat der Ankerfrage. Die empirischen Signifikanzniveaus (p-Werte) der Mittelwertdifferenzen wurden auf Grundlage von einseitigen T-Tests (Hypothese: $H < P$) berechnet.

Auch innerhalb der Gruppen ergeben sich in Abhängigkeit des Skalenformats deutliche Unterschiede (Intra-Gruppenvergleich, Differenzen -14,82 und -7,70). Gruppe 1 hat in der ersten Welle die Ankerfrage im Prozent- und in der zweiten im Häufigkeitsformat beantwortet.

28 Vgl. hierzu Studien zur Überschätzung eigener Risiken und zum *Optimismus/Pessimismus Bias* von Dolinski et al. (1987), Franic/Pathak (2000), Hoorens/Buunk (1993), Klar et al. (1996), Pinkerton et al. (2000).

figkeitsformat bearbeitet. Bei Gruppe 2 ergibt sich entsprechend die umgekehrte Reihenfolge. Innerhalb beider Gruppen zeigt sich jeweils, dass Befragte, die die allgemeine Entdeckungswahrscheinlichkeit auf der Häufigkeitsskala beantwortet haben, geringere Werte für diese Variable angeben als solche, die die Prozentskala genutzt haben. Die Datenanalyse zeigt über alle Gruppenvergleiche hinweg starke Effekte der Häufigkeitsskala auf die allgemeine Entdeckungswahrscheinlichkeit in die erwartete Richtung. Somit kann Hypothese 1 bestätigt werden.

Bei den subjektiven Entdeckungswahrscheinlichkeiten zwischen den Gruppen fallen die erwarteten Kontexteffekte schwächer aus (Inter-Gruppenvergleich): In der ersten Erhebungswelle zeigt sich der erwartete Effekt des Skalenformats der Ankerfrage auf die darauffolgende subjektive Entdeckungswahrscheinlichkeit. Befragte, die die Kontextfrage mit Häufigkeitsskala bearbeitet haben, zeigen bei der Zielfrage signifikant niedrigere Werte als solche, die zuvor mit Hilfe der Prozentskala geantwortet haben (Differenz $-8,85$). Dieser Kontexteffekt des Skalenformats der Ankerfrage auf die Zielfrage zeigt in der zweiten Welle ebenso in die erwartete Richtung, fällt jedoch recht schwach aus (Differenz $-2,81$) und ist zudem nicht signifikant (auf den konventionellen 1 %-, 5 %- bzw. 10 %-Niveaus). Tabelle 3 visualisiert diese Befunde.

Tabelle 3 Differenzen der Mittelwerte der subjektiven Entdeckungswahrscheinlichkeit in Abhängigkeit des Skalenformats, der Welle und Befragtengruppe

		Gruppe 1 (G1)	Gruppe 2 (G2)	Differenz (H-P)	p-Wert (H<P)
Welle 1	Prozentanker (P)	41,81 (100)			
	Häufigkeitsanker (H)		32,96 (138)	-8,85	0,01
Welle 2	Prozentanker (P)		37,44 (80)		
	Häufigkeitsanker (H)	34,63 (79)		-2,81	0,28
Differenz (H-P)		-7,18	-4,48		
p-Wert (H<P)		0,05	0,14		

Anmerkung: Dargestellt sind die Mittelwerte (Fallzahlen in Klammern) der subjektiven Entdeckungswahrscheinlichkeit, unterteilt nach Prozent- versus Häufigkeitsformat der Ankerfrage. Die empirischen Signifikanzniveaus (p-Werte) der Mittelwertdifferenzen wurden auf Grundlage von einseitigen T-Tests (Hypothese: $H < P$) berechnet.

Auffällig ist die Differenz ($37,44 - 41,81 = -4,37$) bei der subjektiven Entdeckungswahrscheinlichkeit mit jeweils vorgeschaltetem Prozentanker zwischen den Grup-

pen.²⁹ Befragte der Gruppe 2, denen der Prozentanker in Welle 2 vorgelegt wurde, zeigen eine geringere subjektive Entdeckungswahrscheinlichkeit (nicht signifikant: einseitiger T-Test; $p=0,16$) verglichen mit Befragten der Gruppe 1, denen der Prozentanker bereits in Welle 1 vorgelegt wurde. Somit kann ein Kontexteffekt von Welle 1 auf Welle 2 nicht ausgeschlossen werden: Es besteht die Vermutung, dass in Gruppe 2 durch die Benutzung des Häufigkeitsankers und der sich daraus ergebenden geringen subjektiven Entdeckungswahrscheinlichkeit in Welle 1, diese auch in Welle 2 – trotz Prozentanker – nach unten korrigiert wurde. Solch ein Reihenfolgeeffekt (Häufigkeiten in Welle 1 und Prozente in Welle 2) wäre eine mögliche Erklärung des nur schwachen Kontexteffektes in Gruppe 2.

Innerhalb der Gruppen zeigen alle Vergleiche der subjektiven Wahrscheinlichkeiten ebenso in die erwartete Richtung (Intra-Gruppenvergleich), wobei auch hier die angenommenen Effekte (verglichen mit Hypothese 1) schwächer ausfallen. Befragte der Gruppe 1, welche die Ankerfrage im Häufigkeits- statt Prozentformat beantwortet haben, zeigen eine geringere subjektive Entdeckungswahrscheinlichkeit (Differenz $-7,18$). Auch bei Befragten der Gruppe 2 zeigt sich der Effekt in die erwartete Richtung, allerdings schwächer (Differenz $-4,48$) und nicht signifikant ($p=0,14$).

Erneut drängt sich die Vermutung auf, dass ein Reihenfolgeeffekt (Gruppe 1: Prozente in Welle 1 und Häufigkeiten in Welle 2; Gruppe 2: Häufigkeiten in Welle 1 und Prozente in Welle 2) den Kontexteffekt der Anker- auf die Zielfrage moderiert. Tabelle 2 und Tabelle 3 zeigen, dass die Schätzungen in der zweiten Welle sowohl durch das Format der ersten wie der zweiten Welle beeinflusst werden. Befragte der Gruppe 1, die in der ersten Welle im Prozentformat antworteten, berichten in der zweiten Welle höhere Werte im Häufigkeitsformat als Befragte der Gruppe 2, die in der ersten Welle unmittelbar im Häufigkeitsformat antworteten.³⁰ Umgekehrt resultiert in Gruppe 2 das Häufigkeitsformat in Welle 1 in niedrigeren Werten im Prozentformat in Welle 2.³¹ Dies deutet auf einen Ankereffekt über Welle und Format hin. Dieser Befund ist konsistent mit Ergebnissen bisheriger Studien, die zeigen, dass Manipulationen zum Zeitpunkt $t1$ die kognitive Repräsentation verändern, auf die zum Zeitpunkt $t2$ zurückgegriffen wird (vgl. Carlston 1980; Sherman

29 Bezüglich der Häufigkeitsanker sind die Angaben zur subjektiven Entdeckungswahrscheinlichkeit für die Gruppen nahezu identisch (Differenz: $34,63 - 32,96 = 1,67$; einseitiger T-Test: $p=0,34$).

30 In Tabelle 2 beträgt die Differenz $1,73$ (einseitiger T-Test; $p=0,28$) und in Tabelle 3 beträgt die Differenz $1,67$ (einseitiger T-Test; $p=0,34$).

31 In Tabelle 2 beträgt die Differenz $-8,85$ (einseitiger T-Test; $p=0,01$) und in Tabelle 3 beträgt die Differenz $-4,37$ (einseitiger T-Test; $p=0,16$).

1980).³² Allerdings zeigen drei von vier Tests keine signifikanten Differenzen. Die von uns aufgestellten Vermutungen zum Reihenfolgeeffekt sollten daher in einer weiteren Untersuchung (mit höherer Fallzahl) repliziert werden.

Zusammenfassend kann auch Hypothese 2 der Tendenz nach bestätigt werden, wobei weitere Untersuchungen eines möglichen Interaktionseffektes von (1) Skalenformat (Häufigkeitsanker versus Prozentanker) und (2) Präsentationsabfolge der Formate (Abfolge: Häufigkeitsanker-Prozentanker versus Prozentanker-Häufigkeitsanker) auf die Einschätzung der subjektiven Entdeckungswahrscheinlichkeit lohnenswert scheinen.³³

5 Diskussion

Im theoretischen Teil der Arbeit wurde die Problematik der Messung von subjektiven Wahrscheinlichkeiten besprochen. Auffällig sind die sich aus ihr ergebenden zahlreichen Fehleinschätzungen, wie *Überschätzung* und *die Nicht-Berücksichtigung von Basisraten*. Die Gründe hierfür sind zum einen in den kognitiven Prozessen der

32 Als eine mögliche Erklärung für diesen Effekt diskutiert Sherman (1980: 218) die erhöhte Verfügbarkeit von stereotypen Antwortsequenzen (sog. kognitiven ‚Skripten‘), die bei der erstmaligen Beantwortung der Frage gebildet, anschließend im Gedächtnis gespeichert und zu späteren Zeitpunkten (bei ähnlichen Fragestellungen) leichter aktiviert werden können. Ähnlich argumentiert Carlston (1980: 324): „Cognitive processes occurring after stimulus observation can alter the information subjects have available for making later impression judgements, and consequently, can alter the impression judgments that are made.“

33 Neben dem Vergleich der konditionalen Mittelwerte wurden zudem 2 multiple OLS-Regressionen mit Interaktionstermen geschätzt: Im ersten Modell ist die abhängige Variable die allgemeine Entdeckungswahrscheinlichkeit (AEW), im zweiten Modell ist die abhängige Variable die subjektive Entdeckungswahrscheinlichkeit (SEW). In beiden Modellen stehen jeweils auf der rechten Seite der Gleichung die beiden Experimentalvariablen Skalenformat (SF: 1 = Häufigkeiten, 0 = Prozente) und Präsentationsabfolge (PA: 1 = Prozente zuerst, 0 = Häufigkeiten zuerst) sowie der entsprechende Interaktionsterm Skalenformat (SF) * Präsentationsabfolge (PA). Im ersten Modell zeigen die Regressionskoeffizienten der beiden Experimentalvariablen in die theoretisch erwartete Richtung: $AEW = 30,17 - 7,70 * SF + 8,85 * PA - 7,12 * SF * PA$. Die Koeffizienten für das Skalenformat ($p=0,01$) und die Präsentationsabfolge ($p=0,01$) sind jeweils signifikant, der Interaktionsterm Skalenformat * Präsentationsabfolge ($p=0,12$) ist dagegen nicht signifikant. Im zweiten Modell zeigen die Regressionskoeffizienten der beiden Experimentalvariablen ebenfalls in die theoretisch erwartete Richtung: $SEW = 37,44 - 4,48 * SF + 4,37 * PA - 2,70 * SF * PA$. Im Vergleich zum Prozentformat, resultiert die Fragebogenversion mit der Ankerfrage im Häufigkeitsformat in niedrigere subjektive Entdeckungswahrscheinlichkeiten. Im Vergleich zur Reihenfolge ‚Häufigkeiten zuerst‘, resultiert die Reihenfolge ‚Prozente zuerst‘ in höhere subjektive Entdeckungswahrscheinlichkeiten. Im Vergleich zum ersten Modell deuten die Koeffizienten im zweiten Modell auf schwächere Effekte hin. Zudem sind die Koeffizienten nicht signifikant ($p=0,28$ für das Skalenformat; $p=0,32$ für die Präsentationsabfolge; $p=0,66$ für den Interaktionsterm Skalenformat * Präsentationsabfolge). In einer separaten Analyse wurden zudem robuste Standardfehler geschätzt (Huber-White-Sandwich-Varianzschätzer). Die hieraus resultierenden p-Werte bleiben nahezu unverändert.

Befragten zu finden, die ihre Einschätzungen häufig mittels einfacher Heuristiken, wie der *Repräsentativitätsheuristik* und der *Verfügbarkeitsheuristik*, vornehmen. Zum anderen führt aber auch die Anwendung inadäquater Präsentations- und Abfrageformate zu den Fehleinschätzungen. Wie die Durchsicht früherer theoretischer und empirischer Forschungsarbeiten gezeigt hat, stellen die in zahlreichen Surveys verwendeten, mit *Vague Quantifiers* verbalisierten Kategorienskalen eine nur ungenaue und häufig inkonsistente Form der Risikoeinschätzung und -kommunikation dar (vgl. Bradburn/Miles 1979; Krumpal et al. 2008; Wright et al. 1994). Ausgangspunkt des vorliegenden Artikels war es deshalb, adäquatere Alternativen zur Erfassung von Wahrscheinlichkeiten zu untersuchen und experimentell zu vergleichen. Es wurden zwei numerische Formate (Häufigkeitsskala versus Prozentskala) vergleichend untersucht. Hierbei zeigte sich ein Einfluss des Formats auf die Wahrscheinlichkeitseinschätzungen der Befragten. Im Vergleich zum Prozentformat führt das Häufigkeitsformat zu niedrigeren Wahrscheinlichkeitseinschätzungen hinsichtlich des Eintretens seltener Ereignisse.

Im Rahmen einer Schwarzfahrerstudie wurde ein experimentelles Design angewendet, welches in einem zweistufigen Ansatz die Abfrage allgemeiner Entdeckungswahrscheinlichkeiten vor der Abfrage subjektiver Entdeckungswahrscheinlichkeiten erfordert und das Wirken von Kontexteffekten annimmt. Die Ergebnisse dieses Methodenexperiments zeigen, dass die Ankerfrage im Häufigkeitsformat zu niedrigeren allgemeinen Wahrscheinlichkeitseinschätzungen führt. Überdies ergab sich ein Kontexteffekt der Ankerfrage mit Häufigkeitsskala auf die Zielfrage, wobei die subjektive Entdeckungswahrscheinlichkeit hier kleiner eingeschätzt wurde als bei der Ankerfrage mit Prozentskala.

Die vorgestellten Ergebnisse können für die empirische Sozialforschung und die Umfragemethodologie von Nutzen sein: Erstens können sie in der Phase der Fragebogenentwicklung bzw. der Konstruktion von Messinstrumenten zur Erfassung und Kommunikation allgemeiner sowie subjektiver Entdeckungswahrscheinlichkeiten Anregungen liefern. Zweitens wurde eine Alternative zu den häufig verwendeten *Vague Quantifiers* bei der Erfassung von Entdeckungswahrscheinlichkeiten aufgezeigt. Drittens wurde argumentiert, dass bei der Informationspräsentation sowie Abfrage von Entdeckungswahrscheinlichkeiten in numerischen Formaten die Nutzung von natürlichen Häufigkeiten intuitiver und verständlicher sei und die subjektive Überschätzung von kleinen Wahrscheinlichkeiten abnehme, wenn Befragte Risiken in Häufigkeiten anstatt Prozenten schätzen. Dies könnte zunächst als Nachteil gedeutet werden, wenn man davon ausgeht, dass viele Umfragen durchgeführt werden, um aus den Antworten der Stichprobe auf Wahrnehmungen in der Population zu schließen. So wäre denkbar, dass Häufigkeitsformate die

Wahrnehmungen und Antworten in der Stichprobe derart beeinflussen, dass diese von den spontanen Wahrnehmungen der Personen in der Population abweichen und dadurch als handlungsrelevante Variablen an Erklärungskraft verlieren würden. Bisherige Forschungsarbeiten deuten allerdings darauf hin, dass das Häufigkeitsformat dem alltäglichen Denken näher ist:

„Natural frequencies facilitate inferences because they carry implicit information about base rates (...). They also correspond to the way in which humans have experienced statistical information over most of their history.“ (Hoffrage et al. 2000: 2261)

Literatur

- Ajzen, I., 1977: Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology* 35: 303-314.
- ALLBUS 1990/2000: Fragebögen und Codebücher. [http://www.gesis.org/dienstleistungen/daten/umfragedaten/allbus/suche-in-fragetexten/?L= \(02.12.2008](http://www.gesis.org/dienstleistungen/daten/umfragedaten/allbus/suche-in-fragetexten/?L= (02.12.2008)).
- Bar-Hillel, M., 1980: The base-rate fallacy in probability judgments. *Acta Psychologica* 44: 211-233.
- Billiet, J. B., L. Waterplas und Geert Loosveldt, 1992: Context effects as substantive data in social surveys. S. 131-147 in: N. Schwarz und S. Sudman (Hg.): *Context effects in social and psychological research*. New York: Springer-Verlag.
- Birnbaum, M. H., 1999: How to show that $9 > 221$. Collect judgments in a between-subjects design. *Psychological Methods* 4: 243-249.
- Black, W. C., R. F. Nease Jr. und A. N. A. Tosteson, 1995: Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *Journal of the National Cancer Institute* 87: 720-731.
- Bradburn, N. M. und C. Miles., 1979: Vague quantifiers. *Public Opinion Quarterly* 43: 92-101.
- Brase, G. L., 2002: Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making* 15: 381-401.
- Brase, G. L., L. Cosmides und J. Tooby, 1998: Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology General* 127: 3-21.
- Bruine de Bruin, W., B. Fischhoff, S. G. Millstein und B. L. Halpern-Felsher, 2000: Verbal and numerical expressions of probability: "It's a fifty-fifty chance". *Organizational Behavior and Human Decision Processes* 81: 115-131.
- Carlston, D. E., 1980: The recall and use of traits and events in social inference processes. *Journal of Experimental Social Psychology* 16: 303-328.
- Cosmides, L. und J. Tooby, 1996: Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58: 1-73.
- Coutts-Heller, E., 2002: Context effects in the measurement of subjective probabilities in surveys. Universität Konstanz: Diplomarbeit.
- Dolinski, D., W. Gromski und E. Zawisza, 1987: Unrealistic pessimism. *Journal of Social Psychology* 127: 511-516.
- Dominitz, J. und C. F. Manski, 1997: Perceptions of economic insecurity. Evidence from the survey of economic expectations. *Public Opinion Quarterly* 61: 261-287.

- Esser, H., 1986: Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 314-336.
- Evans, J. S., S. J. Handley, N. Perham, D. E. Over und V. A. Thompson, 2000: Frequency versus probability formats in statistical word problems. *Cognition* 77: 197-213.
- Feldman, J. M., 1992: Constructive processes as a source of context effects in survey research. Explorations in self-generated validity. S. 49-61 in: N. Schwarz und S. Sudman (Hg.): *Context effects in social and psychological research*. New York: Springer-Verlag.
- Fischhoff, B., A. M. Parker, W. B. De Bruin, J. Downs, C. Palmgren, R. Dawes und C. F. Manski, 2000: Teen expectations for significant life events. *Public Opinion Quarterly* 64: 189-205.
- Franic, D. M. und D. S. Pathak, 2000: Communicating the frequency of adverse drug reactions to female patients. *Drug Information Journal* 34: 251-272.
- Gigerenzer, G., 1996a: On narrow norms and vague heuristics. A reply to Kahneman und Tversky (1996). *Psychological Review* 103: 592-596.
- Gigerenzer, G., 1996b: The psychology of good judgment. Frequency formats and simple algorithms. *Medical Decision Making* 16: 273-280.
- Gigerenzer, G., 1998: Ecological intelligence. An adaption for frequencies. S. 9-29 in: D. D. Cummins und C. Allen (Hg.): *The evolution of mind*. New York: Oxford University Press.
- Gigerenzer, G., 2000: *Adaptive thinking. Rationality in the real world*. New York: Oxford University Press.
- Gigerenzer, G., U. Hoffrage und H. Kleinbölting, 1991: Probabilistic mental models. A Brunswikian theory of confidence. *Psychological Review* 98: 506-528.
- Gigerenzer, G. und U. Hoffrage, 1995: How to improve Bayesian reasoning without instruction. Frequency formats. *Psychological Review* 102: 684-704.
- Gigerenzer, G. und U. Hoffrage, 1999: Overcoming difficulties in Bayesian reasoning. A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review* 106: 425-430.
- Gigerenzer, G., W. Hell und H. Blank, 1988: Presentation and content. The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance* 14: 513-525.
- Ginosar, Z. und Y. Trope, 1980: The effects of base rates and individuating information on judgment about another person. *Journal of Experimental Social Psychology* 16: 228-242.
- Griffin, D. und R. Buehler, 1999: Frequency, probability, and prediction. Easy solutions to cognitive illusions? *Cognitive Psychology* 38: 48-78.
- Hendrickx, L., C. Vlek und H. Oppewal, 1989: Relative importance of scenario information and frequency information in the judgment of risk. *Acta Psychologica* 72: 41-63.
- Hoffrage, U., G. Gigerenzer, S. Krauss und L. Martignon, 2002: Representation facilitates reasoning. What natural frequencies are and what they are not. *Cognition* 84: 343-352.
- Hoffrage, U., S. Lindsey, R. Hertwig und G. Gigerenzer, 2000: Communicating statistical information. *Science* 290: 2261-2262.
- Hoorens, V. und B. P. Buunk, 1993: Social comparison of health risks. Locus of control, the person-positivity bias, and unrealistic optimism. *Journal of Applied Social Psychology* 23: 291-302.
- Johnson, J. E. V. und A. C. Bruce, 2001: Calibration of subjective probability judgments in a naturalistic setting. *Organizational Behavior and Human Decision Processes* 85: 265-290.
- Jones, S. K., K. T. Jones und D. Frisch, 1995: Biases of probability assessment. A comparison of frequency and single-case judgments. *Organizational Behavior and Human Decision Processes* 61: 109-122.
- Kahneman, D. und A. Tversky, 1972: Subjective probability. A judgment of representativeness. *Cognitive Psychology* 3: 430-454.
- Kahneman, D. und A. Tversky, 1973: On the psychology of prediction. *Psychological Review* 80: 237-251.
- Kahneman, D. und A. Tversky, 1982a: On the study of statistical intuitions. *Cognition* 11: 123-141.

- Kahneman, D. und A. Tversky, 1982b: Variants of uncertainty. *Cognition* 11: 143-157.
- Kahneman, D. und A. Tversky, 1996: Theoretical notes on the reality of cognitive illusions. *Psychological Review* 103: 582-591.
- Klar, Y., A. Medding und D. Sarel, 1996: Nonunique Invulnerability. Singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. *Organizational Behavior and Human Decision Processes* 67: 229-245.
- Knowles, E. S., M. C. Coker, D. A. Cook, S. R. Diercks, M. E. Irwin, E. J. Lundeen, J. W. Neville und M. E. Sibicky, 1992: Order effects within personality measures. S. 221-236 in: N. Schwarz und S. Sudman (Hg.): *Context effects in social and psychological research*. New York: Springer-Verlag.
- Koehler, J. J., 1996: The base rate fallacy reconsidered. Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences* 19: 1-53.
- Koehler, J. J., 2001: When are people persuaded by DNA match statistics? *Law and Human Behavior* 25: 493-513.
- Kreuter, F., 2002: *Kriminalitätsfurcht. Messung und methodische Probleme*. Opladen: Leske+Budrich.
- Krumpal, I., H. Rauhut, D. Böhr und E. Naumann, 2008: Wie wahrscheinlich ist „wahrscheinlich“? Zursubjektiven Einschätzung und Kommunikation von Viktimisierungswahrscheinlichkeiten. *Methoden, Daten und Analysen: Zeitschrift für empirische Sozialforschung* 2: 3-27. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.2_Heft_1/2008_MDA1_Krumpal_et_al.pdf (15.5.2009).
- Lippman-Hand, A. und F. C. Fraser, 1979: Genetic counselling. Provision and Reception of information. *American Journal of Medical Genetics* 3: 113-127.
- Mellers, B. A. und A. Peter McGraw, 1999: How to improve Bayesian reasoning. Comment on Gigerenzer and Hoffrage (1995). *Psychological Review* 106: 417-424.
- Nisbett, R. und L. Ross, 1980: *Human inference. Strategies and shortcomings of social judgment*. Englewood Cliffs: Prentice-Hall.
- Phillips, L. D., 1970: The 'true probability' problem. *Acta Psychologica* 34: 254-264.
- Pinkerton, S. D., L. I. Wagner-Raphael, C. A. Craun und P. R. Abramson, 2000: A quantitative study of the accuracy of college students' HIV risk estimates. *Journal of Applied Behavioral Research* 5: 1-25.
- Price, P. C., 1998: Effects of a relative-frequency elicitation question on likelihood judgment accuracy. The case of external correspondence. *Organizational Behavior and Human Decision Processes* 76: 277-297.
- Reeves, T. und R. S. Lockhart, 1993: Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology* 122: 207-226.
- Reuband, K.-H., 2002: Subjektive Wahrscheinlichkeiten und Antwortmuster. Der Einfluss von Personenbezug und Skalierungsart. *ZA-Information* 50: 46-58.
- Rottenstreich, Y. und A. Tversky, 1997: Unpacking, repacking, and anchoring. *Advances in support theory. Psychological Review* 104: 406-415.
- Schapira, M. M., A. B. Nattinger und C. A. McHorney, 2001: Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making* 21: 459-467.
- Schnell, R. und F. Kreuter, 2000: Das DEFECT-Projekt. Sampling-Errors und Nonsampling-Errors in komplexen Bevölkerungsstichproben. *ZUMA-Nachrichten* 47, 89-101. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_47.pdf (15.5.2009).
- Schwarz, N., 1998: Accessible content and accessibility experiences. The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review* 2: 87-99.
- Schwarz, N. und S. Sudman, 1992: *Context effects in social and psychological research*. New York: Springer-Verlag.

- Schwarz, N., F. Strack, D. J. Hilton und G. Naderer, 1991: Base-rates, representativeness, and the logic of conversation. The contextual relevance of "irrelevant" information. *Social Cognition* 9: 67-84.
- Schwarz, N., H.-J. Hippler, B. Deutsch und F. Strack, 1985: Response scales. Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly* 49: 388-395.
- Shadish, W. R., T. D. Cook und D. T. Campbell, 2002: *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sherman, S. J., 1980: On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology* 39: 211-221.
- Slovic, P., M. L. Finucane, E. Peters und D. G. MacGregor, 2004: Risk as analysis and risk as feelings. Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24: 311-322.
- Slovic, P. und J. Monahan, 1995: Probability, danger, and coercion. A study of risk perception and decision making in Mental Health Law. *Law and Human Behavior* 19: 49-65.
- Strack, F., 1992: „Order effects" in survey research. Activation and information functions of preceding questions. S. 23-34 in: N. Schwarz und S. Sudman (Hg.): *Context effects in social and psychological research*. New York: Springer-Verlag.
- Teigen, K. H., 1974: Overestimation of subjective probabilities. *Scandinavian Journal of Psychology* 15: 56-62.
- Tourangeau, R., 1992: Context effects on responses to attitude questions. Attitudes as memory structures. S. 35-47 in: N. Schwarz und S. Sudman (Hg.): *Context effects in social and psychological research*. New York: Springer-Verlag.
- Tourangeau, R., L. J. Rips und K. Rasinski, 2000: *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tversky, A. und D. Kahneman, 1973: Availability. A heuristic for judging frequency and probability. *Cognitive Psychology* 5: 207-232.
- Tversky, A. und D. Kahneman, 1974: Judgment under uncertainty. Heuristics and biases. *Science* 185: 1124-1131.
- Tversky, A. und D. Kahneman, 1983: Extensional versus intuitive reasoning. The conjunction fallacy in probability judgment. *Psychological Review* 90: 293-315.
- Warr, M., 1980: The accuracy of public beliefs about crime. *Social Forces* 59: 456-470.
- Windschitl, P. D., 2002: Judging the accuracy of a likelihood judgment. The case of smoking risk. *Journal of Behavioral Decision Making* 15: 19-35.
- Wright, D. B., G. D. Gaskell und C. A. O'Muircheartaigh, 1994: How much is "Quite a bit"? Mapping between numerical values and vague quantifiers. *Applied Cognitive Psychology* 8: 479-496.
- Yamagishi, K., 1997: Upward versus downward anchoring in frequency judgments of social facts. *Japanese Psychological Research* 39: 124-129.

Korrespondenzadresse: Ivar Krumpal
Universität Leipzig
Institut für Soziologie
Beethovenstraße 15
04107 Leipzig
krumpal@sozio.uni-leipzig.de

Anhang Empirische Häufigkeitsverteilungen der allgemeinen und subjektiven Entdeckungswahrscheinlichkeiten (absolute Häufigkeiten)

Wahrscheinlichkeiten (Wertebereich)	Allgemeine Entdeckungs- wahrscheinlichkeit (Häufigkeitsformat)	Subjektive Entdeckungs- wahrscheinlichkeit (Häufigkeitsformat)	Allgemeine Entdeckungs- wahrscheinlichkeit (Prozentformat)	Subjektive Entdeckungs- wahrscheinlichkeit (Prozentformat)
0	1	1	0	2
1-4	26	17	6	5
5	29	27	12	15
6-9	8	4	3	2
10	29	25	24	22
11-14	2	0	1	1
15	13	8	6	6
16-19	1	0	0	0
20	30	24	15	15
21-24	0	0	1	1
25	6	13	9	10
26-29	1	0	1	1
30	19	16	19	12
31-34	2	1	3	2
35	6	2	2	2
36-39	0	0	2	1
40	7	6	12	9
41-44	0	0	0	0
45	2	0	1	0
46-49	1	1	0	0
50	11	23	24	20
51-54	0	0	0	0
55	0	0	1	0
56-59	1	1	0	1
60	6	4	9	9
61-64	0	0	0	0
65	2	0	4	0
66-69	0	0	0	0
70	6	13	10	11
71-74	0	0	0	0
75	2	4	5	4
76-79	0	0	0	0
80	4	12	7	10
81-84	0	0	0	0
85	0	1	1	3
86-89	0	0	0	0
90	1	6	2	5
91-94	0	0	0	0
95	1	2	0	3
96-99	0	0	0	1
100	0	6	0	7
N	217	217	180	180
25 % Quantil	5	10	12	10
50 % Quantil	17	25	30	30
75 % Quantil	30	50	50	60