

**Experimental Studies in
Decision Making and Management Control**

DISSERTATION



zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften an der Universität Passau

vorgelegt von
Dipl.-Kfm. Christian Meier

Erstgutachter: Prof. Dr. Robert Obermaier

Zweitgutachterin: Prof. Dr. Marina Fiedler

Dezember 2014

Tag der Disputation: 20. Januar 2016

Erstgutachter: Prof. Dr. Robert Obermaier

Zweitgutachterin: Prof. Dr. Marina Fiedler

Acknowledgements

The process of writing a cumulative dissertation is a long journey. From the first seed of thought to the uncountable ideas and rewritings to a state of work, which may be named “publishable”, it takes years, and I am grateful to many people who supported me.

First of all, I would like to thank my supervisor and co-author, Robert Obermaier, for his continuous support, encouragement and patience. He gave me the freedom I needed to develop my own ideas and he was always there to talk about them. His analytical rigor and his critical views were essential for developing my own research skills and shaping my ideas. I would also like to thank Marina Fiedler for her support as my second assessor and Markus Diller for chairing the doctoral committee.

I would like to thank my co-authors Franziska Himml, Tamara Jakob, and Christine Selbitschka. They constantly supported me in designing, programming, and conducting the experiments of three studies of this thesis. I would also like to thank all the conference and workshop participants during the last years for their helpful comments and suggestions.

I would like to thank my colleagues at the Chair of Accounting and Control of the University of Passau for their constant support and the enjoyable work environment during the last four years. I will miss the discussions on management accounting issues, the lunches at the Greek restaurant and especially the frequent exercise sessions in the gym. I am especially grateful to Markus Dirmhirn, Ulrike Haberl, Florian Kaiser, Andreas Neumeier, Josef Schosser, and Felix Weißmüller for their help, suggestions, and cooperation in proofreading the studies of this dissertation as well as their support in preparing and conducting the experiments.

My very special thanks go to my parents, Antonie and Peter, my brother Alexander, and all of my in-laws. My thanks also go out to my late grandmothers Notburga and Waltraud. Their love and their faith in me is what have shaped me to be the person I am today. Thank you for everything.

Finally, and most importantly, I would like to thank my wife Marina for making me a better person without changing me into someone other than myself. Words cannot express my gratitude. □

Content

Preface	I
“Trust the Numbers!?” – The Relation of Information Processing and Decision Quality for a Complex Multi-Criteria Decision Problem (in German)	1
The Effects of Information Aggregation and Visualization on Judgment Quality for Complex Multiattribute Judgment Tasks in Performance Evaluation	48
Effort Allocation in Multi-Task Environments – The Interplay of Performance Measure Characteristics and Task Difficulty	94
Budget Negotiation Structure, Fairness, and Agreement – An Experimental Analysis of Performance, Perception of Performance, and Reputation	138

Preface

Management accounting information is used in organizations to facilitate decision making and to influence actions for management control (Zimmerman 2013).¹ Decisions and actions include the allocation of resources within a firm, coordination across organizational units, costing, pricing, compensation, and incentives. There are many ways how management accounting information is provided, including performance measurement, budgeting, capital budgeting, valuation, inventory systems, product-costing systems, and transfer pricing systems (Lambert 2001, 2007).

Management accounting systems should supply information that enhances individuals' abilities to make decisions and supports them to achieve organizational goals and objectives (Caplan 1988, Horngren et al. 2010). After making decisions, employees must implement them in an organization. However, just because superiors announce a decision does not automatically ensure that employees will implement it, as it is a common assumption in economic theory that employees maximize their own utility, whereas owners of a firm generally want to maximize firm value (Zimmerman 2013). This can lead to conflicts between employees and firm owners, when they have different goals (Ross 1973, Jensen and Meckling 1976). Therefore, management accounting systems should provide information that helps align the interests of employees with the interests of owners by directing an employee's effort to activities that increase firm value (Atkinson et al. 1997, Lambert 2001). Consequently, management accounting systems have two fundamental functions in an organization: (1) provision of information necessary for making decisions and planning (*decision making*), and (2) monitoring and motivating individuals in organizations (*management control*) (Merchant and van der Stede 2012, Zimmerman 2013). Accordingly, these two functions of management accounting information have been referred to as the decision-facilitating function and the decision-influencing function (Demski and Feltham 1976, Sprinkle and Williamson 2007).

¹ The concept referred to by the term "management accounting", which has found wide adoption internationally, might not fully correspond with the "controlling" concept, which prevails in German-speaking countries (Wagenhofer 2006, Obermaier and Müller 2008). While both concepts coincide to a large extent, they might differ primarily in that "controlling" (or the controllers') tasks more frequently refer to problem-solving and system-building activities, while management accounting might have a larger focus on scorecard-keeping and system-coupling activities (Obermaier and Müller 2008). For the remainder of this thesis, the term "management accounting" is employed to address this thesis to a broader international audience.

Decision facilitating is the provision of information to reduce predecision uncertainty. Hence, the information is provided to the decision maker before he makes a decision, to help resolve some form of uncertainty in the given decision problem (Demski and Feltham 1976). Three common uses of the decision-facilitating function are: (1) the use of information to reduce ex ante uncertainty, (2) believe revision use of management accounting information, and (3) problem-solving use (Simon et al. 1954, Demski and Feltham 1976, Baiman 1982, Tiessen and Waterhouse 1983, Sprinkle and Williamson 2007, Obermaier and Müller 2008). Despite the perfect rationality assumption underlying economic agency models and most other models of economic behavior (see e.g. Baiman 1982, 1990, Christensen and Feltham 2008) extensive empirical evidence shows that both producers and users of information are boundedly rational decision makers. Their rationality is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make a decision (Simon 1957, 1982, Newell and Simon 1972, Gigerenzer and Selten 2001, Kahneman 2003). Therefore, research in management accounting is required to assess the quality (and variations in the quality) of judgment and decision making, as well as analyze modifications of the task environment to improve judgment and decision-making performance (Bonner 2008).

Decision influencing is the use of information for monitoring and motivating employees. In the decision-influencing function information is provided after a decision maker selects and performs an action. He only knows the measurement method prior to his action. Hence, performance measures are used to evaluate a decision maker's performance, with the purpose of motivating his action selections. Therefore, he is informed as to how his performance will be measured and how that measurement will affect the outcomes that are relevant to his preferences (Demski and Feltham 1976). Three common uses of the decision influencing function are: (1) use of information to reduce ex post uncertainty, (2) performance evaluation use, and (3) scorecard-keeping and attention-directing use (Simon et al. 1954, Demski and Feltham 1976, Baiman 1982, Tiessen and Waterhouse 1983, Sprinkle and Williamson 2007, Obermaier and Müller 2008). An organization's management accounting system is used for management control to monitor and motivate employees to align actions and effort with the goal of creating firm value (Baiman 1982, Young and Lewis 1995, Zimmerman 2013). Research in management accounting can be utilized to measure the extent to which management control systems motivate individuals within an organization and help re-

duce the discrepancy of interests between employees and owners. More specifically, research can examine how features of management information, accounting, and compensation systems affect incentive problems and help mitigate agency problems of moral hazard and adverse selection (Lambert 2001, 2007). Additionally, despite the self-interest assumption underlying agency models and most models of economic behavior (Baiman 1982, 1990, Christensen and Feltham 2008), empirical evidence shows that concerns for fairness and reciprocity motivate a substantial number of people (e.g. Fehr and Schmidt 2003, 2004, Falk et al. 2008). In this regard, research in management accounting can analyze the extent to which social motives and individual values interact with management control systems and how to align actions of employees in the best interest of the firm (Sprinkle and Williamson 2007).

It is important to study empirically how both, the decision-facilitating and the decision-influencing function of management accounting information affect the behavior of individuals in organizations. Especially, as these two functions can be highly interdependent, one and the same piece of information (e. g. overhead costs) can be used for an extensive cost analysis (decision facilitating) as well as a performance measure in an incentive system (decision influencing) (Obermaier and Müller 2008). Ultimately, firms consist of boundedly rational individuals and firm value is fundamentally linked to their judgments, decisions, and actions. Moreover, an organization's management accounting system is central to motivating employees and facilitating their judgment and decision making. Consequently, it is worthwhile to analyze both roles of management accounting information, and controlled laboratory experiments are a very useful scientific method for examining whether and how management accounting systems affect the behavior of individuals within an organization (Sprinkle and Williamson 2007).²

² It is frequently difficult to employ field or archival data to analyze the effects of an organization's management accounting system on the behavior of its people. Archival-empirical and field research in management accounting often has methodological and econometric problems (Ittner and Larcker 2001). Archival data may be not available or hard to obtain. Also, independent variables being examined may be flawed, because of sample-selection biases and self-selection biases. Finally, variables frequently are imprecisely measured and, hence, can contain random noise and systematic bias. Taken together, these shortcomings can threaten construct validity, internal validity, and statistical conclusion validity of archival or field research (Sprinkle and Williamson 2007). Controlled laboratory experiments can aid overcoming these problems. The experimental method exploits the precise manipulation and measurement of variables, thus, allowing the researcher to create a research setting and generate data (Sprinkle and Williamson 2007). Manipulating the independent variables and using the principle of randomization constitute an experimentalist's comparative advantage at disentangling variables that are confounded in natural settings and measuring intervening processes to draw strong causal inferences (Libby et al. 2002). Clearly designed experiments are thus useful mechanisms for effectively studying cause-effect relations (Kerlinger and Lee 2000). They allow drawing strong causal conclusions with respect to the relationship between the variables of interest and control for threats to the validity of causal inferences (Campbell and Stanley 1963, Cook and Campbell 1979, Kerlinger and Lee 2000, Sprinkle and Williamson 2007).

The first two studies of this thesis are centered on the decision-facilitating function of management accounting information. Both studies utilize experiments to examine multi-criteria decision problems, a class of problems with high theoretical and practical relevance (Wallenius et al. 2008, Dyer et al. 1992), and raise the question whether aggregation and visualization of information can improve judgment and decision-making performance.

In the first study, “Trust the Numbers!?” – The Relation of Information Processing and Decision Quality for a Complex Multi-Criteria Decision Problem”³, aggregation of information without loss of information content significantly increases procedural consistency and procedural speed for a complex multi-criteria choice problem. Although a purely graphical presentation speeds up the decision process, it reduces procedural consistency compared to a table. Contrary to current trends for combined presentation formats in management reporting (dashboards, cockpits, etc.), tabular presentation formats are more suitable for the present class of problems to ensure a consistent process.

For the second study, “The Effects of Information Aggregation and Visualization on Judgment Quality for Complex Multiattribute Judgment Tasks in Performance Evaluation”⁴, results show that information aggregation without loss of information content increases accuracy, consistency, and speed of judgments for a complex multiattribute judgment task in performance evaluation. Graphs as a presentation format, either with or without data values, result in the most accurate judgments, whereas judgments based on tables are less accurate. “Pure” graphs result in a more consistent judgment process compared to graphs with data values. However, neither pure graphs nor graphs with data values lead to a higher procedural consistency compared to tables and procedural speed is not influenced by presentation format.

The results of both studies are of interest for providers of information, e.g. designers of management information systems, as well as recipients of information, e.g. managers and auditors. In both studies, irrespective of problem class, aggregation of information increases procedural consistency and procedural speed. Furthermore, in contrast to current trends for combined presentation formats in management reporting, results show that pure presentation formats are more suitable for the respective problem class to ensure a consistent process.

³ This paper is joint work with Robert Obermaier and Franziska Himml and is accepted for publication in *Die Betriebswirtschaft* (DBW).

⁴ This paper is joint work with Robert Obermaier and Tamara Jakob.

The last two studies of this thesis delve into the decision-influencing function of management accounting information. Both studies raise the question how management accounting systems, specifically performance measurement and budgeting, can be utilized to motivate employees to act in the interest of firm owners.

The third study, “Effort Allocation in Multi-Task Environments – The Interplay of Performance Measure Characteristics and Task Difficulty” examines the interplay of performance measure characteristics, that is, sensitivity and precision, and task difficulty on effort allocation in a multi-task laboratory experiment. Results show that task difficulty does not influence the effect of sensitivity on effort allocation, for both, a fixed-wage and a performance-based compensation system. However, task difficulty moderates the effect of precision on effort allocation for both compensation systems. Results imply that the effect of sensitivity on effort allocation is reliably predictable, whereas the effect of precision on effort allocation is difficult to predict, when underlying tasks differ with respect to task difficulty. These results are of interest for the design of optimal management control, information, and compensation systems.

The fourth study is “Budget Negotiation Structure, Fairness, and Agreement – An Experimental Analysis of Performance, Perception of Performance, and Reputation”⁵. It addresses three issues associated with participative budgeting: effects of budget negotiation structure on subordinates’ performance, superiors’ perception of subordinates’ performance, as well as effects of negotiation agreement and subordinates’ performance on superiors’ assessment of subordinates’ reputation. Results show that subordinates’ performance is high, except for subordinates making the initial proposal and superiors having final authority. In this situation subordinates frame the negotiation outcome as an unfair loss and their performance decreases. Also, superiors’ perception of subordinates’ performance differs significantly from subordinates’ performance. This can be attributed to superiors evaluating subordinates’ performance based on their own preferences. Finally, for superiors’ assessment of subordinates’ reputation, a trade-off between accuracy and leniency is identified. Reputation is only positively related to subordinates’ performance, when negotiation agreement is low. This study contributes to the literature identifying how structural features of information, accounting, and compensation systems affect incentive problems, which is fundamental for the design of management control systems. Also, this study contributes to a better understanding of how social preferences affect incentive problems and assessment of reputation.

⁵ This paper is joint work with Robert Obermaier and Christine Selbitschka.

The four studies in this thesis are independent and stand-alone contributions with their own introduction and conclusion. They contain all information necessary for their understanding. As it is common in experimental research, each study consists of a theoretical part reviewing previous literature and deriving testable hypotheses from relevant theory and an empirical part describing the experimental design, results, discussion, implications, and limitations.

References

- Anthony, R. N. 1965. *Planning and control systems: a framework for analysis*. Cambridge, MA: Harvard University Press.
- Atkinson, A. A., R. D. Banker, R. S. Kaplan, and S. M. Young. 1997. *Management accounting*. Upper Saddle River, NJ: Prentice Hall.
- Baiman, S. 1982. Agency research in managerial accounting: A survey. *Journal of Accounting Literature* 1 (1): 154-213.
- Baiman, S. 1990. Agency research in managerial accounting: A second look. *Accounting, Organizations and Society* 15 (4): 341-371.
- Bonner, S. E. 2008. *Judgment And Decision Making in Accounting*. Upper Saddle River, NJ: Pearson/ Prentice Hall.
- Campbell, D. T., and J.C. Stanley. 1963. *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.
- Caplan, E. H. 1988. Behavioral accounting research-an overview. In: K. R. Ferris (Ed.): *Behavioral Accounting Research: A Critical Analysis*. Columbus, OH: Century VII Publishing Co.
- Christensen, P. O., and G. A. Feltham. 2008. *Performance evaluation. Economics of accounting*. 2. Edition. New York, NY: Springer.
- Cook, T. D., and D. T. Campbell. 1979. *Quasi-experimentation: design & analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Demski, J. S., and G. A. Feltham. 1976. *Cost determination: a conceptual approach*. Ames, IA: Iowa State University Press.
- Dyer, J. S., P. C. Fishburn, R. E. Steuer, J. Wallenius, and S. Zionts. 1992. Multiple criteria decision making, multiattribute utility theory: the next ten years. *Management Science* 38 (5): 645-654.
- Falk, A., E. Fehr, and U. Fischbacher. 2008. Testing theories of fairness—Intentions matter. *Games and Economic Behavior* 62 (1): 287-303.
- Fehr, E., and K. M. Schmidt. 2003. Theories of Fairness and Reciprocity - Evidence and Economic Applications. In: M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Ed.): *Advances in Economic Theory: Eighth World Congress of the Economic Society*. Cambridge, U.K.: Cambridge University Press.
- Fehr, E., and K. M. Schmidt. 2004. Fairness and incentives in a multi-task principal-agent model. *Scandinavian Journal of Economics* 106 (3): 453-474.
- Gigerenzer, G., and R. Selten. 2001. *Bounded rationality. The adaptive toolbox*. Cambridge, MA: MIT Press.
- Horngren, C. T., G. Foster, S. M. Datar, M. Rajan, C. Ittner, and A. A. Baldwin. 2010. Cost accounting: A managerial emphasis. *Issues in Accounting Education* 25 (4): 789-790.
- Ittner, C. D., and D. F. Larcker. 2001. Assessing empirical research in managerial accounting: a value-based management perspective. *Journal of Accounting and Economics* 32 (1): 349-410.

- Jensen, M. C., and W. H. Meckling. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3 (4): 305-360.
- Kahneman, D. 2003. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist* 58 (9): 697-720.
- Kerlinger, F. N., and H. B. Lee. 2000. *Foundations of behavioral research*. Orlando, FL: Harcourt.
- Lambert, R. A. 2001. Contracting theory and accounting. *Journal of Accounting and Economics* 32 (1): 3-87.
- Lambert, R. A. 2007. Agency Theory and Management Accounting. In: C. S. Chapman, A. G. Hopwood, and M. D. Shields (Ed.): *Handbook of Management Accounting Research 1*. Oxford, U.K.: Elsevier: 247-268.
- Libby, R., R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27 (8): 775-810.
- Merchant, K. A., and W. A. van der Stede. 2012. *Management control systems. Performance measurement, evaluation and incentives*. 3. Edition. Harlow, U.K.: Pearson, Financial Times/ Prentice Hall.
- Newell, A., and H. A. Simon. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Obermaier, R., and F. Müller. 2008. Management accounting research in the lab – method and applications. *Zeitschrift für Planung & Unternehmenssteuerung* 19 (3): 325-351.
- Ross, S. A. 1973. The Economic Theory of Agency: The Principal's Problem. *American Economic Review* 63 (2): 134-139.
- Simon, H. A. 1957. *Models of Man: Social and Rational*. New York, NY: Wiley.
- Simon, H. A. 1982. *Models of bounded rationality: behavioral economics and business organization*. Cambridge, MA: MIT Press.
- Simon, H. A., G. Kozmetsky, H. Guetzkow, and G. Tyndall. 1954. *Centralization vs. Decentralization in Organizing the Controller's Department*. New York, NY: The Controllershship Foundation.
- Sprinkle, G. B., and M. G. Williamson. 2007. Experimental Research in Managerial Accounting. In: C. S. Chapman, A. G. Hopwood, and M. D. Shields (Ed.): *Handbook of Management Accounting Research 1*. Oxford, U.K.: Elsevier: 415-444.
- Tiessen, P., and J. H. Waterhouse. 1983. Towards a descriptive theory of management accounting. *Accounting, Organizations and Society* 8 (2): 251-267.
- Wagenhofer, A. 2006. Management Accounting Research in German-Speaking Countries. *Journal of Management Accounting Research* 18 (1): 1-19.
- Wallenius, J., J. S. Dyer, P. C. Fishburn, R. E. Steuer, S. Zionts, and K. Deb. 2008. Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science* 54 (7): 1336-1349.
- Young, S. M., and B. L. Lewis. 1995. Incentive contracting research in management accounting. In: R. H. Ashton, and A. H. Ashton (Ed.): *Judgment and Decision Making Research in Accounting and Auditing*. Cambridge, U.K.: Cambridge University Press.

Zimmerman, J. 2013. Accounting for Decision Making and Control. New York, NY:
McGraw-Hill Education.

„Trust the Numbers!?“ – Zum Verhältnis von Informationsaufbereitung und Entscheidungsgüte bei komplexen multikriteriellen Entscheidungsproblemen

“Trust the Numbers!?“ – The Relation of Information Processing and Decision Quality for a Complex Multi-Criteria Decision Problem

Robert Obermaier / Christian Meier / Franziska Himml[✉]

Zusammenfassung

Die vorliegende Studie überprüft experimentell, ob Entscheidungsprozess und -ergebnis eines komplexen multikriteriellen Entscheidungsproblems unter Sicherheit durch Aggregation und Visualisierung entscheidungsrelevanter Information verbessert werden können. Informationsaggregation erhöht signifikant Konsistenz und Geschwindigkeit des Entscheidungsprozesses. Visualisierung in Form einer rein graphischen Darstellung beschleunigt zwar den Prozess, verringert aber dessen Konsistenz. Entgegen aktueller Trends zu kombinierten Darstellungsformaten, bspw. im betrieblichen Berichtswesen (Dashboards, Cockpits etc.), sind rein tabellarische Darstellungsformate besser für die vorliegende Problemklasse geeignet, um einen konsistenten Entscheidungsprozess zu gewährleisten.

Abstract

This study experimentally examines whether decision process and result quality for a complex multi-criteria decision problem under certainty can be improved through aggregation and visualization of crucial information. Aggregation significantly increases procedural consistency and procedural speed. Although a purely graphical presentation speeds up the process, it reduces procedural consistency. Contrary to current trends for combined presentation formats (dashboards, cockpits, etc.), for instance in management reporting, tabular presentation formats are more suitable for the present class of problems to ensure a consistent process.

Stichwörter: Multikriterielles Entscheidungsproblem, Laborexperiment, Informationsaufbereitung, Visualisierung, Prozedurale Rationalität.

Keywords: Multi-criteria decision problem, laboratory experiment, information processing, visualization, procedural rationality.

✉ Univ.-Prof. Dr. *Robert Obermaier*, Dipl.-Kfm. *Christian Meier*, M. Sc. *Franziska Himml*, Universität Passau, Lehrstuhl für Betriebswirtschaftslehre mit Schwerpunkt Accounting und Controlling, Innstraße 27, 94032 Passau, E-Mail: controlling@uni-passau.de.

1 Problemstellung

Komplexe mehrkriterielle Entscheidungsprobleme sind eine Problemklasse mit hoher theoretischer und praktischer Relevanz (vgl. Wallenius et al., 2008; Dyer et al., 1992). Sie können durch menschliches Urteilsvermögen intuitiv oder durch Rückgriff auf normative Entscheidungsmodelle gelöst werden (vgl. Stock/Watson, 1984, S. 192 f.; Libby, 1981, S. 105). Normative Entscheidungsmodelle gehen davon aus, dass mehr Information per se zu besseren Entscheidungen führt. Demgegenüber wies Simon (1957) mit dem Begriff der „bounded rationality“ bereits früh darauf hin, dass limitierte Ressourcen, limitierte kognitive Fähigkeiten sowie habituelles Entscheidungsverhalten objektiv-rationale Entscheidungen von Individuen praktisch unmöglich machen.¹

Auf den ersten Blick mag es daher nahe liegen, die Güte einer Entscheidung an einer objektiven Ergebnisqualität festzumachen, d.h. empirisch zu überprüfen, ob sich das Ergebnis eines Entscheidungssubjekts mit einem objektiv-rationalem Ergebnis im Sinne einer (quasi-) dominanten Handlungsalternative deckt. Allerdings kann die ausschließliche Betrachtung der Ergebnisperspektive kein ausreichendes Kriterium zur Ermittlung einer umfassend verstandenen Entscheidungsgüte sein, wenn z.B. im Fall mehrkriterieller Probleme aus methodischen Gründen auf die mitunter artifizielle Konstruktion einer dominanten Alternative als Beurteilungsmaßstab verzichtet werden soll oder subjektive Artenpräferenzrelationen im Entscheidungskalkül explizit Berücksichtigung finden sollen.

Ein weiteres Problem ist, dass bei einer relativ hohen Anzahl an Alternativen oder Kriterien die Informationsverarbeitungskapazität eines Entscheiders rasch überschritten werden kann (vgl. Driver/Streifert, 1969, S. 274 f.; Schroder et al., 1967, S. 36-39). Abweichungen von einem rationalen Ergebnis sind die mögliche Folge.

Die Entscheidungstheorie greift daher auf das Konzept der prozeduralen Rationalität zurück, um einen weiteren Maßstab zur Beurteilung der Entscheidungsgüte festzulegen (vgl. Bamberg et al., 2012, S. 3 f., Eisenführ et al., 2010, S. 5, Obermaier/Saliger, 2013, S. 243 ff.). Rationales Verhalten setzt dann voraus, dass bereits bei der Aufnahme und Verarbeitung von Information im Entscheidungsprozess Anforderungen an die Konsis-

¹ Weitergehende Einblicke in die Ursachen für „bounded rationality“ geben die Arbeiten von Tversky/Kahneman (1974), die als erste auf eine Reihe von systematischen Rationalitätsdefiziten (biases) und die dafür ursächlichen spezifischen Kontextbedingungen hinwiesen. Aktuelle Forschungsbemühungen gehen so weit, zu fragen, unter welchen Bedingungen Heuristiken und habituelles Entscheidungsverhalten nicht sogar zu „besseren“ Entscheidungen als der klassische Rationalkalkül führen können (vgl. u.a. Gigerenzer/Gaissmaier, 2011).

tenz der Präferenzen des Entscheiders gestellt werden. Für die Sicherstellung eines rationalen Entscheidungsprozesses und -ergebnisses ist es daher von fundamentalem Interesse, ob und wie die Informationsverarbeitungskapazität durch die Art der Informationsaufbereitung beeinflusst wird. Neben der Aggregation von komplexen Informationen wird insbesondere Visualisierung als Möglichkeit betrachtet, die Güte von Entscheidungen zu verbessern (vgl. u.a. Tufte, 2001; DeSanctis, 1984).

Studien im Rahmen mehrkriterieller Entscheidungsprobleme belegen, dass das Darstellungsformat in Abhängigkeit vom Aufgabentyp die objektive Ergebnisqualität beeinflussen kann (vgl. u.a. Bassler, 2010; So/Smith, 2004; Speier/Morris, 2003; Denis/Carte, 1998; Smelcer/Carmel, 1997; Wright, 1995; Nibbelin et al., 1992; Benbasat/Dexter, 1986). Weitere Studien im mehrkriteriellen Kontext zeigen, dass Informationsaggregation die objektive Ergebnisqualität verbessern kann (vgl. u.a. Otley/Dias, 1982; Harvey et al., 1979; Tiessen, 1976; Chervany/Dickson, 1974). Jüngste Studien aus dem Bereich von Investitionen in Finanzanlagen zeigen ferner, dass Informationsaggregation zu verändertem Risikoverhalten und auch zu besserem Risikoverständnis führen kann (vgl. Anagol/Gamble 2013; Beshears et al., 2013; Kaufmann/Weber, 2013).

Dabei fällt auf, dass bestehende empirische Studien im hier relevanten Bereich komplexer mehrkriterieller Entscheidungsprobleme stets eine objektive Ergebnisqualität in Abhängigkeit von Informationsaggregation und Darstellungsformat untersuchen. Damit einher geht die nicht unproblematische Annahme einer quasi-dominanten, d.h. objektiv-optimalen Alternative, oder die Vorgabe von Zielgewichten. Hingegen existiert bisher keine Studie, die explizit eine auf subjektiven Präferenzen basierende Ergebnistransitivität und prozedurale Konsistenz zur Beurteilung der Entscheidungsgüte heranzieht. Auch Wechselwirkungen von Informationsaggregation und Darstellungsformat und deren Effekte auf die Entscheidungsgüte wurden bislang nicht ausreichend untersucht.

Im Fokus der vorliegenden Arbeit stehen daher die Auswirkungen unterschiedlicher Formen von Informationsaufbereitung auf die Rationalität komplexer multikriterieller Entscheidungen unter Sicherheit, ohne die Annahme einer dominanten Handlungsalternative und unter expliziter Berücksichtigung subjektiver Präferenzrelationen. Unter Informationsaufbereitung wird dabei jener Teilbereich der Informationsversorgung verstanden, der sich mit der Auswahl, Aggregation und Darstellung von Informationen

zum Zweck der Entscheidungsvorbereitung beschäftigt. Konkret sollen folgende Forschungsfragen untersucht werden:

(1) Kann der Entscheidungsprozess und das Ergebnis durch die visuelle Darstellung von quantitativen Informationen verbessert werden?

(2) Beeinflusst der Aggregationsgrad der dem Entscheidungsträger zur Verfügung stehenden Information die Güte des Entscheidungsprozesses und des Ergebnisses?

Zur Beantwortung dieser Fragen dient ein Laborexperiment mit den unabhängigen Variablen Darstellungsformat und Informationsaggregation. Die Hypothesen werden auf Basis der Cognitive Fit-Theorie von Vessey (2006, 1994, 1991) und anhand der Theorie der menschlichen Informationsverarbeitung (vgl. Driver/Streufert, 1969; Schroder et al., 1967) abgeleitet.

Die vorliegende Studie liefert einen Beitrag zur Untersuchung der Auswirkungen von Informationsaufbereitung im Kontext eines komplexen multikriteriellen Entscheidungsproblems unter Berücksichtigung subjektiver Artenpräferenzrelationen. Im Ergebnis führt eine rein tabellarische Darstellung zur höchsten prozeduralen Konsistenz, während eine rein graphische Darstellung lediglich zur höchsten prozeduralen Geschwindigkeit führt. Informationsaggregation verbessert prozedurale Konsistenz und prozedurale Geschwindigkeit. Die subjektive Ergebnistransitivität kann hingegen weder durch Visualisierung noch durch Informationsaggregation erhöht werden. Die Analyse des Darstellungsformaten bei variierender Informationsaggregation zeigt, dass eine Tabelle und eine Kombination aus Tabelle und Graphik bei hoher Informationsaggregation mit höherer prozeduraler Konsistenz und Geschwindigkeit einhergehen als bei niedriger Informationsaggregation. Bei rein graphischer Darstellung wird durch Informationsaggregation lediglich die prozedurale Geschwindigkeit verbessert.

Diese Studie geht dabei auch methodisch über bestehende Studien hinaus, indem sie auf den Analytischen Hierarchieprozess (vgl. Saaty, 1994) zurückgreift und damit erstmals die Offenlegung und Integration subjektiver Artenpräferenzrelationen im Rahmen der Entscheidungsanalyse ermöglicht. Erst diese Herangehensweise erlaubt gerade bei Nichtvorliegen einer dominanten Alternative die Analyse nicht-trivialer Entscheidungsprobleme aus Prozessperspektive (konsistente Artenpräferenzen) und aus Ergebnisperspektive (transitive Präferenzfunktionen). Darüber hinaus wird die bestehende Literatur im Rahmen der vorliegenden Problemklasse ergänzt durch die Untersuchung der Wechselwirkungen von Informationsaggregation und Darstellungsformat und deren Auswir-

kungen auf prozedurale Konsistenz und subjektive Ergebnistransitivität. Neben reinen Darstellungsformaten werden zudem die Effekte einer Kombination aus Tabelle und Graphik analysiert. Schließlich kann diese Studie darlegen, wie Informationsaufbereitung Entscheidungsprozess und -ergebnis sowohl positiv als auch negativ beeinflussen kann, um so Handlungsempfehlungen zur Nutzung von Aggregation und Visualisierung abzuleiten.

Der Beitrag ist wie folgt aufgebaut: In Kapitel 2 werden Hypothesen zur Wirkung von Darstellung und Informationsaggregation auf die Güte von Entscheidungen entwickelt, die im Rahmen des in Kapitel 3 darzustellenden Laborexperimentes überprüft werden. Kapitel 4 präsentiert die Ergebnisse. In Kapitel 5 folgen die Diskussion und Limitationen. Kapitel 6 schließt mit einem Fazit.

2 Theorie und Hypothesen

2.1 Kriterien zur Beurteilung der Entscheidungsgüte

Die Beurteilung der Rationalität komplexer Entscheidungen ist selbst ein komplexes Problem, das sowohl aus Prozess- als auch aus Ergebnisperspektive betrachtet werden kann. Zur Messung der Güte einer Entscheidung erfordern beide Perspektiven jeweils einen als rational zu bezeichnenden Maßstab, mit dem sowohl der Entscheidungsprozess als auch das Entscheidungsergebnis eines begrenzt rationalen Entscheiders verglichen werden können (vgl. Bonner, 2008, S. 29; Simon, 1979, S. 507-509; Simon, 1976, S. 130-132).

In laborexperimentellen Studien zur Beurteilung der Entscheidungsgüte wird daher regelmäßig eine (quasi-) dominante Handlungsalternative verwendet (vgl. u.a. Hirsch/Volnhals, 2012; Bassler, 2010; Chan, 2001; Dennis/Carste, 1998; Smelcer/Carmel, 1997; Benbasat/Dexter, 1986; Otley/Dias, 1982; Casey, 1980; Tiessen, 1976; Chervany/Dickson, 1974; Barefield, 1972; Ronen, 1971). Die Entscheidungsgüte wird mithin danach beurteilt, ob und wie das Ergebnis des faktischen Entscheidungsverhaltens von dieser objektiv-dominanten Alternative abweicht. Das entsprechende Rationalitätskriterium wird meist als (objektive) *Entscheidungsqualität* bezeichnet (vgl. Bonner, 2008, S. 30).

Bei einem komplexen multikriteriellen Entscheidungsproblem unter Sicherheit kann indes nur dann sinnvoll auf die objektive Entscheidungsqualität abgestellt werden, wenn eine solche dominante Alternative existiert und die Zielgewichte der Attribute vorab

festgelegt sind. Das vermeintlich komplexe Entscheidungsproblem degeneriert damit aber zu einem mehr oder weniger aufwendigen „Rechenproblem“. Existiert de facto hingegen keine dominante Alternative und sind die Zielgewichte an subjektive Artenpräferenzrelationen geknüpft, was der empirische Regelfall sein dürfte, muss zunächst individuell ermittelt werden, wie hoch der Grad der Abweichung der tatsächlich beobachtbaren (faktischen) Entscheidung eines Individuums von der Entscheidung eines (fiktiv) rationalen Entscheidungsträgers ist. Das Ergebnis einer derart fiktiv-rationalen Entscheidung setzt die Aufdeckung der Präferenzen des Entscheidungsträgers unter der Annahme von Rationalität voraus und kann unter Rückgriff auf ausgewählte multikriterielle Entscheidungsverfahren (z.B. den Analytic Hierarchy Process) gewonnen werden. Diesen Weg wählt die vorliegende Studie. Die hierzu erforderlichen Minimalanforderungen an das Entscheidungsmitglied hinsichtlich der Bildung von konsistenten Präferenzvorstellungen über die Alternativen werden durch das Ordnungs- und das Transitivitätsaxiom beschrieben (vgl. die klassischen Arbeiten von Luce/Raiffa, 1957 und von Von Neumann/Morgenstern, 1947, 1944). Der Vergleich von faktischer und fiktiv-rationaler Entscheidung erlaubt schließlich die Beurteilung der *Ergebnistransitivität*.

Mit Blick auf den Entscheidungsprozess erschweren vor allem Zielkonflikte die Entscheidungsfindung. Im Kontext kompensatorischer Entscheidungsverfahren sind konsistente Zielgewichte bedeutsam, da sie Artenpräferenzrelationen widerspiegeln. Zur Herstellung prozeduraler Rationalität ist daher sicherzustellen, dass Entscheidungsmitglieder auf konsistente Artenpräferenzrelationen zurückgreifen, die unter bestimmten Bedingungen die Grenzrate der Substitution zwischen den Attributen angeben (vgl. Bamberg et al., 2012, S. 52 f.). Dieses Rationalitätskriterium wird im Folgenden als *prozedurale Konsistenz* bezeichnet.

Eine weitere Anforderung an die prozedurale Rationalität ist ein angemessener Informationsaufwand. Zeit ist knapp, da sie stets alternativ verwendet werden kann. Wer alle zur Verfügung stehenden Alternativen einer Entscheidung erkunden und dann ermitteln will, welche von den Alternativen die Beste im Hinblick auf ein Ziel oder mehrere Ziele ist, braucht Zeit. Ein Entscheider der diesen Zeitaufwand mit den Opportunitätskosten des knappen Faktors Zeit bewertet, wird erkennen, dass es rational sein kann, den schnelleren Entscheidungsprozess zu präferieren, wenn sich zwei Entscheidungsprozesse c. p. nur in der Zeitspanne unterscheiden. Dieses Rationalitätskriterium wird als *prozedurale Geschwindigkeit* bezeichnet.

Grundthese der vorliegenden Arbeit ist, dass die Informationsaufbereitung und damit insbesondere die Auswahl, Aggregation und Darstellung von Informationen die Fähigkeit von Entscheidungsträgern zur Lösung von Entscheidungsproblemen determiniert und infolgedessen die Entscheidungsgüte beeinflusst. Dabei soll die Entscheidungsgüte jeweils danach beurteilt werden, ob bei einem vorliegenden multikriteriellen Entscheidungsproblem ein hoher Grad an prozeduraler Konsistenz, prozeduraler Geschwindigkeit und Ergebnistransitivität vorliegt.

2.2 Entscheidungsgüte und Darstellungsformat

2.2.1 Die Theorie des Cognitive Fit

Grundgedanke der Cognitive Fit-Theorie ist, dass Kompatibilität von Problemtyp und Problemrepräsentation den Entscheider befähigt, Informationsverarbeitungsprozesse zu nutzen, die hinsichtlich Problemrepräsentation und -typ auf die gleiche Art von Information zurückgreifen. Diese Übereinstimmung wird als „Cognitive Fit“ bezeichnet und erzeugt eine konsistente mentale Repräsentation, die den Entscheidungsprozess unterstützt und zu einer präzisen und schnellen Problemlösung führt. Stimmen Problemrepräsentation und -typ nicht überein, verarbeitet der Entscheider Problemrepräsentation und -typ mit unterschiedlichen Informationsverarbeitungsprozessen. Die mentale Repräsentation muss dann transformiert werden, weswegen die Problemlösung mit einer höheren kognitiven Anstrengung einhergeht und mit hoher Wahrscheinlichkeit unpräziser und zeitaufwendiger ist (vgl. Vessey 2006, 1994, 1991).

Zentrale Annahme ist, dass die Eigenschaften der Problemrepräsentation (z.B. Darstellung in Form von Tabelle oder Graphik) und der Typ des Problems (z.B. Beschaffung oder Auswertung von Information) unabhängig voneinander sind, sodass die Problemlösung mit unterschiedlichen Darstellungsformaten erreicht werden kann, wobei tabellarische und graphische Darstellung auf den gleichen Ausgangsdaten basieren müssen (vgl. Vessey, 1991, S. 224).

Die Cognitive Fit-Theorie kann auf Problemrepräsentationen in Form von Tabellen und Graphiken angewendet werden. Graphiken stellen räumlich miteinander verbundene Information dar und werden im Rahmen der Cognitive Fit-Theorie als räumliches Darstellungsformat eingeordnet. Graphiken bringen Zusammenhänge zwischen den Daten zur Geltung, stellen jedoch diskrete Datenwerte nicht unmittelbar dar. Sie ermöglichen eine überblicksartige Betrachtung der Information ohne einzelne Datenelemente analytisch zu adressieren. Die Information in einer Graphik wird mit wahrnehmenden

Informationsverarbeitungsprozessen abgerufen (vgl. Vessey 1994, 1991). Wahrnehmende Informationsverarbeitungsprozesse beinhalten die visuelle Erfassung eines Objekts und die Ableitung von dessen Bedeutung (vgl. Bolles, 1991, S. 152 f.).

Tabellen stellen symbolische Information dar und werden als symbolisches Darstellungsformat eingeordnet. Tabellen unterstützen die Extraktion und Weiterverarbeitung spezifischer Datenwerte, stellen aber Zusammenhänge zwischen den Daten nicht unmittelbar dar. Die Information in einer Tabelle wird mit analytischen Informationsverarbeitungsprozessen abgerufen (vgl. Vessey 1994, 1991). Analytische Informationsverarbeitungsprozesse gehen einher mit kognitiver Anstrengung, um ein zu betrachtendes Phänomen zu verstehen und zu bewerten (vgl. Speier, 2006, S. 1116; Simon/Lea, 1974, S. 108-111). Wahrnehmende Prozesse werden effektiver im Rahmen von räumlichen Darstellungsformaten eingesetzt, während analytische Prozesse effektiver im Rahmen von symbolischen Darstellungsformaten eingesetzt werden (vgl. Vessey 1994, 1991).

Im Einklang mit Einhorn/Hogarth (1981, S. 9) teilt auch die Cognitive Fit-Theorie Entscheidungsprobleme in Informationsbeschaffungs- und Informationsauswertungsprobleme ein, die jeweils als räumlich oder symbolisch einzustufen sind und entsprechend mit wahrnehmenden oder analytischen Informationsverarbeitungsprozessen verarbeitet werden (vgl. Vessey, 2006, S. 145 f.; Larkin/Simon, 1987, S. 66). Ein Informationsbeschaffungsproblem (d.h. eine Suche nach Informationen im Umfeld der Aufgabe und im Kurzzeitgedächtnis) wird entsprechend mit Informationsbeschaffungsprozessen gelöst, die entweder (1) räumlich (d.h. räumliche Identifikation von Strukturen und Beziehungen) oder (2) symbolisch (d.h. diskrete Datenwerte werden analytisch extrahiert) sein können. Ein Informationsauswertungsproblem (d.h. die Nutzung einer Strategie, um Information zu verstehen und zu verarbeiten) wird wiederum mit Informationsauswertungsprozessen gelöst, die (1) räumlich (d.h. Assoziationen zwischen Informationen werden räumlich verarbeitet, z.B. zur Trenderkennung) oder (2) symbolisch (d.h. diskrete Datenwerte werden analytisch ausgewertet und verarbeitet, z.B. im Rahmen einer Nutzwertanalyse) sein können (vgl. Speier, 2006, S. 1116). Ein Cognitive Fit existiert immer dann, wenn räumliche Informationsbeschaffungs- und Informationsauswertungsprobleme mit räumlichen Darstellungsformaten einhergehen und symbolische Informationsbeschaffungs- und Informationsauswertungsprobleme mit symbolischen Darstellungsformaten verknüpft werden (vgl. Vessey, 1991, S. 227).

2.2.2 Die Cognitive Fit-Theorie im Rahmen komplexer multikriterieller Entscheidungsprobleme unter Sicherheit

In empirischen Studien findet die Cognitive Fit-Theorie im Rahmen von Informationsbeschaffungsproblemen und einfachen Informationsauswertungsproblemen durchweg Unterstützung (Literaturüberblicke finden sich bei Kelton et al., 2010; Speier, 2006; Vessey, 2006).

Gegenstand der vorliegenden Untersuchung ist hingegen ein komplexes multikriterielles Problem unter Sicherheit. Es erfordert, dass ein Entscheider (1) die Struktur der Kriterien und die Beziehungen zwischen den Alternativen identifiziert, um sich einen Überblick über die gegebenen Alternativen und die Ausprägungen der Kriterien verschaffen zu können und (2) die Kriterien gewichtet, bei der Auswertung und Analyse der Information numerische Datenwerte extrahiert und weiterverarbeitet, um eine optimale Alternative gemäß subjektiver Artenpräferenzrelationen bestimmen zu können.

Ein derartiges multikriterielles Entscheidungsproblem unter Sicherheit setzt sich demnach aus zwei Teilproblemen zusammen: (1) einem räumlichen Informationsbeschaffungsproblem und (2) einem gut-strukturierten symbolischen Informationsauswertungsproblem. Das symbolische Informationsauswertungsproblem ist dann als gut-strukturiert einzuordnen, wenn unter Sicherheit (1) die Zielfunktion des Entscheiders bei Kenntnis seiner Präferenzen gut spezifizierbar ist (z.B. in Form einer Zielgewichtung), (2) alle numerischen Werte bekannt sind und (3) durch die Verknüpfung der numerischen Werte mit der Zielfunktion eine optimale Lösung auf Basis eines numerischen Wertes (Präferenzfunktional) bestimmt werden kann (vgl. Simon/Newell, 1958, S. 4 f.). Die Bestimmung von Zielgewichten, deren Verknüpfung mit den zugehörigen Datenwerten und die Ermittlung einer optimalen Alternative setzen ein hohes Maß an analytischem Denkvermögen voraus. Diese Schritte sind zudem zeitaufwendig, sodass die kognitiven Ressourcen signifikant beansprucht werden. Multikriterielle Entscheidungsprobleme unter Sicherheit sind daher für einen begrenzt rationalen Entscheider hinsichtlich des Schwierigkeitsgrades generell als komplexe Probleme einzustufen (vgl. Vessey, 1994; Einhorn/Hogarth, 1981; Newell/Simon, 1972).

Im Entscheidungsprozess ist daher im Hinblick auf das symbolische Informationsauswertungsproblem zu erwarten, dass eine Tabelle, die analytische Informationsverarbeitungsprozesse unterstützt, zu konsistenteren Artenpräferenzrelationen führt als eine Graphik. Dagegen wird eine Graphik, welche die räumliche Informationsbeschaffung

unterstützt, mit Blick auf das gesamte Entscheidungsproblem zu einer höheren prozeduralen Geschwindigkeit führen (vgl. Vessey, 2006, S. 149). Aus Ergebnisperspektive erscheint es unwahrscheinlich, dass für ein komplexes multikriterielles Entscheidungsproblem unter Sicherheit ein Cognitive Fit eintritt, wenn die Darstellung rein tabellarisch (d.h. symbolisch), oder rein graphisch (d.h. räumlich) ist. In diesem Fall werden die entsprechenden zugrundeliegenden Informationsverarbeitungsprozesse entweder nur die räumliche Informationsbeschaffung oder nur die symbolische Informationsauswertung unterstützen, sodass keine Verbesserung der Ergebnistransitivität in Abhängigkeit der Art der Darstellung zu erwarten ist (vgl. Vessey, 2006, S. 148). Folglich können folgende Hypothesen formuliert werden:

***HI(a):** Bei einem komplexen multikriteriellen Problem unter Sicherheit führt eine Tabelle zu einer höheren prozeduralen Konsistenz als eine Graphik.*

***HI(b):** Bei einem komplexen multikriteriellen Problem unter Sicherheit führt eine Graphik zu einer höheren prozeduralen Geschwindigkeit als eine Tabelle.*

***HI(c):** Bei einem komplexen multikriteriellen Problem unter Sicherheit unterscheiden sich eine Tabelle und eine Graphik hinsichtlich der Ergebnistransitivität nicht.*

In bestehenden Studien wird die Cognitive Fit-Theorie im Rahmen mehrkriterieller Entscheidungsprobleme in Bezug auf prozedurale Geschwindigkeit und objektive Entscheidungsqualität überwiegend unterstützt. Dennis/Carte (1998) untersuchen bei einem komplexen, symbolischen multikriteriellen Entscheidungsproblem unter Sicherheit die Entscheidungsqualität in Abhängigkeit des Darstellungsformates. Eine Tabelle führt zu einer präziseren und eine Graphik zu einer schnelleren Problemlösung. Smelcer/Carmel (1997) hingegen analysieren komplexe, räumliche multikriterielle Entscheidungsprobleme. In deren Studie kann die prozedurale Geschwindigkeit durch eine graphische Darstellung signifikant verbessert werden. Bei Speier/Morris (2003) führt im Rahmen eines symbolischen mehrkriteriellen Problems mit hoher Komplexität eine Graphik zur höchsten Entscheidungsqualität, was für ein limitierendes Problem spricht (vgl. Vessey, 2006, S. 175). Entgegen der Cognitive Fit-Theorie führt in deren Studie eine symbolische Darstellung zur höchsten Entscheidungsgeschwindigkeit.

Die Theorie des Cognitive Fit kann auf kombinierte Darstellungsformate erweitert werden (vgl. Vessey, 1991, S. 229). Eine Kombination aus Tabelle und Graphik wäre im Rahmen eines räumlichen Problems vorteilhaft gegenüber einer rein tabellarischen Darstellung, da die zusätzliche graphische Darstellung wahrnehmende Informationsver-

arbeitungsprozesse unterstützt. Bei einem symbolischen Problem, welches analytische Informationsverarbeitung erfordert, wären hingegen keine Verbesserungen zu erwarten (vgl. Vessey, 1991, S. 229).

In Bezug auf das vorliegende komplexe multikriterielle Problem unter Sicherheit ist daher zu erwarten, dass eine kombinierte Darstellung, im Vergleich zu einer Graphik, die symbolische Informationsauswertung und damit die prozedurale Konsistenz verbessert. Im Vergleich zu einer Tabelle, die ohnehin analytische Informationsverarbeitungsprozesse unterstützt, ist keine Verbesserung zu erwarten. Hinsichtlich der prozeduralen Geschwindigkeit und der Ergebnistransitivität sollte die Kombination aus Tabelle und Graphik vorteilhaft gegenüber den reinen Darstellungsformaten sein, da sowohl die räumliche Informationsbeschaffung durch die graphische Darstellung als auch die symbolische Informationsauswertung durch die tabellarische Darstellung unterstützt werden. Folgende Hypothesen können formuliert werden:

***H2(a):** Bei einem komplexen multikriteriellen Problem unter Sicherheit führt eine Kombination aus Tabelle und Graphik zu einer höheren prozeduralen Konsistenz als eine Graphik und unterscheidet sich nicht von einer Tabelle.*

***H2(b):** Bei einem komplexen multikriteriellen Problem unter Sicherheit führt eine Kombination aus Tabelle und Graphik zu einer höheren prozeduralen Geschwindigkeit als eine Tabelle oder Graphik.*

***H2(c):** Bei einem komplexen multikriteriellen Problem unter Sicherheit führt eine Kombination aus Tabelle und Graphik zu einer höheren Ergebnistransitivität als eine Tabelle oder Graphik.*

Die bestehende empirische Evidenz zur Cognitive Fit-Theorie zeigt im Rahmen kombinierter Darstellungsformate gemischte Resultate im Hinblick auf die objektive Entscheidungsqualität auf. So führt bei Benbasat/Dexter (1986) eine Kombination aus Tabelle und Graphik bei räumlicher Informationsbeschaffung- und symbolischer Informationsauswertung zwar zu einer höheren Entscheidungsqualität als eine Graphik jedoch nur zur gleichen Entscheidungsqualität wie eine Tabelle. Ohne Zeitdruck führt das Kombinationsformat - theoriekonform - zu einer höheren Entscheidungsgeschwindigkeit als eine Tabelle. Bassler (2010) analysiert ein komplexes, mehrkriterielles symbolisches Entscheidungsproblem. Im Vergleich zu einer Tabelle führt eine Kombination aus Tabelle und Graphik zu einer höheren Entscheidungsqualität, die prozedurale Geschwindigkeit wird nicht beeinflusst. Wright (1995) untersucht ein multikriterielles,

räumliches Entscheidungsproblem mit hoher Komplexität. Entsprechend der Cognitive Fit-Theorie ist die Entscheidungsqualität bei einer Kombination aus Tabelle und Graphik höher als bei rein tabellarischer Darstellung. In der Studie von Nibbelin et al. (1992) hingegen ist bei einer mehrkriteriellen räumlichen Aufgabe die Entscheidungsqualität bei rein tabellarischer Darstellung vorteilhaft gegenüber einer Kombination aus Tabelle und Graphik. Frownfelter-Lohrke (1998) variiert im Rahmen eines mehrkriteriellen Entscheidungsproblems die Darstellung zwischen Tabelle, Graphik und Kombination aus Tabelle und Graphik. Weder bei einer räumlichen noch bei einer symbolischen Aufgabe wird die Entscheidungsqualität vom Darstellungsformat beeinflusst. Unabhängig vom Aufgabentyp führt die Graphik zu einer höheren prozeduralen Geschwindigkeit als die kombinierte Darstellung, welche wiederum schneller als die Tabelle ist. So/Smith (2004) analysieren ein mehrkriterielles räumliches Problem. In deren Studie führt eine Tabelle bei hoher Komplexität zu einer höheren Entscheidungsqualität als eine Kombination aus Tabelle und Graphik.

2.3 Entscheidungsgüte und Informationsauswahl

Die Theorie der menschlichen Informationsverarbeitung betrachtet Individuen im Rahmen von Entscheidungsproblemen als Informationsverarbeitungssysteme (vgl. Driver/Streifert, 1969; Schroder et al., 1967). Mit einem Anstieg der Informationsmenge erhöht sich zunächst die Menge an Information, die in den Problemlösungsprozess integriert wird. Sie erreicht an einem bestimmten Punkt ein Maximum und nimmt danach ab. Die optimale Informationsmenge beschreibt den Punkt, an dem das Informationsverarbeitungsniveau des Entscheiders und seine Entscheidungsgüte dieses Maximum erreicht. Wird dieser Punkt überschritten, tritt für den Entscheider eine Informationsüberlastung ein und die Informationsmenge, die in die mentale Repräsentation integriert werden kann, sinkt. Dies hat zur Folge, dass der Entscheider ab diesem Punkt nicht mehr alle zur Verfügung stehenden Informationen berücksichtigen kann und infolgedessen die Entscheidungsgüte abnimmt (vgl. Schroder et al., 1967, S. 36). Die Beziehung zwischen der Informationsmenge, die dem Entscheider zur Verfügung steht, und der Informationsmenge, die ein Entscheider in den Problemlösungsprozess integrieren kann, wird auch als invertiert u-förmiger Zusammenhang bezeichnet (vgl. Eppler/Mengis, 2004, S. 276).

Aggregation von Information stellt eine bekannte Maßnahme gegen Informationsüberlastung dar (vgl. Eppler/Mengis, 2004, S. 289). Sie kann in zwei unterschiedliche

Typen eingeteilt werden. Informationsaggregation vom Typ I reduziert zwar die Informationsmenge, aber nicht den Informationsgehalt in Bezug auf ein Entscheidungsproblem. Typ I aggregiert daher die zur Verfügung stehende Information, ändert aber nicht deren Informationsgehalt. Informationsaggregation vom Typ II hingegen reduziert sowohl Informationsmenge als auch Informationsgehalt. Wird ausgehend von disaggregierter Information die Information graduell aggregiert, steigt die Entscheidungsgüte, solange es sich um Typ I handelt, da bei konstantem Informationsgehalt die Komplexität des Entscheidungsproblems abnimmt und das Informationsverarbeitungsniveau steigt. Sobald die Aggregation als Typ II einzustufen ist und mit einer Verringerung des Informationsgehalts einhergeht, nimmt die Entscheidungsgüte ab, da die zur Problemlösung notwendige Information fehlt. Damit lässt sich der oben angesprochene, invertierte u-förmige Zusammenhang auf die Informationsaggregation übertragen (vgl. Otley/Dias, 1982, S. 172).

Um den Entscheidungsprozess und das -ergebnis im Rahmen eines komplexen multikriteriellen Entscheidungsproblems unter Sicherheit zu verbessern, ist Informationsaggregation vom Typ I erforderlich. Zunehmende Aggregation vom Typ I reduziert die Informationsmenge, der Informationsgehalt hingegen bleibt konstant. Damit sinkt die Anzahl der zu verarbeitenden Informationseinheiten, die Problemkomplexität wird reduziert und die Anforderungen an die Informationsverarbeitung sinken. Dies wiederum entlastet das Kurzzeitgedächtnis, erhöht die Konsistenz der mentalen Repräsentation, beschleunigt den Entscheidungsprozess, und verbessert die subjektive Ergebnistransitivität. Folgende Hypothesen werden formuliert:

***H3(a):** Bei einem komplexen multikriteriellen Problem unter Sicherheit erhöht Informationsaggregation bei konstantem Informationsgehalt die prozedurale Konsistenz.*

***H3(b):** Bei einem komplexen multikriteriellen Problem unter Sicherheit erhöht Informationsaggregation bei konstantem Informationsgehalt die prozedurale Geschwindigkeit.*

***H3(c):** Bei einem komplexen multikriteriellen Problem unter Sicherheit erhöht Informationsaggregation bei konstantem Informationsgehalt die Ergebnistransitivität.*

Die Studien von Ronen (1971), Barefield (1972), Chervany/Dickson (1974) und Tiessen (1976) untersuchen Informationsaggregation vom Typ I. Ronen (1971) zeigt im Rahmen einer Entscheidung zwischen alternativen Produktionsplänen, dass Informationsaggregation die Entscheidungsqualität verbessern kann. Barefield (1972) hingegen

findet bei aggregierter Information lediglich eine marginale Verbesserung der Entscheidungsqualität und eine Verringerung der Konsistenz von Entscheidungen. Bei Chervany/Dickson (1974) führt aggregierte Information zu höherer Entscheidungsqualität als Rohdaten. Tiessen (1976) zeigt ebenfalls positive Effekte eines steigenden Aggregationsgrades. Harvey et al. (1979) hingegen untersuchen Typ II und ändern in ihrer experimentellen Studie systematisch den Informationsgehalt. Im Ergebnis der Studie können Finanzanalysten Investitionsentscheidungen mit disaggregierter Information besser interpretieren als mit aggregierter Information vom Typ II. Otley/Dias (1982) untersuchen beide Arten der Informationsaggregation und können den invertiert u-förmigen Zusammenhang in Bezug auf die Entscheidungsqualität experimentell belegen. In experimentellen Studien konnte ferner gezeigt werden, dass eine Verringerung der Anzahl an Alternativen und/oder Kriterien die prozedurale Geschwindigkeit erhöht (vgl. Iselin, 1988; Casey, 1980; Jacoby et al., 1974a, 1974b).

Gemäß der Cognitive Fit-Theorie unterstützt eine Tabelle die Extraktion und Verarbeitung spezifischer Datenwerte. Zusammenhänge zwischen Daten werden nicht hergestellt. Ausgehend von Aggregation vom Typ I erhöht Disaggregation die Informationsmenge, der Informationsgehalt hingegen bleibt konstant. Tabellen besitzen keine integrativen Eigenschaften, d.h. der kognitive Aufwand der analytischen Informationsverarbeitungsprozesse erhöht sich bei der Verarbeitung der zusätzlichen Informationseinheiten. Im Rahmen des vorliegenden komplexen multikriteriellen Problems wird daher bei tabellarischer Darstellung mit Disaggregation die Geschwindigkeit des Entscheidungsprozesses abnehmen und die Fehlerhäufigkeit steigen, sodass sich die Konsistenz der mentalen Repräsentation verringert und die subjektive Ergebnistransitivität sinkt (vgl. Umanath/Vessey, 1994, S. 810). Folgende Hypothesen können formuliert werden:

H4(a): *Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Tabelle ist die prozedurale Konsistenz bei niedriger Informationsaggregation geringer als bei hoher Informationsaggregation.*

H4(b): *Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Tabelle ist die prozedurale Geschwindigkeit bei niedriger Informationsaggregation geringer als bei hoher Informationsaggregation.*

H4(c): *Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Tabelle ist die Ergebnistransitivität bei niedriger Informationsaggregation geringer als bei hoher Informationsaggregation.*

Entsprechend der Cognitive Fit-Theorie stellt eine Graphik Zusammenhänge zwischen Datenwerten her, ohne diskrete Datenwerte direkt abzubilden und ermöglicht die überblicksartige Betrachtung der Information. Graphiken besitzen integrative Eigenschaften, d.h. die wahrnehmenden Informationsverarbeitungsprozesse unterstützen den Entscheider bei der Verarbeitung der zusätzlichen Informationseinheiten bei Disaggregation. Deswegen wird zwar die Geschwindigkeit des Entscheidungsprozesses abnehmen, die Konsistenz der mentalen Repräsentation und die subjektive Ergebnistransitivität hingegen werden nur unwesentlich beeinflusst (vgl. Umanath/Vessey, 1994, S. 810). Folgende Hypothesen werden überprüft:

***H5(a):** Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Graphik ändert sich die prozedurale Konsistenz durch Informationsaggregation mit konstantem Informationsgehalt nicht.*

***H5(b):** Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Graphik ist die prozedurale Geschwindigkeit bei niedriger Informationsaggregation geringer als bei hoher Informationsaggregation.*

***H5(c):** Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Graphik ändert sich die Ergebnistransitivität durch Informationsaggregation mit konstantem Informationsgehalt nicht.*

Im Kombinationsformat unterstützt die Graphik für das vorliegende komplexe multikriterielle Entscheidungsproblem die räumliche Informationsbeschaffung, die Tabelle hingegen fördert die symbolische Informationsauswertung. Da die Tabelle, welche keine integrativen Elemente besitzt, die symbolische Informationsauswertung unterstützt, kommen die integrativen Elemente der Graphik hinsichtlich prozeduraler Konsistenz und Ergebnistransitivität nicht zum Tragen. Analog zur reinen Graphik wird daher hinsichtlich des räumlichen Informationsbeschaffungsproblems bei Disaggregation die prozedurale Geschwindigkeit infolge der Verarbeitung zusätzlicher Informationseinheiten abnehmen. Ebenfalls, analog zur rein tabellarischen Darstellung, verringert sich für die symbolische Informationsauswertung durch Disaggregation die prozedurale Konsistenz und Ergebnistransitivität (vgl. Umanath/Vessey, 1994, S. 810). Folgende Hypothesen werden formuliert:

H6(a): *Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Kombination aus Tabelle und Graphik ist die prozedurale Konsistenz bei niedriger Informationsaggregation geringer als bei hoher Informationsaggregation.*

H6(b): *Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Kombination aus Tabelle und Graphik ist die prozedurale Geschwindigkeit bei niedriger Informationsaggregation geringer als bei hoher Informationsaggregation.*

H6(c): *Bei einem komplexen multikriteriellen Problem unter Sicherheit mit Darstellung in Form einer Kombination aus Tabelle und Graphik ist die Ergebnistransitivität bei niedriger Informationsaggregation geringer als bei hoher Informationsaggregation.*

Umanath/Vessey (1994) untersuchen im Rahmen einer multiattributiven holistischen Insolvenzprognoseaufgabe die Auswirkungen von Darstellungsformat und Informationsmenge auf die Entscheidungsqualität. Die Informationsmenge wird erhöht durch die Hinzunahme von Variablen, die zur Problemlösung nicht benötigt werden, sodass der Informationsgehalt konstant bleibt. Sowohl bei graphischer als auch tabellarischer Darstellung verringert sich mit Anstieg der Informationsmenge zwar die prozedurale Geschwindigkeit, die Entscheidungsqualität hingegen bleibt gleich. Bei Abdel-Khalik (1973) hat Informationsaggregation vom Typ I im Rahmen einer multiattributiven Insolvenzprognoseaufgabe mit tabellarischer Darstellung keinen positiven Effekt auf die Entscheidungsqualität. Anagol/Gamble (2013) und Beshears et al. (2011) analysieren Typ I bei mehrkriteriellen Investitionsentscheidungen von Finanzanlagen. Beide Studien untersuchen, ob eine disaggregierte Darstellung einzelner Wertpapierrenditen zu einer geringeren Entscheidungsqualität im Sinne einer geringeren Risikobereitschaft führt als eine aggregierte Darstellung in Form von Portfoliorenditen. Die Darstellung erfolgt jeweils graphisch. Anagol/Gamble (2013) finden im Durchschnitt keine Auswirkung der Informationsaggregation auf die Entscheidungsqualität. Beshears et al. (2011) finden im Rahmen eines Feldexperimentes ebenfalls keine Auswirkung der Informationsaggregation auf die Risikobereitschaft. Auch Kaufmann/Weber (2013) analysieren Typ I im Rahmen von mehrkriteriellen Investitionsentscheidungen von Finanzanlagen. Die Darstellung erfolgt in einer Kombination aus Graphik und numerischen Werten. Untersucht wird wiederum, ob eine disaggregierte Darstellung von Wertpapierrenditen zu einer geringeren Risikobereitschaft führt als eine aggregierte Darstellung in Form

von Portfoliorenditen. Im Ergebnis verbessert Informationsaggregation die Risikobereitschaft, die prozedurale Geschwindigkeit hingegen ändert sich nicht. Swink/Speier (1999) untersuchen im Rahmen eines multikriteriellen Netzwerk-Design-Problems mit graphischer Darstellung und Informationsaggregation vom Typ II. Die prozedurale Geschwindigkeit erhöht sich durch Informationsaggregation. Werden die Daten graphisch stark zerstreut dargestellt, nimmt die Entscheidungsqualität mit zunehmender Aggregation ab.

Zusammenfassend zeigt sich empirisch bisher, dass Informationsaggregation vom Typ I bei kombinierter Darstellung die objektive Entscheidungsqualität verbessern kann, bei rein tabellarischer oder graphischer Darstellung hingegen nicht. Bei rein tabellarischer Darstellung wird die prozedurale Geschwindigkeit durch Informationsaggregation erhöht, bei kombinierter Darstellung nicht.

3 Methodik und Aufbau des Experiments

3.1 Aufbau des Experiments

Um die Effekte von Darstellung und Informationsaggregation auf die Entscheidungsgüte zu untersuchen, wird ein vollfaktorieller 2×3-Between-Subjects Experimentaufbau verwendet (vgl. Obermaier/Müller, 2008, S. 335 f.). Der Aggregationsgrad der Information wird über zwei Ausprägungen (*hohe Informationsaggregation* und *niedrige Informationsaggregation*) und das Darstellungsformat über drei Ausprägungen (*Tabelle*, *Graphik*, und *Kombination aus Tabelle und Graphik*) variiert. Die Gruppen gliedern sich wie folgt:

Tabelle 1: Übersicht über den Experimentaufbau (n = Anzahl der Teilnehmer)

	Tabelle	Graphik	Tabelle und Graphik
Hohe Aggregation	n = 19	n = 19	n = 19
Niedrige Aggregation	n = 19	n = 19	n = 19

Das Between-Subjects Design eignet sich zur Untersuchung der Hypothesen, die auf der Cognitive Fit-Theorie und der Theorie der menschlichen Informationsverarbeitung basieren, indem es die Wahrscheinlichkeit minimiert, dass Teilnehmende Fehler und Inkonsistenzen im Entscheidungsprozess korrigieren, und somit ihren natürlichen

Denkprozess möglichst wenig beeinflusst (vgl. Libby et al., 2002, S 804 f.; Kahneman/Tversky, 1996, S. 587.). Entscheidungsgüte wird aus Prozessperspektive untersucht anhand der abhängigen Variablen *prozedurale Konsistenz* und *prozedurale Geschwindigkeit* und aus Ergebnisperspektive anhand der abhängigen Variable *Ergebnistransitivität*.

3.2 Experimentaufgabe und unabhängige Variablen

Die experimentelle Aufgabe ist ein komplexes multikriterielles Entscheidungsproblem mit räumlicher Informationsbeschaffung und symbolischer Informationsauswertung. Aufgabe der Teilnehmenden war das Bilden einer Rangfolge von Alternativen gemäß den eigenen Präferenzen. Für die Alternativenreihung mußten (1) die Struktur der Kriterien und die Beziehungen zwischen den Alternativen identifiziert werden (räumliche Informationsbeschaffung) und (2) die Merkmalsausprägungen der Kriterien extrahiert und die Datenwerte weiterverarbeitet werden (symbolische Informationsauswertung).

Konkret war es Aufgabe fünf Smartphones auf Basis einer multikriteriellen Entscheidungsvorlage mit Bewertungen der Kriterien für die Alternativen, gemäß den eigenen Präferenzen, in eine ordinale Rangfolge zu bringen. Damit wurde ein generisches Entscheidungsproblem konstruiert, mit dem die Teilnehmenden vertraut sein sollten, um fach- oder erfahrungsspezifische Verzerrungen hinsichtlich einer bestimmten Studierendengruppe auszuschließen.

Die fünf Smartphones wurden in den Instruktionen anhand detaillierter Beschreibungen von acht Oberkriterien und zugehöriger zwanzig Unterkriterien charakterisiert. Sämtliche Kriterien und Bewertungen basierten auf dem Test „Smartphones und Multimediahandys“ der Stiftung Warentest (2011, Nr. 5, S. 38-43). Dadurch sollte zu einem hohen Grad sichergestellt werden, dass keine faktisch relevanten Entscheidungskriterien ausgelassen wurden. Ferner wurde auf die Angabe von Marken explizit verzichtet, um eine Verzerrung der Entscheidung aufgrund von Markenpräferenzen verhindern zu können und die Entscheidung auf mehr oder weniger objektive Merkmale stützen zu können. Die einführenden Instruktionen und die Beschreibungen der Oberkriterien und der zugehörigen Unterkriterien waren für alle Teilnehmenden identisch. Ausschließlich die Entscheidungsvorlage unterschied sich hinsichtlich der unabhängigen Variablen Informationsaggregation und Darstellung.

3.2.1 Informationsaggregation

Zur Prüfung der Hypothesen ist es notwendig, dass die Informationsaggregation vom Typ I ist, d.h. dass durch Aggregation zwar die Informationsmenge reduziert wird, nicht aber der Informationsgehalt. Dies bedeutet, dass allen Teilnehmenden die zur Erstellung einer Rangfolge notwendige Information zur Verfügung stehen muss. Um einen konstanten Informationsgehalt zu gewährleisten, standen allen Teilnehmenden die Instruktionen mit den Beschreibungen der acht Oberkriterien und den zugehörigen zwanzig Unterkriterien in identischer Form zur Verfügung.

In Bezug auf die Informationsmenge wurde in der Entscheidungsvorlage die Anzahl an Alternativen und Anzahl an Oberkriterien für alle Experimentgruppen konstant gehalten. Gewichtungen der Kriterien und Gesamtwerte für jede Alternative standen den Teilnehmenden explizit nicht zur Verfügung. Die Gruppen mit *hoher Informationsaggregation* und *niedriger Informationsaggregation* unterschieden sich ausschließlich hinsichtlich des Aggregationsgrades der Kriterien. In den Gruppen mit *hoher Informationsaggregation* standen den Teilnehmenden die Datenwerte der acht Oberkriterien zur Verfügung. In den Gruppen mit *niedriger Informationsaggregation* standen den Teilnehmenden die Datenwerte der zwanzig Unterkriterien zur Verfügung. Damit betrug die Gesamtanzahl an Datenwerten bei *hoher Informationsaggregation* vierzig Datenwerte und bei *niedriger Informationsaggregation* einhundert Datenwerte.

Da allen Teilnehmenden neben der Entscheidungsvorlage die Instruktionen mit den Beschreibungen aller Ober- und Unterkriterien zur Verfügung standen, bestand für alle die Möglichkeit, auf Basis der gegebenen Information in der Entscheidungsvorlage, eine subjektive Rangfolge gemäß der eigenen Präferenzen zu erstellen, sodass die Informationsaggregation als Typ I einzustufen ist.

3.2.2 Darstellung

Die Information in der Entscheidungsvorlage wurde in drei unterschiedlichen Darstellungsformaten präsentiert (*Tabelle, Graphik, Tabelle und Graphik*). Die Entscheidungsvorlage war jeweils auf einer Seite abgebildet, sodass die kognitiven Prozesse zur Informationsbeschaffung und -auswertung besser kontrolliert werden können (vgl. Speier, 2006, S. 1122). Die Reihenfolge sämtlicher Kriterien entspricht dem Test der Stiftung Warentest (2011) und wird zwischen den unterschiedlichen Darstellungsformaten nicht variiert. Konkret wurden die drei Darstellungsformate in der Entscheidungsvorlage folgendermaßen gestaltet:

(1) Für die Darstellung in Form einer *Tabelle* sind die Kriterien mit den zugehörigen numerischen Bewertungen zeilenweise und die fünf Alternativen spaltenweise abgetragen. In der *Tabelle mit hoher Informationsaggregation* (Anhang 1) befinden sich in den Zeilen die acht Oberkriterien mit numerischen Werten für die fünf Alternativen. In der *Tabelle mit niedriger Informationsaggregation* (Anhang 2) befinden sich in den Zeilen jeweils nach einem der acht Oberkriterien die Unterkriterien mit numerischen Werten für die fünf Alternativen.

(2) Für die Darstellung in Form einer *Graphik* sind die Kriterien mit den zugehörigen Bewertungen in Form von Säulendiagrammen zeilenweise abgetragen. In jeder Zeile befindet sich neben der Nennung des Kriteriums ein Diagramm bestehend aus fünf Säulen, eine Säule pro Alternative. Eine höhere Säule stellt eine bessere Bewertung dar. In der *Graphik mit hoher Informationsaggregation* (Anhang 3) befinden sich in den Zeilen die acht Oberkriterien mit einem Säulendiagramm je Zeile. In der *Graphik mit niedriger Informationsaggregation* (Anhang 4) befinden sich in den Zeilen jeweils nach einem der acht Oberkriterien die Unterkriterien mit einem Säulendiagramm je Zeile.

(3) Für die Darstellung in Form einer Kombination aus *Tabelle und Graphik* sind die Kriterien mit den zugehörigen numerischen Bewertungen zeilenweise und die fünf Alternativen spaltenweise abgetragen. In jeder Zeile befindet sich neben den numerischen Werten zusätzlich ein Diagramm bestehend aus fünf Säulen, eine Säule pro Alternative. Eine höhere Säule stellt eine bessere Bewertung dar. In der *Tabelle und Graphik mit hoher Informationsaggregation* (Anhang 5) befinden sich in den Zeilen die acht Oberkriterien mit numerischen Werten für die fünf Alternativen und ein Säulendiagramm je Zeile. In der *Tabelle und Graphik mit niedriger Informationsaggregation* (Anhang 6) befinden sich in den Zeilen jeweils nach einem der acht Oberkriterien die Unterkriterien mit numerischen Werten für die fünf Alternativen und ein Säulendiagramm je Zeile.

3.3 Abhängige Variablen

3.3.1 Ergebnistransitivität

Aufgabe der Experimentteilnehmenden war es, eine Rangfolge gemäß den eigenen Präferenzen zu erstellen. *Ergebnistransitivität* als abhängige Variable vergleicht daher die von den Teilnehmenden im Experiment gewählte faktische Rangfolge der Alternativen mit einer rekonstruierten fiktiv-rationalen Reihung der Alternativen auf Basis einer mehrkriteriellen Präferenzfunktion je Alternative. Im Gegensatz zu bestehenden Studien

mit (quasi-)dominanter Handlungsalternative kann durch diese Herangehensweise auch bei Berücksichtigung subjektiver Artenpräferenzrelationen die Rationalität des Ergebnisses beurteilt werden.

Um einen rationalen Gesamtnutzenwert für jede Alternative ermitteln zu können, werden Informationen hinsichtlich der Artenpräferenzrelationen der Probanden benötigt. Deren Ermittlung erfolgt in der vorliegenden Studie auf Basis des Analytischen Hierarchieprozesses (AHP) (vgl. Saaty, 1994, 1990, 1980). Hierfür werden unmittelbar nach Erstellung der subjektiven Rangfolge paarweise Vergleiche zwischen den Kriterien abgefragt, um basierend auf der Eigenvektormethode Zielgewichte für alle Kriterien ermitteln zu können. Das Verfahren unterstellt eine Zielhierarchie, in der das Entscheidungsziel in Oberkriterien aufgeteilt und in hierarchisch untergeordnete Kriterien untergliedert werden kann. Diese Bedingungen decken sich mit der Struktur der Experimentaufgabe, bei der die Kriterien (Gesamtnutzenwert als Entscheidungsziel) in Oberkriterien und Unterkriterien strukturiert sind. Dementsprechend werden durch paarweise Vergleiche der Kriterien die Artenpräferenzrelationen der Teilnehmenden rekonstruiert. Zu deren Rekonstruktion mußten die Gruppen mit hoher Informationsaggregation acht Oberkriterien durch paarweisen Vergleich bewerten, die Gruppen mit geringer Informationsaggregation mußten zusätzlich die 20 zugrundeliegenden Unterkriterien bewerten.

So kann eine fiktiv-rationale Reihung der Alternativen erzeugt werden, indem auf Basis der Artenpräferenzrelationen für die Ober- und Unterkriterien Zielgewichte ermittelt werden, aus denen sich in Kombination mit den numerischen Werten der Kriterien ein Gesamtnutzenwert je Alternative ergibt. Die Merkmalsausprägungen der Alternativen werden damit faktisch als quasi-kardinale Werte behandelt (Zur Diskussion ihrer Eignung im Rahmen des AHP vgl. Schneeweiß, 1991, S. 173-175). Die so gewonnene Reihung der Alternativen fungiert im Sinne einer prozedural-rationalen Entscheidung als Vergleichsmaßstab für die faktisch getroffene Entscheidung.

Dementsprechend wird die Ergebnistransitivität für jeden Probanden als Summe der absoluten Abweichungen der Ränge zwischen dessen faktischer Rangfolge der Alternativen $Rang_{faktisch}$ und der fiktiv-rationalen rekonstruierten Rangfolge der Alternativen

$Rang_{fiktiv-rational}$ ermittelt:

$$Ergebnistransitivität = \sum_{i=1}^5 \left| Rang_{faktisch}^i - Rang_{fiktiv-rational}^i \right| \quad (1)$$

wobei i den Laufindex der fünf Alternativen bezeichnet. Der Grad der Ergebnistransitivität ist dabei umso höher, je geringer die absolute Summe der Abweichungen zwischen den Rängen ist. Die Summe der Abweichungen kann minimal 0 und maximal 12 Punkte ergeben, wobei eine Übereinstimmung von faktischem und fiktiv-rationalem Ergebnis einer Ergebnistransitivität in Höhe von 0 entspricht. Die Eignung des vorgeschlagenen Konstrukts wird durch *Kendalls* τ in Höhe von -,966 ($p < ,001$) bestätigt.

3.3.2 Prozedurale Konsistenz

Prozedurale Konsistenz als abhängige Variable misst die Konsistenz der Paarvergleiche zwischen den Kriterien. Die Ermittlung der prozeduralen Konsistenz erfolgt für alle Gruppen identisch anhand der Paarvergleiche der acht Oberkriterien. Die durch Paarvergleiche rekonstruierten Artenpräferenzrelationen schlagen sich in sogenannten Paarvergleichsmatrizen nieder. Dabei kann eine Paarvergleichsmatrix entweder konsistent sein, oder nicht. Zur Messung des Grades der prozeduralen Konsistenz kann eine Konsistenzprüfung durchgeführt werden. Dazu wird eine sogenannte Durchschnittsmatrix als Vergleichsbasis erzeugt, deren durchschnittlicher Eigenwert λ_{\max} in die Ermittlung eines Konsistenzindex CI eingeht (vgl. Saaty, 1994, S. 41 f.):

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (2)$$

wobei n die Anzahl der Zeilen bzw. Spalten der Durchschnittsmatrix darstellt. Wird dieser Konsistenzindex CI ins Verhältnis zu Werten R entsprechender zufallsgesteuerter Testuntersuchungen gesetzt, ergibt sich die Konsistenzkennzahl CR (vgl. Saaty, 1994, S. 41 f.):

$$CR = \frac{CI}{R} \quad (3)$$

wobei R den durchschnittlichen Zufallskonsistenzindex in Abhängigkeit von der Matrixgröße darstellt und Saaty (1994, S. 42) entnommen werden kann. Der Grad der prozeduralen Konsistenz als abhängige Variable wird gemessen anhand der Konsistenzkennzahl CR .

Ein Wert der Konsistenzkennzahl CR in Höhe von Null impliziert vollständige prozedurale Konsistenz. Der Wert der Konsistenzkennzahl ist umso höher, je inkonsistenter die Vergleichsurteile eines Entscheiders sind, sodass mit steigender Konsistenzkennzahl

die prozedurale Konsistenz abnimmt. Überschreitet die Konsistenzkennzahl einen Wert von 0,1 gelten die Präferenzen eines Entscheiders gemeinhin als inkonsistent (vgl. Saaty, 1990, S. 13).

3.3.3 Prozedurale Geschwindigkeit

Die *prozedurale Geschwindigkeit* als abhängige Variable wurde automatisiert als Zeitdauer in Sekunden ermittelt und umfasst die Zeitspanne, die sich über die individuelle Erstellung der subjektiven Reihung durch die Probanden erstreckt.

3.4 Teilnehmer und Experimentablauf

Im Rahmen des Experimentes galt es ein generisches multikriterielles Entscheidungsproblem zu lösen. Hierfür nahmen 114 Studierende an dem rechnergestützten Laborexperiment² teil. Deren Durchschnittsalter betrug 22,65 Jahre. Die Probanden bilden einen Querschnitt unterschiedlicher Studiengänge und Fakultäten. Tabelle 2 zeigt einen Überblick über den Experimentablauf.

Tabelle 2: Übersicht über den Experimentablauf

Abschnitt 1	Abschnitt 2	Abschnitt 3	Abschnitt 4
Einführende Instruktionen	Komplexes mehrkriterielles Entscheidungsproblem unter Sicherheit als experimentelle Aufgabe:	Paarweise Vergleiche der acht Oberkriterien	Post-Test Fragebogen
Beschreibung der Ober- und Unterkriterien	Erstellen einer Rangfolge von Alternativen auf Basis einer mehrkriteriellen Entscheidungsvorlage	Paarweise Vergleiche der zwanzig Unterkriterien (nur in den Gruppen mit niedriger Informationsaggregation)	Auszahlung an die Teilnehmenden
Instruktionen zum Entscheidungsproblem			

Im ersten Abschnitt des rechnergestützten Experimentes wurde den Teilnehmenden, nach einführenden Instruktionen, das multikriterielle Entscheidungsproblem erläutert. Sie erhielten einen Ausdruck der Beschreibung der Ober- und Unterkriterien und Instruktionen zum Entscheidungsproblem. Um einen Anreiz für die bestmögliche Erfüllung der Aufgabe zu bieten, wurde in der Aufgabenbeschreibung eine Auszahlung am Ende des Experiments in Aussicht gestellt.

² Das Experiment wurde im Juli 2011 im ökonomischen Experimentallabor Paula der Universität Passau unter Verwendung der Software *z-Tree* (Fischbacher, 2007) durchgeführt.

Im zweiten Abschnitt des Experimentes war es Aufgabe der Teilnehmenden, das Entscheidungsproblem in Form des Erstellens der Rangfolge der Smartphones zu lösen. Hierfür wurden sie aufgefordert die Entscheidungsvorlage, die am Arbeitsplatz zunächst noch verdeckt war, aufzudecken.

Nach Erstellung der Rangfolge folgten im dritten Abschnitt die paarweisen Vergleiche der Kriterien. Auf Basis der aus dem Analytischen Hierarchieprozess bekannten Bewertungsskala von Saaty (1980) war für jedes Kriterium anzugeben, wie bedeutend dieses Kriterium bei der Erstellung der Rangfolge im Vergleich zu einem anderen Kriterium war. Alle Teilnehmenden mußten zunächst die acht Oberkriterien durch paarweisen Vergleich bewerten. Die Gruppen mit *geringer Informationsaggregation* mußten im Anschluß zusätzlich die zwanzig Unterkriterien paarweise vergleichen. Um die Anzahl der notwendigen Paarvergleiche zu begrenzen, wurde Reziprozität der Urteile unterstellt.

Im vierten und letzten Abschnitt hatten die Probanden noch einen Post-Test-Fragebogen zu beantworten. Am Ende des Experiments erhielten alle Teilnehmer eine feste Auszahlung in Höhe von 5 €. Die Experimentdauer in den einzelnen Sessions bewegte sich zwischen 30 und 45 Minuten und betrug im Mittel ca. 40 Minuten.

4 Ergebnisse

4.1 Manipulationschecks

Zur Überprüfung eines generellen Verständnisses der Experimentaufgabe wurde im Post-Test Fragebogen das allgemeine Aufgabenverständnis abgefragt. Alle 114 Teilnehmenden beantworteten die dichotome Frage: „Ich habe die Anweisungen und die Aufgabenstellung (Erstellung eines Rankings auf Basis eines Qualitätstests) verstanden.“ mit „ja“.

Um zu überprüfen, ob die Manipulation der Informationsaggregation für die komplexe Aufgabe erfolgreich war, wurde analysiert, wie sich die Experimentgruppen mit *hoher* und *niedriger Informationsaggregation* in ihrem wahrgenommenen *Überlastungs-* und *Mengenempfinden* unterschieden haben (Tabelle 6).

Überlastungsempfinden wird gemessen auf einer 7-stufigen Likert-Skala anhand der Aussage: „Ich fühlte mich von der Menge der bereitgestellten Informationen im Qualitätstest überfordert.“ mit „1 = trifft nicht zu“ und „7 = trifft voll zu“. Ein Mittelwertvergleich zeigt, dass die Gruppen mit *hoher Informationsaggregation* (2,30) die Informa-

tionsüberlastung signifikant niedriger einschätzten als die Gruppen mit *niedriger Informationsaggregation* (3,21) (Mann-Whitney U; $z = -2,82$; $p = ,005$; zweiseitig). Das Ergebnis zeigt, dass die Informationsaggregation einen signifikanten Einfluß auf das subjektive *Überlastungsempfinden* der Teilnehmenden hat und bestätigt die Manipulation der Informationsaggregation.

Mengenempfinden wird gemessen auf einer 7-stufigen Likert-Skala anhand der Aussage: „Die Menge der zur Verfügung stehenden Informationen im Qualitätstest empfand ich als...“ mit „1 = viel zu gering“ und „7 = viel zu hoch“. Ein Mittelwertvergleich zeigt, dass sowohl die Gruppen mit *hoher Informationsaggregation* (3,93) als auch die Gruppen mit *niedriger Informationsaggregation* (4,21) die Informationsmenge ähnlich hoch einschätzten (Mann-Whitney U; $z = -1,47$; $p = ,142$; zweiseitig), sodass die Experimentaufgabe als komplex eingestuft werden kann.

4.2 Deskriptive Statistik und Hypothesentests

Analog zu Umanath/Vessey (1994, S. 814), wird zunächst überprüft, ob die *Ergebnistransitivität* der Teilnehmenden im Durchschnitt von einer rein zufälligen Reihung abweicht. Die *Ergebnistransitivität* der Teilnehmenden ist in jedem Treatment signifikant höher als die rein zufällige *Ergebnistransitivität* in Höhe von $8,0^3$ (alle $t > 3,99$, alle $p < ,001$). Dieses Ergebnis deutet darauf hin, dass die Reihungen im Experiment nicht zufällig gewählt wurden.

Im Anschluß wird erkundet, ob die abhängigen Variablen miteinander korreliert sind, da dies Analysen in Form von multivariaten Tests zur Folge hätte. *Prozedurale Konsistenz* und *prozedurale Geschwindigkeit* (Spearman's Rho; $\rho = ,023$; $p = ,884$), *prozedurale Konsistenz* und *Ergebnistransitivität* (Spearman's Rho; $\rho = -,120$; $p = ,203$), sowie *prozedurale Geschwindigkeit* und *Ergebnistransitivität* (Pearson-Korrelation; $r = 0,012$; $p = ,901$) korrelieren nicht signifikant miteinander. Die weitere Analyse wird daher in Form von univariaten Tests durchgeführt.

Die Effekte der unabhängigen Variablen Darstellung und Informationsaggregation auf *prozedurale Konsistenz*, *prozedurale Geschwindigkeit* und *Ergebnistransitivität* werden mit SPSS auf Basis verallgemeinerter linearer Modelle analysiert.

³ Bei fünf Alternativen ergeben sich $5! = 120$ verschiedene Kombinationen an Vergleichen zwischen faktischer und subjektiv rationaler Reihung der Alternativen. Mögliche Ausprägungen der abhängigen Variable *Ergebnistransitivität* sind 0, 2, 4, 6, 8, 10, 12. Der Mittelwert aller 120 möglichen Kombinationen ist dann 8,0.

Tabelle 3: Deskriptive Statistik zu prozeduraler Konsistenz, prozeduraler Geschwindigkeit und Ergebnistransitivität

Darstellung ^d	Abhängige Variable	Informationsaggregation ^c		
		Hohe Aggregation (Oberkategorien)	Niedrige Aggregation (Ober- und Unterkategorien)	Randmittel
Tabelle	Konsistenz	0,104 [0,053] (n = 19)	0,181 [0,181] (n = 19)	0,143 [0,137] (n = 38)
	Geschwindigkeit	239,89 [87,23] (n = 19)	350,37 [146,40] (n = 19)	295,13 [131,39] (n = 38)
	Transitivität	3,89 [2,35] (n = 19)	4,32 [3,22] (n = 19)	4,11 [2,79] (n = 38)
Graphik	Konsistenz	0,257 [0,264] (n = 19)	0,347 [0,295] (n = 19)	0,302 [0,280] (n = 38)
	Geschwindigkeit	184,58 [98,67] (n = 19)	290,84 [126,54] (n = 19)	237,71 [124,20] (n = 38)
	Transitivität	4,21 [2,97] (n = 19)	3,58 [3,02] (n = 19)	3,89 [2,98] (n = 38)
Tabelle und Graphik	Konsistenz	0,153 [0,184] (n = 19)	0,275 [0,352] (n = 19)	0,214 [0,284] (n = 38)
	Geschwindigkeit	256,58 [93,92] (n = 19)	330,63 [119,06] (n = 19)	293,61 [112,23] (n = 38)
	Transitivität	5,16 [2,85] (n = 19)	4,32 [2,43] (n = 19)	4,74 [2,65] (n = 38)
Randmittel	Konsistenz	0,172 [0,196] (n = 57)	0,268 [0,288] (n = 57)	0,220 [0,250] (n = 114)
	Geschwindigkeit	227,02 [96,82] (n = 57)	323,95 [131,21] (n = 57)	275,48 [124,69] (n = 114)
	Transitivität	4,42 [2,74] (n = 57)	4,07 [2,88] (n = 57)	4,25 [2,81] (n = 114)

^a Die abhängige Variable *prozedurale Konsistenz* wird gemessen anhand der Konsistenzkennzahl, die im Rahmen des AHP-Verfahrens ermittelt wird. Die Werte der Konsistenzkennzahl sind umso größer, je inkonsistenter die Vergleichsurteile der Experimentteilnehmenden sind.

^b Die abhängige Variable *prozedurale Geschwindigkeit* wird automatisiert in Sekunden ermittelt und umfasst die Zeitspanne, die sich über die individuelle Erstellung der subjektiven Reihung durch die Probanden erstreckt.

^c Die abhängige Variable *Ergebnistransitivität* wird für jeden Probanden als Summe der absoluten Abweichungen der Ränge zwischen dessen faktischer (begrenzt rationaler) Reihung der Alternativen und der fiktiv-rationalen rekonstruierten Reihung der Alternativen ermittelt. Je höher die Ausprägung der Variable, desto geringer die Ergebnistransitivität. Die Summe der Abweichungen kann minimal 0 und maximal 12 Punkte ergeben.

^d Die unabhängige Variable *Darstellung* umfasst die drei Faktorstufen *Tabelle*, *Graphik*, sowie *Tabelle und Graphik*. Sie wird zwischen den Gruppen (between-subjects) anhand des Darstellungsformates der Entscheidungsvorlage variiert. Für die Gruppen *Tabelle* erfolgt die Darstellung in tabellarischer Form mit numerischen Einzelwerten. Die Gruppen *Graphik* erhalten eine graphische Darstellung mit Säulendiagrammen. Die Gruppen *Tabelle und Graphik* erhalten eine Kombination aus der Darstellung der Gruppen *Tabelle* und *Graphik*, welche sowohl die numerischen Einzelwerte als auch die Säulendiagramme enthält.

^e Die unabhängige Variable *Informationsaggregation* umfasst die zwei Faktorstufen *hohe Aggregation* und *niedrige Aggregation*. Sie wird zwischen den Gruppen (between-subjects) anhand der Anzahl der Kriterien in der Entscheidungsvorlage des Laborexperimentes variiert. Die Anzahl der Alternativen wurde für alle Treatments konstant gehalten. In den Gruppen mit hoher Aggregation werden fünf Alternativen mit Einzelwerten für acht Oberkriterien (d.h. 40 Informationseinheiten) versehen. In den Gruppen mit niedriger Aggregation werden fünf Alternativen mit Einzelwerten für zwanzig Unterkriterien (d.h. 100 Informationseinheiten), welche den acht Oberkriterien zugeordnet sind, versehen.

Es können sechs distinktive Gruppenmittelwerte ermittelt werden: *hohe Informationsaggregation* und *geringe Informationsaggregation* mit Darstellung in Form einer *Tabelle*, *Graphik* oder Kombination aus *Tabelle und Graphik*.

Verallgemeinerte lineare Modelle stellen Globaltests der Effekte der unabhängigen Variablen dar. Da sich die Darstellung aus drei Faktorstufen zusammensetzt, werden weiterführende Untersuchungen in Form von Mittelwertvergleichen durchgeführt, um die Effekte von Darstellung und Informationsaggregation näher zu untersuchen.

Tabelle 4: Verallgemeinerte lineare Modelle mit prozeduraler Konsistenz, prozeduraler Geschwindigkeit und Ergebnistransitivität als abhängige Variablen

Abhängige Variable	Prozedurale Konsistenz ^b			Prozedurale Geschwindigkeit ^c			Ergebnistransitivität ^d		
Verteilung	Gamma			Normal			Normal		
Link-Funktion	Log			Identität			Identität		
Faktoren	df	Wald- χ^2	p-Wert ^e	df	Wald- χ^2	p-Wert ^e	df	Wald- χ^2	p-Wert ^e
Aggregation (A)	1	9,23	,002***	1	21,80	<,001***	1	0,46	,496
Darstellung (D)	2	16,19	<,001***	2	6,63	,036**	2	1,93	,381
A x D Interaktion	2	,66	,721	2	0,61	,736	2	1,15	,563
	<i>N = 114</i>			<i>N = 114</i>			<i>N = 114</i>		

^a Das verallgemeinerte lineare Modell ist eine Erweiterung des allgemeinen linearen Modells. Die abhängige Variable steht mit einer festgelegten Link-Funktion in linearem Zusammenhang zu den unabhängigen Faktoren. Es ist mit diesem Modell möglich, dass die Verteilung der abhängigen Variable von einer Normalverteilung abweichen kann. Die abhängige Variable kann metrisch skaliert oder binär sein, bzw. in Form von Häufigkeiten vorliegen. Annahmegemäß sind die unabhängigen Faktoren kategorial. Das verallgemeinerte lineare Modell deckt verschiedenste statistische Modelle ab, wie beispielsweise die lineare Regression bzw. Varianzanalyse für normalverteilte Daten, logistische Modelle für binäre Daten und loglineare Modelle für Häufigkeitsdaten, sowie viele andere statistische Modelle (IBM 2012, S. 52).

^b *Prozedurale Konsistenz* als abhängige Variable ist gammaverteilt. Diese Verteilung eignet sich für Variablen mit positiven Skalenwerten, die in Richtung größerer positiver Werte verzerrt sind. Die Link-Funktion wird als Log festgelegt, so dass $f(x) = \log(x)$ (IBM 2012, S. 55 f.).

^c *Prozedurale Geschwindigkeit* als abhängige Variable ist normalverteilt. Diese Option eignet sich für metrische Variablen, deren Werte eine symmetrische, glockenförmige Verteilung um einen Mittelwert aufweisen. Die Link-Funktion wird als Identität festgelegt, so dass $f(x) = x$. Die abhängige Variable wird nicht transformiert (IBM 2012, S. 55 f.).

^d *Ergebnistransitivität* als abhängige Variable ist normalverteilt. Diese Option eignet sich für metrische Variablen, deren Werte eine symmetrische, glockenförmige Verteilung um einen Mittelwert aufweisen. Die Link-Funktion wird als Identität festgelegt, so dass $f(x) = x$. Die abhängige Variable wird nicht transformiert (IBM 2012, S. 55 f.).

^e * 10 % Signifikanzniveau (zweiseitig).

** 5 % Signifikanzniveau (zweiseitig).

*** 1 % Signifikanzniveau (zweiseitig).

Die deskriptive Statistik ist in Tabelle 3 abgebildet. Die verallgemeinerten linearen Modelle werden in Tabelle 4 dargestellt und die detaillierten Mittelwertvergleiche befinden sich in Tabelle 5.

4.2.1 Prozedurale Konsistenz

Zur Untersuchung der *prozeduralen Konsistenz* wird ein verallgemeinertes lineares Modell verwendet (Tabelle 4). Die Verteilung der prozeduralen Konsistenz ist gamma-verteilt und rechtsschief, daher wird analog zu Cardinaels (2008), Waller et al. (1999) und Gupta/King (1997) eine Log-Link-Funktion verwendet. Darstellung und Informationsaggregation werden als unabhängige Variablen festgelegt.

Der Haupteffekt der Darstellung auf die *prozedurale Konsistenz* ist signifikant (Wald- $\chi^2 = 16,19$; $p < ,001$). Auch der Haupteffekt der Informationsaggregation auf die *prozedurale Konsistenz* ist signifikant (Wald- $\chi^2 = 9,23$; $p = ,002$). Der Interaktionseffekt zwischen Darstellung und Informationsaggregation hingegen ist nicht signifikant (Wald- $\chi^2 = ,66$; $p = ,721$).

Um die Effekte der Darstellung auf die *prozedurale Konsistenz* detaillierter zu analysieren, werden Mittelwertvergleiche auf Basis einer Kontrastanalyse durchgeführt (Tabelle 5) (vgl. Buckless/Ravenscroft, 1990). Ein Mittelwertvergleich zeigt, dass eine *Tabelle* (Tabelle 3, Mittelwert der Konsistenzkennzahl = ,143) zu einer signifikant höheren *prozeduralen Konsistenz* führt als eine *Graphik* (,302) (Wald- $\chi^2 = 12,91$; $p < ,001$). Dieses Ergebnis stützt H1(a). Die *prozedurale Konsistenz* ist bei Kombination aus *Tabelle und Graphik* (,214) marginal signifikant höher als bei einer *Graphik* (,302) (Wald- $\chi^2 = 3,55$; $p = ,059$) und signifikant geringer als bei einer *Tabelle* (,143) (Wald- $\chi^2 = 4,07$; $p = ,044$). H2(a) kann nicht bestätigt werden.

Nun wird die Wirkung der Informationsaggregation auf die *prozedurale Konsistenz* untersucht (Tabelle 5). Ein Mittelwertvergleich zeigt, dass die *prozedurale Konsistenz* bei *hoher Informationsaggregation* (,172) signifikant höher ist als bei *niedriger Informationsaggregation* (,268) (Wald- $\chi^2 = 8,42$; $p = ,004$). Globaltest des verallgemeinerten linearen Modells und Mittelwertvergleich stützen H3(a).

Schließlich wird der Effekt der Informationsaggregation auf die *prozedurale Konsistenz* bei gegebenem Darstellungsformat untersucht (Tabelle 5). Bei *tabellarischer Darstellung* führt die *hohe Informationsaggregation* (,104) zu einer marginal signifikant höheren *prozeduralen Konsistenz* als die *niedrige Informationsaggregation* (,181)

(Wald- $\chi^2 = 3,66$; $p = ,056$). Das Ergebnis bestätigt H4(a) in der Tendenz. Bei rein *graphischer Darstellung* unterscheidet sich die *prozedurale Konsistenz* nicht signifikant zwischen *hoher* (,257) und *niedriger Informationsaggregation* (,347) (Wald- $\chi^2 = 1,16$; $p = ,282$). H5(a) muss abgelehnt werden. Bei Kombination aus *Tabelle und Graphik* wiederum ist die *prozedurale Konsistenz* bei *hoher Informationsaggregation* (,153) signifikant höher als bei *niedriger Informationsaggregation* (,275) (Wald- $\chi^2 = 3,98$; $p = ,046$). H6(a) findet Unterstützung.

4.2.2 Prozedurale Geschwindigkeit

Zur Untersuchung der *prozeduralen Geschwindigkeit* wird ein verallgemeinertes lineares Modell verwendet (Tabelle 4). Die Variable ist normalverteilt mit der Identität als Link-Funktion. Darstellung und Informationsaggregation werden als unabhängige Variablen festgelegt.

Der Haupteffekt der Darstellung auf die *prozedurale Geschwindigkeit* ist signifikant (Wald- $\chi^2 = 6,63$; $p = ,036$). Der Haupteffekt der Informationsaggregation auf die *prozedurale Geschwindigkeit* ist ebenfalls signifikant (Wald- $\chi^2 = 21,80$; $p < ,001$). Der Interaktionseffekt zwischen Darstellung und Informationsaggregation wiederum ist nicht signifikant (Wald- $\chi^2 = ,61$; $p = ,736$).

Zur näheren Analyse des Effektes der Darstellung auf die *prozedurale Geschwindigkeit* werden ebenfalls Mittelwertvergleiche auf Basis einer Kontrastanalyse durchgeführt (Tabelle 5). Anhand eines Mittelwertvergleiches wird ersichtlich, dass eine *Graphik* (Tabelle 3, Mittelwert der Entscheidungsdauer in Sekunden = 237,71) zu einer signifikant höheren *prozeduralen Geschwindigkeit* führt als eine *Tabelle* (295,13) (Wald- $\chi^2 = 5,10$; $p = ,024$). H1(b) findet Stützung. Die *prozedurale Geschwindigkeit* ist bei Kombination aus *Tabelle und Graphik* (293,61) signifikant geringer als bei einer *Graphik* (237,71) (Wald- $\chi^2 = 4,83$; $p = ,028$) und unterscheidet sich nicht von einer *Tabelle* (295,13) (Wald- $\chi^2 = ,00$; $p = ,952$). Die Ergebnisse widersprechen H2(b).

Es folgt die Untersuchung der Wirkung der Informationsaggregation auf die *prozedurale Geschwindigkeit* (Tabelle 5). Ein Mittelwertvergleich zeigt, dass bei *hoher Informationsaggregation* (227,02) im Vergleich zu *niedriger Informationsaggregation* (323,95) die *prozedurale Geschwindigkeit* signifikant höher ist (Wald- $\chi^2 = 21,80$; $p < ,001$). Gemeinsam mit dem Globaltest des verallgemeinerten linearen Modells stützt das Ergebnis H3(b).

Tabelle 5: Mittelwertvergleiche für prozedurale Konsistenz, prozedurale Geschwindigkeit und Ergebnistransitivität

Prozedurale Konsistenz ^a						Prozedurale Geschwindigkeit ^b						Ergebnistransitivität								
H _A	Hypothese ^d	Treat- ment 1	Treat- ment 2	Wald- χ^2	p-Wert	H _A	Hypothese ^d	Treat- ment 1	Treat- ment 2	Wald- χ^2	p-Wert	H _A	Hypothese ^d	Treat- ment 1	Treat- ment 2	Wald- χ^2	p-Wert			
Darstellung						Darstellung						Darstellung								
H1(a)	T > G	☑	,137	,299	12,91	<,001***	H1(b)	G > T	☑	237,71	295,13	5,10	,024**	H1(c)	G = T	☑	3,89	4,11	,11	,739
H2(a)	TG > G	☑	,205	,299	3,55	,059*	H2(b)	TG > G		293,61	237,71	4,83	,028**	H2(c)	TG > G		4,74	3,89	1,78	,182
H2(a)	TG = T		,205	,137	4,07	,044**	H2(b)	TG > T		293,61	295,13	,00	,952	H2(c)	TG > T		4,74	4,11	1,00	,317
Informationsaggregation						Informationsaggregation						Informationsaggregation								
H3(a)	IA _H > IA _N	☑	,160	,259	8,42	,004***	H3(b)	IA _H > IA _N	☑	227,02	323,95	21,80	<,001***	H3(c)	IA _H > IA _L		4,42	4,07	,46	,496
Darstellung bei variierender Informationsaggregation						Darstellung bei variierender Informationsaggregation						Darstellung bei variierender Informationsaggregation								
H4(a)	T _H > T _N	☑	,104	,181	3,66	,056*	H4(b)	T _H > T _N	☑	239,89	350,37	9,44	,002***	H4(c)	T _H > T _N		3,89	4,32	,22	,637
H5(a)	G _H = G _N	☑	,257	,347	1,16	,282	H5(b)	G _H > G _N	☑	184,58	290,84	8,73	,003***	H5(c)	G _H = G _N	☑	4,21	3,58	,50	,479
H6(a)	TG _H > TG _N	☑	,153	,275	3,98	,046**	H6(b)	TG _H > TG _N	☑	256,58	330,63	4,24	,039**	H6(c)	TG _H > TG _N		5,16	4,32	,89	,345

^a Die Mittelwerte für *prozedurale Konsistenz* werden verglichen anhand einer Kontrastanalyse mit einfachen Kontrasten, die auf einem verallgemeinerten linearen Modell mit *prozeduraler Konsistenz* als abhängiger Variable basiert, wie in Tabelle 4 spezifiziert. Bei den abgebildeten Mittelwerten handelt es sich um geschätzte Randmittel.

^b Die Mittelwerte für *prozedurale Geschwindigkeit* werden verglichen anhand einer Kontrastanalyse mit einfachen Kontrasten, die auf einem verallgemeinerten linearen Modell mit *prozeduraler Geschwindigkeit* als abhängiger Variable basiert, wie in Tabelle 4 spezifiziert. Bei den abgebildeten Mittelwerten handelt es sich um geschätzte Randmittel.

^c Die Mittelwerte für *Ergebnistransitivität* werden verglichen anhand einer Kontrastanalyse mit einfachen Kontrasten, die auf einem verallgemeinerten linearen Modell mit *Ergebnistransitivität* als abhängiger Variable basiert, wie in Tabelle 4 spezifiziert. Bei den abgebildeten Mittelwerten handelt es sich um geschätzte Randmittel.

^d Hypothesen wie in Kapitel 2 festgelegt. Treatment 1 bezieht sich auf die linke Seite der Gleichung der entsprechenden Hypothese, Treatment 2 auf die rechte Seite. Abkürzungen sind: T = Tabelle, G = Graphik, TG = Tabelle und Graphik, IA_H = Hohe Informationsaggregation, IA_N = Niedrige Informationsaggregation, T_H = Tabelle bei hoher Informationsaggregation, T_N = Tabelle bei niedriger Informationsaggregation, G_H = Graphik bei hoher Informationsaggregation, G_N = Graphik bei niedriger Informationsaggregation, TG_H = Tabelle und Graphik bei hoher Informationsaggregation; TG_N = Tabelle und Graphik bei niedriger Informationsaggregation. Ein ☑ Symbol deutet an, dass die zugehörige Hypothese gestützt wird.

^e ***, **, * repräsentiert ein 1 %, 5 %, 10 % Signifikanzniveau (jeweils zweiseitig).

Ebenfalls wird der Effekt der Informationsaggregation auf die *prozedurale Geschwindigkeit* bei gegebenem Darstellungsformat analysiert (Tabelle 5). Bei einer *Tabelle* führt die *hohe Informationsaggregation* (239,89) zu einer signifikant höheren *prozeduralen Geschwindigkeit* als die *niedriger Informationsaggregation* (350,37) (Wald- $\chi^2 = 9,44$; $p = ,002$). Das Ergebnis zeigt eine Bestätigung von H4(b) auf. Auch bei rein *graphischer Darstellung* ist die *prozedurale Geschwindigkeit* signifikant höher bei *hoher Informationsaggregation* (184,58) im Vergleich zu *niedriger Informationsaggregation* (290,84) (Wald- $\chi^2 = 8,73$; $p = ,003$). H5(b) wird bestätigt. Auch bei Kombination aus *Tabelle und Graphik* ist die *prozedurale Geschwindigkeit* bei *hoher Informationsaggregation* (256,58) signifikant höher als bei *niedriger Informationsaggregation* (330,63) (Wald- $\chi^2 = 4,24$; $p = ,039$). Das Ergebnis stützt H6(b).

4.2.3 Ergebnistransitivität

Zur Untersuchung der Ergebnistransitivität wird ein verallgemeinertes lineares Modell verwendet (Tabelle 4). Die Variable ist normalverteilt mit der Identität als Link-Funktion. Darstellung und Informationsaggregation werden als unabhängige Variablen festgelegt.

Weder der Haupteffekt der Darstellung auf die *Ergebnistransitivität* (Wald- $\chi^2 = 0,46$; $p = ,496$), noch der Haupteffekt der Informationsaggregation auf die *Ergebnistransitivität* sind signifikant (Wald- $\chi^2 = 1,93$; $p = ,381$). Der Interaktionseffekt zwischen Darstellung und Informationsaggregation ist ebenso nicht signifikant. (Wald- $\chi^2 = 1,15$; $p = ,563$).

Erneut werden Mittelwertvergleiche auf Basis einer Kontrastanalyse durchgeführt, um mögliche Effekte auf die *Ergebnistransitivität* im Vergleich einzelner Darstellungsformate zu untersuchen (Tabelle 5). Ein Mittelwertvergleich zeigt, dass *Tabelle* (Tabelle 3, Mittelwert der Abweichung zwischen faktischer und subjektiv rationaler Reihung = 4,11) und *Graphik* (3,89) sich hinsichtlich der *Ergebnistransitivität* nicht signifikant unterscheiden (Wald- $\chi^2 = ,11$; $p = ,739$). Dieses Ergebnis stützt H1(c). Die *Ergebnistransitivität* unterscheidet sich auch bei Kombination aus *Tabelle und Graphik* (4,74) weder signifikant von einer *Graphik* (3,89) (Wald- $\chi^2 = 1,78$; $p = ,182$) noch signifikant von einer *Tabelle* (4,11) (Wald- $\chi^2 = 1,00$; $p = ,317$). H2(c) steht im Gegensatz zu den Ergebnissen.

Hinsichtlich der Wirkung der Informationsaggregation auf die *Ergebnistransitivität* zeigt ein Mittelwertvergleich, dass sich die *Ergebnistransitivität* bei *hoher Information-*

saggregation (4,42) nicht signifikant von der *niedrigen Informationsaggregation* (4,07) unterscheidet (Tabelle 5; Wald- $\chi^2 = ,46$; $p = ,496$). Sowohl der Globaltest des verallgemeinerten linearen Modells als auch der Mittelwertvergleich stehen im Widerspruch zu H3(c).

Zuletzt wird der Effekt der Informationsaggregation auf die *Ergebnistransitivität* bei gegebenem Darstellungsformat untersucht (Tabelle 5). Bei *tabellarischer Darstellung* unterscheidet sich die *hohe Informationsaggregation* (3,89) nicht von der *niedrigen Informationsaggregation* (4,32) (Wald- $\chi^2 = ,22$; $p = ,637$). Das Ergebnis bestätigt H4(c). Auch bei rein *graphischer Darstellung* unterscheidet sich die *Ergebnistransitivität* nicht signifikant zwischen *hoher* (4,21) und *niedriger Informationsaggregation* (3,58) (Wald- $\chi^2 = ,50$; $p = ,479$). H5(c) findet Unterstützung. Schließlich ist für die *Ergebnistransitivität* bei Kombination aus *Tabelle und Graphik* ebenso kein signifikanter Unterschied zwischen *hoher Informationsaggregation* (5,16) und *niedriger Informationsaggregation* (4,32) zu konstatieren (Wald- $\chi^2 = ,89$; $p = ,345$). H6(c) muss abgelehnt werden.

4.3 Weitere Analysen

Um ausschließen zu können, dass sich die Teilnehmenden im Durchschnitt nur auf die oberen Kriterien der Entscheidungsvorlage fokussieren und ihre Entscheidungen z.B. anhand der ersten vier Kriterien treffen, wurden die durchschnittlichen Gewichte aller Kriterien auf Basis der Paarvergleiche ermittelt. Die Gewichte der Oberkriterien sind im Durchschnitt aller Teilnehmenden: Telefon (,20), Kamera (,10), Musikspieler (,08), Internet und PC (,12), GPS (,06), Handhabung (,14), Stabilität (,13) und Akku (,17) (Reihenfolge gemäß der Entscheidungsvorlage von oben nach unten). Die Mittelwerte deuten nicht darauf hin, dass die Reihenfolge der Kriterien mit einer abnehmenden Gewichtung einhergeht.

Im Post-Test Fragebogen wurde außerdem abgefragt, ob in den Gruppen mit Kombination aus *Tabelle und Graphik* auch beide Formate zur Entscheidungsfindung genutzt wurden. Von 38 Probanden gaben fünf an, bei der Auswertung stärker auf die Graphik geachtet zu haben. Acht Teilnehmende nutzten sowohl Tabelle als auch Graphik. 15 Personen achteten stärker auf die Tabelle als auf die Graphik und zehn nur auf die Tabelle. Insgesamt gaben damit fast 75 % dieser Teilnehmenden an, tatsächlich auf beide Darstellungsformen zur Erstellung der Rangfolge zurückgegriffen zu haben.

Tabelle 6 bietet schließlich einen Überblick der deskriptiven Statistik zu *Mengen- und Überlastungsempfinden*.

Tabelle 6: Deskriptive Statistik zu Mengen- und Überlastungsempfinden

Durchschnittliches Mengen ^a - und Überlastungsempfinden ^b je Treatment (Mittelwert, [Standardabweichung], Anzahl der Beobachtungen)				
Darstellung ^c	Item/ Variable	Informationsaggregation ^d		Randmittel
		Hohe Aggregation (Oberkategorien)	Niedrige Aggregation (Ober- und Unterkategorien)	
Tabelle	Mengenempfinden	3,79 [1,03] (n = 19)	4,47 [1,26] (n = 19)	4,13 [1,19] (n = 38)
	Überlastungsempfinden	2,11 [1,45] (n = 19)	3,68 [1,53] (n = 19)	2,89 [1,67] (n = 38)
Graphik	Mengenempfinden	3,58 [,84] (n = 19)	4,00 [1,45] (n = 19)	3,79 [1,19] (n = 38)
	Überlastungsempfinden	1,89 [1,52] (n = 19)	3,21 [2,02] (n = 19)	2,55 [1,88] (n = 38)
Tabelle und Graphik	Mengenempfinden	4,42 [1,07] (n = 19)	4,16 [,90] (n = 19)	4,29 [,98] (n = 38)
	Überlastungsempfinden	2,89 [1,85] (n = 19)	2,74 [1,66] (n = 19)	2,82 [1,74] (n = 38)
Randmittel	Mengenempfinden	3,93 [1,03] (n = 57)	4,21 [1,22] (n = 57)	4,07 [1,13] (n = 114)
	Überlastungsempfinden	2,30 [1,65] (n = 57)	3,21 [1,76] (n = 57)	2,75 [1,76] (n = 114)

^a *Mengenempfinden (ME)* wird gemessen auf einer 7-stufigen Likert-Skala anhand der Aussage: „Die Menge der zur Verfügung stehenden Informationen im Qualitätstest empfand ich als...“. Mit „1 = viel zu gering“ und „7 = viel zu hoch“.

^b *Überlastungsempfinden (ÜE)* wird gemessen auf einer 7-stufigen Likert-Skala anhand der Aussage: „Ich fühlte mich von der Menge der bereitgestellten Informationen im Qualitätstest überfordert.“ Mit „1 = trifft nicht zu“ und „7 = trifft voll zu“.

^c *Darstellung* umfasst die drei Faktorstufen *Tabelle*, *Graphik*, sowie *Tabelle und Graphik*. Sie wird zwischen den Gruppen (between-subjects) anhand des Darstellungsformates der Entscheidungsvorlage variiert. Für die Gruppen *Tabelle* erfolgt die Darstellung in tabellarischer Form mit numerischen Einzelwerten. Die Gruppen *Graphik* erhalten eine graphische Darstellung mit Säulendiagrammen. Die Gruppen *Tabelle und Graphik* erhalten eine Kombination aus der Darstellung der Gruppen *Tabelle* und *Graphik*, welche sowohl die numerischen Einzelwerte als auch die Säulendiagramme enthält.

^d *Informationsaggregation* umfasst die zwei Faktorstufen *hohe Aggregation* und *niedrige Aggregation*. Sie wird zwischen den Gruppen (between-subjects) anhand der Anzahl der Kriterien in der Entscheidungsvorlage des Laborexperimentes variiert. Die Anzahl der Alternativen wurde für alle Treatments konstant gehalten. In den Gruppen mit hoher Aggregation werden fünf Alternativen mit Einzelwerten für acht Oberkriterien (d.h. 40 Informationseinheiten) versehen. In den Gruppen mit niedriger Aggregation werden fünf Alternativen mit Einzelwerten für zwanzig Unterkriterien (d.h. 100 Informationseinheiten), welche den acht Oberkriterien zugeordnet sind, versehen.

5 Diskussion

Das Ziel der hier vorgestellten Studie ist die Analyse der Effekte von Darstellungsformat und Informationsaggregation auf die Entscheidungsgüte eines komplexen multi-kriteriellen Entscheidungsproblems, ohne eine quasi-dominante Handlungsalternative einführen zu müssen. Das zugrundeliegende Entscheidungsproblem ist unter Bezugnahme auf die Cognitive Fit-Theorie hinsichtlich Informationsbeschaffung als räumlich und hinsichtlich Informationsauswertung als symbolisch einzustufen.

Die prozedurale Konsistenz und damit die Rationalität des Entscheidungsprozesses kann dabei durch Visualisierung nicht verbessert werden. Dies mag im Rahmen der vorliegenden symbolischen Aufgabe, deren Lösung konkrete Datenwerte erfordert, für eine rein graphische Darstellung nicht verwunderlich sein, denn die Rekonstruktion

numerischer Werte aus einer Graphik ist ohne weitere Hilfsmittel kaum möglich und hat deshalb regelmäßig Einbußen hinsichtlich numerischer Präzision zur Folge. Bemerkenswert erscheint allerdings die geringere prozedurale Konsistenz bei Rückgriff auf eine Kombination aus Tabelle und Graphik. So/Smith (2004) führen unter Bezugnahme auf Smith/Taffler (1995, S. 1196, 1209) an, dass Entscheidungsträger im Rahmen komplexer Entscheidungsprobleme bei Verfügbarkeit von Tabelle und Graphik auf die sog. Repräsentativitätsheuristik (vgl. Kahneman/Tversky 1972, S. 448) zurückgreifen und sich so lediglich auf eine Untermenge der verfügbaren Information stützen. Sie nutzen zur Informationsauswertung vor allem die Graphik, deren Verarbeitung mit einem geringeren kognitiven Aufwand verbunden ist als der Rückgriff auf die tabellarischen Daten. Die prozedurale Konsistenz verringert sich damit bei einem symbolischen Problem im Vergleich zu einem rein tabellarischen Format. Dies hat Folgen: seit längerem erfreuen sich (nicht nur) im Zusammenhang mit sog. Management-Dashboards kombinierte Darstellungsformate großer Beliebtheit. Entscheider, die sich der Verhaltenswirkungen des Kombinationsformates im Kontext der vorliegenden Problemklasse nicht bewußt sind, nehmen durch die vordergründig vorteilhaft erscheinende zusätzliche Visualisierung eine geringere prozedurale Rationalität in Kauf.

Insgesamt stützen die Ergebnisse für die reinen Darstellungsformate die Cognitive Fit-Theorie im Rahmen der vorliegenden Problemklasse hinsichtlich der prozeduralen Konsistenz und in Einklang mit Benbasat/Dexter (1986), Smelcer/Carmel (1997) und Dennis/Carste (1998) auch hinsichtlich der prozeduralen Geschwindigkeit. Die Erweiterung der Cognitive Fit-Theorie auf kombinierte Darstellungsformate hingegen findet hinsichtlich prozeduraler Konsistenz und hinsichtlich prozeduraler Geschwindigkeit, wie auch bei Frownfelter-Lohrke (1998) und Bassler (2010), keine Unterstützung.

Die vorliegenden Ergebnisse zeigen überdies, dass Entscheider im Rahmen komplexer multikriterieller Probleme bei aggregierter Information mit konstantem Informationsgehalt hinsichtlich der subjektiven Artenpräferenzrelationen zwischen den Kriterien signifikant konsistentere Vergleichsurteile bilden. Unter Berücksichtigung der Opportunitätskosten des knappen Faktors Zeit erscheint zudem bedeutsam, dass Aggregation bei gleichem Informationsgehalt auch mit höherer prozeduraler Geschwindigkeit einhergeht. Die Theorie der menschlichen Informationsverarbeitung wird gestützt und die Ergebnisse hinsichtlich der prozeduralen Geschwindigkeit stimmen mit bestehenden empirischen Studien überein (vgl. Iselin, 1988; Casey, 1980; Jacoby et al., 1974a, 1974b).

Die Ergebnisse hinsichtlich Informationsaggregation in Abhängigkeit des Darstellungsformates stützen die Cognitive Fit-Theorie aus Prozessperspektive und decken sich in Bezug auf die prozedurale Geschwindigkeit für die reinen Darstellungsformate mit den Ergebnissen von Umanath/Vessey (1994) und Swink/Speier (1999). Bei hoher Informationsaggregation führen eine Tabelle und eine Kombination aus Tabelle und Graphik zu signifikant höherer prozeduraler Konsistenz und prozeduraler Geschwindigkeit im Vergleich zu niedriger Informationsaggregation. Eine Graphik geht bei hoher Informationsaggregation mit einer signifikant höheren Entscheidungsgeschwindigkeit einher, die prozedurale Konsistenz hingegen wird bei graphischer Darstellung durch Aggregation nicht beeinflusst.

Die Ergebnistransitivität kann indes für keines der vorliegenden Darstellungsformate durch Informationsaggregation verbessert werden. Frühe empirische Studien hingegen konnten für die objektive Entscheidungsqualität meist einen positiven Effekt durch Informationsaggregation vom Typ I zeigen (vgl. Otley/Dias, 1982; Tiessen, 1976; Chervany/Dickson, 1974; Barefield, 1972; Ronen, 1971). Für die Tabelle deckt sich das vorliegende insignifikante Ergebnis mit Umanath/Vessey (1994) und Abdel-Khalik (1973). Hinsichtlich der Graphik entspricht das Ergebnis den Studien von Anagol/Gamble (2013), Beshears et al. (2011) und Umanath/Vessey (1994). Bei Kaufmann/Weber (2013) hingegen verbessert sich die Entscheidungsqualität bei Informationsaggregation und kombinierter Darstellung.

Die vorliegenden Ergebnisse zur Ergebnistransitivität können in der Anwendung von Heuristiken bzw. nicht kompensatorischen Entscheidungsregeln durch die Teilnehmenden begründet sein. Es war keineswegs gefordert, dass alle Kriterien der Entscheidungsvorlage in den Entscheidungsprozess einfließen müssen. Der Rückgriff auf bspw. eine kompensatorische Entscheidungsregel blieb damit offen. Es ist denkbar, dass die Teilnehmenden eine nicht-kompensatorische Entscheidungsregel nutzten, dabei gezielt Informationen vernachlässigten, und sich im Entscheidungsprozess lediglich auf wenige Kriterien stützten (vgl. dazu auch Payne, 1976). In diesem Fall zeigt eine Aggregation hinsichtlich des Ergebnisses keine Wirkung, da die gesamte zur Verfügung stehende Information nur selektiv genutzt wird und damit der Verbesserung der Informationsverarbeitungskapazität durch Informationsaggregation keine signifikante Rolle zukommt.

Die gefundenen Ergebnisse haben Konsequenzen für komplexe mehrkriterielle Entscheidungsprobleme. Entgegen der häufig zu beobachtenden Annahme, dass die Komplexität eines multikriteriellen Problems für das menschliche Urteilsvermögen zu hoch ist, um rein tabellarisch gelöst werden zu können, sollten Tabellen nicht ohne weiteres durch graphische Elemente ersetzt bzw. um solche ergänzt werden. Eine Graphik, die im Vergleich zur Tabelle kognitiv weniger anstrengende wahrnehmende Informationsverarbeitungsprozesse unterstützt, ist bei einem symbolischen Informationsauswertungsproblem nur dann vorteilhaft, wenn die Anforderungen an die analytische Informationsverarbeitung die Informationsverarbeitungskapazität des Entscheiders stark beanspruchen oder sogar übersteigen (vgl. Vessey, 2006, S. 149). Solche Probleme können dann als limitierende Probleme eingestuft werden. Demgegenüber verbleiben nicht nur Entscheidungsprobleme mit niedriger sondern eben auch mit hoher Komplexität, zu deren Lösung sich die Aufbereitung von Daten in tabellarischer Form besser eignet, solange nicht von einem limitierenden Problem ausgegangen werden muß. Eine objektive Abgrenzung eines limitierenden Problems ist freilich schwer möglich. Entscheidend ist daher der Hinweis, nicht zu früh von limitierender Komplexität auszugehen. Die Ergebnisse der vorliegenden Studie zeigen, dass eine Tabelle zu einem konsistenten Entscheidungsprozess führt, eine Graphik lediglich zu einer schnelleren Problemlösung. Nach Vessey (2006, S. 149) ist das vorliegende Problem daher als komplex, allerdings noch nicht als limitierend einzustufen. Daraus lässt sich schließen, dass die Komplexität des Problems noch analytisch beherrschbar, eine rein tabellarische Darstellung mithin vorzuziehen ist und die Grenze, ab der eine Graphik vorteilhaft ist, noch nicht erreicht wurde. Der Wahl des Darstellungsformates sollte daher stets eine Einschätzung der Problemklasse vorausgehen.

Schließlich bleibt zu berücksichtigen, dass die Ergebnisse der Studie Limitationen unterworfen sind, die das Ausmaß, bis zu dem die Ergebnisse verallgemeinert werden können beschränken, zugleich aber Ausgangspunkte künftiger Arbeiten sein könnten.

In dieser Studie wurde die experimentelle Aufgabe bewußt auf Studierende abgestimmt, sodass davon ausgegangen werden konnte, dass die Probanden mit dem Problemkomplex vertraut sind und über das zur Lösung notwendige Wissen („domain knowledge“) verfügen. Dennoch können Unterschiede zwischen Studierenden und Fachleuten nicht ausgeschlossen werden. Studien, die auf erfahrene Fachleute aus der Praxis für vergleichbare Entscheidungsprobleme zurückgreifen, stellen jedoch regelmäßig keine Unterschiede zwischen beiden Gruppen fest (vgl. Eggleton et al., 1992; Stock/Watson,

1984; Moriarity, 1979). Eine Durchführung der Studie mit Managern würde jedoch die explizite Berücksichtigung einer unternehmerischen Aufgabenstellung erlauben, um so die Robustheit der Ergebnisse zu prüfen und die Grundgesamtheit von betrieblichen Entscheidungsträgern umfassender abdecken zu können.

Bestehende Studien (vgl. u. a. Payne, 1976; Newell/Simon, 1972) nutzen Verbalprotokolle im direkten Anschluß an die Lösung eines Entscheidungsproblems, um Probanden nach der Wahl ihrer Entscheidungsregel bzw. Heuristik zu befragen. Diese Vorgehensweise kann Hinweise darauf geben, ob die Probanden bei der faktischen Entscheidung tatsächlich ein additives Modell verwenden. In der vorliegenden Studie sollten die Teilnehmenden direkt im Anschluß an die faktische Reihung paarweise Vergleiche durchführen. Dieser Ablauf wurde bewußt gewählt, um Störfaktoren zwischen den Prozessschritten ausschließen zu können, welche die Rekonstruktion der Präferenzen auf Basis der Paarvergleiche hätten beeinträchtigen können.

Hinsichtlich der Problemstruktur liegt dieser Studie ein zwar komplexes, aber grundsätzlich gut strukturiertes hierarchisches Problem zugrunde. Dies mag in der betrieblichen Praxis womöglich die kleinere Klasse von Entscheidungsproblemen darstellen. Da aber in dem untersuchten Entscheidungsproblem mehrere Alternativen gegeben, verschiedene Attribute zu beurteilen und zu bewerten sind, d.h. in Bezug auf den Schwierigkeitsgrad zweifellos ein komplexes Problem vorliegt, wird jedenfalls innerhalb der Klasse der gut strukturierten Probleme eine Subklasse untersucht, die in der betrieblichen Praxis regelmäßig anzutreffen und zweifellos als relevant einzustufen ist. Die hierarchische Struktur impliziert, dass sich Attribute gegenseitig ausschließen und überschneidungsfrei sind. Erst diese Eingrenzung und die Annahme der Vollständigkeit hinsichtlich der Oberkriterien, Unterkriterien und Alternativen erlaubt den Rückgriff auf ein additives multiattributives Modell zur Rekonstruktion der Artenpräferenzrelationen mittel AHP, zur Ermittlung der Gesamtnutzenwerte und der sich anschließenden Konsistenzprüfung.

6 Fazit

Die „Verlockungen“ graphischer Darstellungsformate liegen auf der Hand: Sie sehen regelmäßig ansprechend aus und bieten vielfältige Darstellungsvarianten. Hinsichtlich der Verbesserung der Entscheidungsgüte durch die visuelle Aufbereitung der Information zeigt sich im Rahmen der vorliegenden Studie anhand eines komplexen multikriteriellen Entscheidungsproblems allerdings ein ernüchterndes Bild: Visualisierung

kann nicht zur Sicherstellung prozeduraler Rationalität beitragen; Informationsaggregation hingegen ist sehr wohl ein probates Mittel, welches den Zweck der Verbesserung des Entscheidungsprozesses erfüllt. Aggregierte Information macht den Entscheidungsprozess konsistenter und schneller. Die Transitivität des Entscheidungsergebnisses, dessen Optimum im Gegensatz zu Vorläuferstudien nicht objektiv festgelegt, sondern durch Rekonstruktion subjektiver Artenpräferenzrelationen bestimmt wurde, wird allerdings weder durch Visualisierung noch durch Informationsaggregation signifikant beeinflusst.

Aus Sicht von Theorie und Praxis ist die Studie relevant, da sie unter anderem Empfehlungen zur Gestaltung von Management-Informationssystemen geben kann. Insbesondere in diesem Kontext werden häufig Kombinationen aus Tabelle und Graphik reinen Darstellungsformaten gegenüber vorgezogen. Die vorliegenden Ergebnisse sind daher bedeutsam vor allem für die betriebliche Entscheidungsunterstützung (z.B. das Controlling) bzw. entsprechende Entscheidungsunterstützungssysteme (z.B. Management-Informationssysteme), deren Aufgabe es ist, entscheidungsrelevante Informationen zu gewinnen, aufzubereiten und adäquat darzustellen. Dazu gehört die Identifizierung der Problemklasse, die Aggregation der Daten und die Wahl des Darstellungsformats. Dieses sollte weder dem Zufall, noch bloßer Intuition, irgendwelchen Moden oder habituellem Verhalten überlassen werden.

Denn zur Sicherstellung der Rationalität im Entscheidungsprozess kann empfohlen werden, dem Entscheidungsträger für die vorliegende Problemklasse angemessen aggregierte Informationen vorzulegen. Entgegen aktueller Trends ist einer konventionellen tabellarischen Darstellungsform solange der Vorzug zu geben, mithin also den „nackten“ Zahlen zu trauen, als nicht von limitational-komplexen Problemen auszugehen ist. Erst ab diesem Punkt ist zu erwarten, dass graphische Formate ihre Vorzüge entfalten können. Dies zu untersuchen, bleibt freilich nachfolgenden Studien ebenso überlassen, wie die Diskussion der Frage, inwieweit rationale Entscheidungsprozesse tatsächlich zu besseren Entscheidungen führen. Zwar hängt die Interpretation, Bewertung und Beurteilung der zur Verfügung gestellten Informationen und damit letztlich das Entscheidungsergebnis offenkundig von der vorgelagerten Entscheidungsunterstützungsfunktion ab, liegt aber doch stets im subjektiven Aufgabenbereich des jeweiligen Entscheidungsträgers.

Anhang

Anhang 1: Entscheidungsvorlage als Tabelle mit hoher Aggregation

Bewertungskategorien	Smartphones				
	A	B	C	D	E
Telefon	2,0	2,3	2,1	2,5	2,2
Kamera	2,8	3,0	2,3	2,1	3,0
Musikspieler	1,8	2,0	1,7	1,8	2,7
Internet und PC	2,3	3,3	1,5	2,4	2,8
GPS	2,0	2,9	1,8	2,1	1,5
Handhabung	2,4	2,9	1,8	2,4	2,1
Stabilität	1,9	2,5	2,2	1,4	1,3
Akku	2,7	2,0	3,0	2,1	2,0

Anhang 2: Entscheidungsvorlage als Tabelle mit niedriger Aggregation

Bewertungskategorien	Smartphones				
	A	B	C	D	E
Telefon					
Sprachqualität	2,3	1,7	2,1	2,0	2,2
Netzempfindlichkeit	1,7	2,5	1,9	3,0	2,0
SMS	2,0	2,7	2,3	2,5	2,4
Kamera					
Fotos bei normaler Beleuchtung	3,5	2,9	1,9	2,4	4,5
Fotos bei geringer Beleuchtung	3,5	2,7	2,0	2,3	2,5
Auslöseverzögerung	1,5	2,7	2,3	1,4	1,5
Video	2,5	3,7	3,0	2,3	3,5
Musikspieler					
Klang	2,2	2,5	2,0	2,2	2,9
Kompatibilität	1,4	1,5	1,4	1,4	2,5
Internet und PC					
Surfen	2,3	4,3	1,3	3,3	3,3
E-Mail	1,5	2,9	1,2	1,5	2,4
Synchronisation	3,1	2,7	2,0	2,4	2,7
GPS					
Schnelligkeit & Genauigkeit der Ortung	2,0	2,9	1,8	2,1	1,5
Handhabung					
Gebrauchsanleitung und Inbetriebnahme	3,5	2,3	1,8	4,0	2,3
Display	1,7	3,5	1,6	1,7	2,0
Tastatur	1,9	3,3	1,9	1,9	2,5
Menü und Bedienung	1,9	3,1	1,0	1,7	2,0
Transport	3,0	2,3	2,7	2,7	2,7
Stabilität					
Bruchfestigkeit, Feuchtigkeitstest, Solidität	1,9	2,5	2,2	1,4	1,3
Akku					
Betriebs-, Ladedauer, Lademöglichkeiten	2,7	2,0	3,0	2,1	2,0

Anhang 3: Entscheidungsvorlage als Graphik mit hoher Aggregation

Bewertungskategorien	Smartphones				
	A	B	C	D	E
Telefon	■	■	■	■	■
Kamera	■	■	■	■	■
Musikspieler	■	■	■	■	■
Internet und PC	■	■	■	■	■
GPS	■	■	■	■	■
Handhabung	■	■	■	■	■
Stabilität	■	■	■	■	■
Akku	■	■	■	■	■

Anhang 4: Entscheidungsvorlage als Graphik mit niedriger Aggregation

Bewertungskategorien	Smartphones				
	A	B	C	D	E
Telefon					
Sprachqualität	■	■	■	■	■
Netzempfindlichkeit	■	■	■	■	■
SMS	■	■	■	■	■
Kamera					
Fotos bei normaler Beleuchtung	■	■	■	■	■
Fotos bei geringer Beleuchtung	■	■	■	■	■
Auslöseverzögerung	■	■	■	■	■
Video	■	■	■	■	■
Musikspieler					
Klang	■	■	■	■	■
Kompatibilität	■	■	■	■	■
Internet und PC					
Surfen	■	■	■	■	■
E-Mail	■	■	■	■	■
Synchronisation	■	■	■	■	■
GPS					
Schnelligkeit & Genauigkeit der Ortung	■	■	■	■	■
Handhabung					
Gebrauchsanleitung und Inbetriebnahme	■	■	■	■	■
Display	■	■	■	■	■
Tastatur	■	■	■	■	■
Menü und Bedienung	■	■	■	■	■
Transport	■	■	■	■	■
Stabilität					
Bruchfestigkeit, Feuchtigkeitstest, Solidität	■	■	■	■	■
Akku					
Betriebs-, Ladedauer, Lademöglichkeiten	■	■	■	■	■

Anhang 5: Entscheidungsvorlage als Kombination aus Tabelle und Graphik mit hoher Aggregation

Bewertungskategorien	Smartphones									
	A	B	C	D	E	A	B	C	D	E
Telefon	2,0	2,3	2,1	2,5	2,2					
Kamera	2,8	3,0	2,3	2,1	3,0					
Musikspieler	1,8	2,0	1,7	1,8	2,7					
Internet und PC	2,3	3,3	1,5	2,4	2,8					
GPS	2,0	2,9	1,8	2,1	1,5					
Handhabung	2,4	2,9	1,8	2,4	2,1					
Stabilität	1,9	2,5	2,2	1,4	1,3					
Akku	2,7	2,0	3,0	2,1	2,0					

Anhang 6: Entscheidungsvorlage als Kombination aus Tabelle und Graphik mit niedriger Aggregation

Bewertungskategorien	Smartphones									
	A	B	C	D	E	A	B	C	D	E
Telefon										
Sprachqualität	2,3	1,7	2,1	2,0	2,2					
Netzempfindlichkeit	1,7	2,5	1,9	3,0	2,0					
SMS	2,0	2,7	2,3	2,5	2,4					
Kamera										
Fotos bei normaler Beleuchtung	3,5	2,9	1,9	2,4	4,5					
Fotos bei geringer Beleuchtung	3,5	2,7	2,0	2,3	2,5					
Auslöseverzögerung	1,5	2,7	2,3	1,4	1,5					
Video	2,5	3,7	3,0	2,3	3,5					
Musikspieler										
Klang	2,2	2,5	2,0	2,2	2,9					
Kompatibilität	1,4	1,5	1,4	1,4	2,5					
Internet und PC										
Surfen	2,3	4,3	1,3	3,3	3,3					
E-Mail	1,5	2,9	1,2	1,5	2,4					
Synchronisation	3,1	2,7	2,0	2,4	2,7					
GPS										
Schnelligkeit & Genauigkeit der Ortung	2,0	2,9	1,8	2,1	1,5					
Handhabung										
Gebrauchsanleitung und Inbetriebnahme	3,5	2,3	1,8	4,0	2,3					
Display	1,7	3,5	1,6	1,7	2,0					
Tastatur	1,9	3,3	1,9	1,9	2,5					
Menü und Bedienung	1,9	3,1	1,0	1,7	2,0					
Transport	3,0	2,3	2,7	2,7	2,7					
Stabilität										
Bruchfestigkeit, Feuchtigkeitstest, Solidität	1,9	2,5	2,2	1,4	1,3					
Akku										
Betriebs-, Ladedauer, Lademöglichkeiten	2,7	2,0	3,0	2,1	2,0					

Verzeichnis der zitierten Literatur

- Abdel-Khalik, A. Rashad (1973): The Effect of Aggregating Accounting Reports on the Quality of the Lending Decision: An Empirical Investigation. In: Journal of Accounting Research 11 (1973), S. 104-138.
- Anagol, Santosh/Gamble, Keith J. (2013): Does Presenting Investment Results Asset by Asset Lower Risk Taking? In: Journal of Behavioral Finance 14 (2013), No. 4, S. 276-300.
- Bamberg, Günter/Coenenberg, Adolf G./Krapp, Michael (2012): Betriebswirtschaftliche Entscheidungslehre. 15. Aufl. München 2012.
- Barefield, Russell M. (1972): The Effect of Aggregation on Decision Making Success: A Laboratory Study. In: Journal of Accounting Research, Vol. 10 (1972), No. 2, S. 229-242.
- Bassler, Anna (2010): Die Visualisierung von Daten im Controlling. Köln 2010.
- Benbasat, Izak/Dexter, Albert S. (1986): An Investigation of the Effectiveness of Color and Graphical Information Presentation Under Varying Time Constraints. In: MIS Quarterly, Vol. 10 (1986), No. 1, S. 59-83.
- Beshears, John/Choi, James/Laibson, David/Madrian, Brigitte (2011): Can psychological aggregation manipulations affect portfolio risk-taking? Evidence from a framed field experiment. In: NBER Working Paper (2011), (w16868).
- Bolles, Edmund B. (1991): A Second way of knowing: The riddle of human perception: Prentice Hall Press 1991.
- Bonner, Sarah E. (2008): Judgment And Decision Making in Accounting. Upper Saddle River, New Jersey 2008.
- Buckless, Frank A./Pickard Ravenscroft, Sue (1990): Contrast Coding: A Refinement of ANOVA in Behavioral Analysis. In: The Accounting Review 65 (1990), No. 4, S. 933-945.
- Cardinaels, Eddy (2008): The interplay between cost accounting knowledge and presentation formats in cost-based decision-making. In: Accounting, Organizations and Society 33 (2008), No. 6, S. 582-602.
- Casey, Cornelius J. JR. (1980): Variation in Accounting Information Load: The Effect on Loan Officers' Predictions on Bankruptcy. In: The Accounting Review, Vol. 55 (1980), No. 1, S. 36-49.
- Chan, Siu Y. (2001): The use of graphs as decision aids in relation to information overload and managerial decision quality. In: Journal of Information Science, Vol. 27 (2001), No. 6, S. 417-425.
- Chervany, Norman L./Dickson, Gary W. (1974): An Experimental Evaluation of Information Overload in a Production Environment. In: Management Science, Vol. 20 (1974), No.10, S. 1335-1344.
- Dennis, Alan R./Carte, Traci A. (1998): Using Geographical Information Dystems for Decision Making: Extending Cognitive Fit Theory to Map-Based Presentations. In: Information Systems Research, Vol. 9 (1998), No. 2, S. 194-203.
- Desanctis, Gerardine (1984): Computer Graphics as Decision Aids: Directions for Research. In: Decision Sciences, Vol. 15 (1984), No. 4, S. 463-487.


- Driver, Michael J./Streufer, Siegfried (1969): Integrative complexity: An approach to individuals and groups as information-processing systems. In: *Administrative Science Quarterly* Vol. 14 (1969), No. 2, S. 272-285.
- Dyer, James S./Fishburn, Peter C./Steuer, Ralph E./Wallenius, Jyrki/Zionts, Stanley (1992): Multiple criteria decision making, multiattribute utility theory: the next ten years. In: *Management Science* 38 (1992), No. 5, S. 645-654.
- Eggleton, I. R.C./Knipping, P./McCulloch, B. (1992): Schematic faces and the communication of multivariate financial information in a bankruptcy prediction task. Unpublished Working Paper. University of Western Australia 1992.
- Einhorn, Hillel J./Hogarth, Robin M. (1981): Behavioral Decision Theory: Processes of Judgment and Choice. In: *Journal of Accounting Research*, Vol. 19 (1981), No. 1, S. 1-31.
- Eisenführ, Franz/Weber, Martin/ Langer, Thomas (2010): *Rationales Entscheiden*. 5. überarbeitete und erweiterte Aufl. Heidelberg 2010.
- Eppler, Martin J./Mengis, Jeanne (2004): The Concept of Information Overload - A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. In: *The Information Society: An International Journal*, Vol. 20 (2004), No. 5, S. 1-20.
- Fischbacher, Urs (2007): z-Tree: Zurich toolbox for ready-made economic experiments. In: *Experimental Economics*, Vol. 10 (2007), S. 171-178.
- Frownfelter-Lohrke, Cynthia (1998): The effects of differing information presentations of general purpose financial statements on users' decisions. In: *Journal of Information Systems* 12 (1998), No. 2, S. 99-107.
- Gupta, Mahendra/King, Ronald R. (1997): An Experimental Investigation of the Effect of Cost Information and Feedback on Product Cost Decisions*. In: *Contemporary Accounting Research* 14 (1997), No. 1, S. 99-127.
- Gigerenzer, Gerd/Gaissmaier, Wolfgang (2011). Heuristic Decision Making. *Annual Review of Psychology*, Nr. 62, S. 451-482.
- Harvey, David W./Rhode, John G./Merchant, Kenneth A. (1979): Accounting aggregation: User Preferences and Decision Making. In: *Accounting, Organizations and Society*, Vol. 4 (1979), No. 3, S. 187-210.
- Hirsch, Bernhard/Volnhals, Martina (2012): Information Overload im betrieblichen Berichtswesen - ein unterschätztes Phänomen. In: *Die Betriebswirtschaft*, 72. Jg (2012), S. 21-54.
- IBM Corporation (2012): *IBM SPSS Advanced Statistics 21* 2012.
- Iselin, Errol R. (1988): The Effects of Information Load and Information Diversity on Decision Quality in a Structured Decision Task. In: *Accounting, Organizations and Society*, Vol. 13 (1988), S. 147-164.
- Jacoby, Jacob/Speller, Donald E./Berning Kohn, Carol A. (1974a): Brand Choice Behavior as a Function of Information Load: Replication and Extension. In: *Journal of Consumer Research*, Vol. 1 (1974), S. 33-42.
- Jacoby, Jacob/Speller, Donald E./Kohn, Carol A. (1974b): Brand Choice Behavior as a Function of Information Overload. In: *Journal of Marketing Research*, Vol. 11 (1974), No. 1, S. 63-69.

- Kahneman, Daniel/ Tversky, Amos (1972): Subjective probability: A judgment of representativeness. In: *Cognitive Psychology*, Vol. 3 (1972), No. 3, S. 430-454.
- Kahneman, Daniel/Tversky, Amos (1996): On the reality of cognitive illusions. In: *Psychological Review* 103 (1996), No. 3, S. 582-591.
- Kaufmann, Christine/Weber, Martin (2013): Sometimes less is more – The influence of information aggregation on investment decisions. In: *Journal of Economic Behavior & Organization* 95 (2013), S. 20-33.
- Kelton, Andrea S./Pennington, Robin R./Tuttle, Brad M. (2010): The effects of information presentation format on judgment and decision making: A review of the information systems research. In: *Journal of Information Systems* 24 (2010), No. 2, S. 79-105.
- Larkin, Jill H./Simon, Herbert A. (1987): Why a diagram is (sometimes) worth ten thousand words. In: *Cognitive science* 11 (1987), No. 1, S. 65-100.
- Libby, Robert (1981): *Accounting and Human Information Processing: Theory and Applications*. EnglewoodCliffs, NJ: Prentice Hall 1981.
- Libby, Robert/Bloomfield, Robert/Nelson, Mark W. (2002): Experimental research in financial accounting. In: *Accounting, Organizations and Society* Vol. 27 (2002), S. 775-810.
- Lucas, Henry C. JR./Nielsen, Norman R. (1980): The Impact of the Mode of Information Presentation on Learning and Performance. In: *Management Science*, Vol. 26 (1980), No. 10, S. 982-993.
- Luce, Robert Duncan/Raiffa, Howard (1957): *Games and decisions: Introduction and critical survey*. New York 1957.
- Moriarty, Shane (1979): Communicating financial information through multidimensional graphics. In: *Journal of Accounting Research*, Vol. 17 (1979), No. 1, S. 205-224.
- Neumann, John von/Morgenstern, Oskar (1944): *The Theory of Games and Economic Behavior*. Princeton: University Press 1944.
- Neumann, John von/Morgenstern, Oskar (1947): *The Theory of Games and Economic Behavior*. 2.Aufl. Princeton: University Press 1947.
- Newell, Allen/Simon, Herbert A. (1972): *Human problem solving*. Englewood Cliffs, New Jersey 1972.
- Obermaier, Robert/Müller, Felix (2008): Management accounting research in the lab – method and applications. In: *Zeitschrift für Planung & Unternehmenssteuerung*, 19 Jg (2008), S. 325-351.
- Obermaier, Robert/Saliger, Edgar (2013): *Betriebswirtschaftliche Entscheidungstheorie*. 6. Aufl. München 2013.
- Otley, David T./Dias, Francisco J. B. (1982): Accounting Aggregation and Decision-Making Performance: An Experimental Investigation. In: *Journal of Accounting Research*, Vol. 20 (1982), No. 1, S. 171-188.
- Payne, John W. (1976): Task complexity and Contingent Processing in Decision Making: An information Search and Protocol Analysis. In: *Organizational behavior and human performance*, Vol. 16 (1976), No. 2, S. 366-387.

- Payne, John W. (1982): Contingent decision behavior. In: Psychological Bulletin 92 (1982), No. 2, S. 382-402.
- Ronen, Joshua (1971): Some Effects of Sequential Aggregation in Accounting on Decision-Making. In: Journal of Accounting Research, Vol. 9 (1971), No. 2, S. 307-332.
- Saaty, Thomas L. (1980): The analytic hierarchy process: Planning, priority setting, resource allocation. New York 1980.
- Saaty, Thomas L. (1990): How to make a decision: The Analytic Hierarchy Process. In: European Journal of Operational Research, Vol. 48 (1990), No. 1, S. 9-26.
- Saaty, Thomas L. (1994): How to Make a Decision: The Analytic Hierarchy Process. In: Interfaces, Vol. 24 (1994), No. 6, S. 19-43.
- Schneeweiß, Christoph (1991): Planung 1. Berlin 1991.
- Schroder, Harold M./Driver, Michael J./Streufert, Siegfried (1967): Human Information Processing. New York 1967.
- Shields, Michael D. (1983): Effects on Information Supply and Demand on Judgment Accuracy: Evidence from Corporate Managers. In: The Accounting Review, Vol. 58 (1983), No. 2, S. 284-303.
- Simon, Herbert A. (1957): Models of Man: Social and Rational. New York 1957.
- Simon, Herbert A./Newell, Allen (1958): Heuristic problem solving: The next advance in operations research. In: Operations research 6 (1958), No. 1, S. 1-10.
- Simon, Herbert A. (1976): From substantive to procedural rationality. Method and Appraisal in Economics. Cambridge et al. 1976.
- Simon, Herbert A. (1979): Rational Decision Making in Business Organizations. In: American Economic Review, Vol. 69 (1979), No. 4, S. 493-513.
- Simon, Herbert A./Lea, Glenn (1974): Problem solving and rule induction: A unified view. In: Knowledge and Cognition 1974, S. 105-128.
- Smelcer, John B./Carmel, Erran (1997): The Effectiveness of Different Representations for Managerial Problem Solving: Comparing Tables and Maps. In: Decision Sciences, Vol. 28 (1997), S. 391-420.
- Smith, Malcolm/Taffler, Richard (1995): The incremental effect of narrative accounting information in corporate annual reports. In: Journal of Business Finance & Accounting, Vol. 22 (1995), No. 8, S. 1195-1210.
- So, Stella/Smith, Malcolm (2004): Multivariate decision accuracy and the presentation of accounting information. Corporate financial communication and voluntary disclosure. In: Accounting Forum 28 (2004), No. 3, S. 283-305.
- Speier, Cheri (2006): The influence of information presentation formats on complex task decision-making performance. In: International Journal of Human-Computer Studies 64 (2006), No. 11, S. 1115-1131.
- Speier, Cheri/Morris, Michael G. (2003): The influence of query interface design on decision-making performance. In: MIS Quarterly 27 (2003), No. 3, S. 397-423.
- Stiftung Warentest (2011): „Smartphones für alle“. In: Stiftung Warentest, Nr. 5 (2011).

- Stock, Duane/Watson, Collin J. (1984): Human judgment accuracy, multidimensional graphics, and humans versus models. In: *Journal of Accounting Research* 22 (1984), No. 1, S. 192-206.
- Tiessen, Peter (1976): An investigation of some effects of data aggregation and cognitive style on decision making performance. Dissertation. Minnesota 1976.
- Tufte, Edward R. (2001): *The Visual Display of Quantitative Information*. Cheshire, Connecticut 2001.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, Nr. 211, Heft 4481, S. 453-458.
- Umanath, Narayan S./Vessey, Iris (1994): Multiattribute Data Presentation and Human Judgment: A Cognitive Fit Perspective. In: *Decision Sciences* 25 (1994), No. 5-6, S. 795-824.
- Vessey, Iris (1991): Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature. In: *Decision Sciences*, Vol. 22 (1991), S. 219-240.
- Vessey, Iris (1994): The effect of information presentation on decision making: A cost-benefit analysis. In: *Information & Management*, Vol. 27 (1994), No. 2, S. 103-119.
- Vessey, Iris (2006): The Theory of Cognitive Fit. In: Zhang, Ping/Galletta, Dennis (Hrsg.): *Human-Computer Interaction and Management Information Systems: Foundations*. Vol. 5. New York 2006, S. 141-183.
- Wallenius, Jyrki/Dyer, James S./Fishburn, Peter C./Steuer, Ralph E./Zionts, Stanley/Deb, Kalyanmoy (2008): Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. In: *Management Science* 54 (2008), No. 7, S. 1336-1349.
- Waller, William S./Shapiro, Brian/Sevcik, Galen (1999): Do cost-based pricing biases persist in laboratory markets? In: *Accounting, Organizations and Society* 24 (1999), No. 8, S. 717-739.


The Effects of Information Aggregation and Visualization on Judgment Quality for Complex Multiattribute Judgment Tasks in Performance Evaluation

Christian Meier / Robert Obermaier / Tamara Jakob 

Abstract

Performance evaluation is one of the most prominent multiple criteria decision-making applications. Cognitive limitations of boundedly rational decision makers increase the risk of adversely evaluating performance, which can have severe consequences. Therefore, it is necessary to facilitate judgments in performance evaluation. Empirical studies have found presentation format and information aggregation to increase judgment quality when implemented properly. However, the interplay of information aggregation and presentation format has not been analyzed so far. Hence, this study examines the interplay of information aggregation and presentation format on judgment quality for a complex multiattribute judgment task in performance evaluation. Cognitive fit theory and information-processing theory are employed to derive the hypotheses. An experiment is conducted where both information aggregation and presentation format are varied between-subjects. Results show that graphs, either with or without data values, lead to the most accurate judgments. Judgments based on tables are less accurate. “Pure” graphs result in a more consistent judgment process compared to graphs with data values. However, graphs without or with data values do not lead to a higher procedural consistency compared to tables. Presentation format does not influence procedural speed. Information aggregation without loss of information content increases accuracy, consistency and speed of judgments. Increasing information aggregation, graphs lead to more accurate and consistent judgments and graphs with data values lead to more accurate judgments. Accuracy and consistency with tables does not differ with a change in information aggregation. For graphs with data values and low information aggregation the judgment process takes significantly more time, whereas for graphs and tables procedural speed does not change. Overall, findings support information-processing theory and provide mixed evidence for cognitive fit theory. The results are of interest for providers of information, e.g. designers of management information systems, as well as recipients of information, e.g. managers and auditors.

Keywords: Information aggregation, problem representation, information processing, performance evaluation, multiattribute judgments, judgment quality.

 Christian Meier and Robert Obermaier: University of Passau, Chair of Accounting and Control, Innstraße 27, 94032 Passau, Germany, E-Mail: controlling@uni-passau.de; Phone: 0851-509-3271. Tamara Jakob: Lämmersreut 25, 94065 Waldkirchen, E-Mail: tamijak91@gmx.de.

1 Introduction

Performance evaluation is one of the most prominent multiple criteria decision-making applications (Hülle et al. 2011). Multi-criteria decision problems are a class of problems with high theoretical and practical relevance (Wallenius et al. 2008, Dyer et al. 1992). They can be solved employing normative analytic models or using human judgment (Stock and Watson 1984, Libby 1981). Multiattribute judgments are regularly performed by boundedly rational decision makers with limits on their abilities to perceive new problems, to remember important facts and to process information properly (Newell and Simon 1972). Generally, performance evaluation judgment tasks with multiple criteria are inherently complex, as a substantial number of information cues have to be processed and an increasing cognitive load has to be evaluated (Speier 2006, Payne 1982). If the number of criteria or alternatives exceeds short-term memory, information processing decreases, problem-solving performance is reduced (Schroder et al. 1967), and the risk of adversely evaluating performance increases. Consequently, employees will react negatively when equally good performances are evaluated differently and resource allocation can be distorted when departments or products are evaluated inaccurately (Merchant and van der Stede 2012).

Therefore, it is necessary to alter the structure of the task environment in desirable ways to facilitate complex multiattribute judgments (Bonner 2008). Aggregation of complex information and choice of presentation format are both ways to modify the task environment and to improve judgment quality (Tufte 2001, DeSanctis 1984). Effects of presentation format on judgment and decision quality in the context of decision making with graphs vs. tables have been studied extensively over the last decades. A central finding of these studies is that presentation format facilitates judgment quality when it fits to the information-processing requirements of the underlying task (Yigitbasioglu and Velcu 2012, Kelton et al. 2010, Vessey 2006). Behavioral effects of information aggregation have also been focus of several studies (Kaufmann and Weber 2013, Otley and Dias 1982, Harvey et al. 1979, Tiessen 1976, Chervany and Dickson 1974, Ronen 1971). A central result is that judgment and decision-making quality increases with information aggregation, if the information content is held constant, and decreases with information aggregation with loss of information content. Furthermore, the interplay of task complexity and presentation format has been focus of prior studies (Speier 2006, So and Smith 2004, Umanath and Vessey 1994). However, the interplay of information aggregation and presentation format for a complex multiattribute judgment task has not

been examined so far. Cognitive fit theory (Vessey 2006, 1994, 1991) predicts that graphs possess integrating capabilities that mitigate adverse effects of increased information load on judgment quality, whereas tables do not (Umanath and Vessey 1994). Accordingly, accuracy of judgments based on tables should benefit from information aggregation. While a growing body of empirical evidence supports the fundamentals of cognitive fit theory for simple tasks, extensions to more complex tasks are sparse (Speier 2006). Thus, to facilitate high judgment quality in performance evaluation, it is necessary to examine whether and how information aggregation and presentation format influence judgment quality for the problem class of complex multiattribute judgment tasks.

Also, prior studies analyzing effects of information aggregation and presentation format focus on result accuracy as an aspect of judgment quality and neglect procedural consistency. However, it is important to examine procedural consistency in a multi-criteria decision problem because consistency of subjects' preferences for the multiple decision criteria is an antecedent to procedural rationality (Simon 1976). Moreover, consistency is important when evaluating the performance of organizational units, as managers may be required to specify weights for various performance measures and inconsistent judgments may distort the assessment of organizational performance (Merchant and van der Stede 2012, Tuttle and Kershaw 1998).

To address these issues, this study examines the effects of information aggregation and presentation format on judgment quality for a complex multiattribute judgment task in performance evaluation. An experiment is conducted where both information aggregation and presentation format are varied between-subjects. The following research question is examined: Do information aggregation and presentation format influence judgment quality for a complex multiattribute judgment task? Judgment quality is measured both from a performance and a procedural perspective with the criteria of result accuracy, procedural consistency, and procedural speed. Cognitive fit theory and information-processing theory (Driver and Streufert 1969, Schroder et al. 1967) as aspects of a general theory of problem solving are used to derive the hypotheses. The results are of interest for providers of information, e.g. management accountants and designers of management information systems and management control systems, as well as recipients of information, e.g. managers and auditors.

Results show that graphs as a presentation format, either with or without data values, result in the most accurate judgments. Judgments based on tables are less accurate. “Pure” graphs result in a more consistent judgment process compared to graphs with data values. However, neither graphs nor graphs with data values lead to a higher procedural consistency compared to tables, and procedural speed is not influenced by presentation format. Information aggregation without loss of information content increases accuracy, consistency, and speed of judgments. Increasing information aggregation, graphs lead to more accurate and consistent judgments and graphs with data values lead to more accurate judgments. Accuracy and consistency with tables do not differ significantly with a change in information aggregation. For graphs with data values and low information aggregation the judgment process takes significantly more time, whereas for graphs and tables procedural speed does not change.

The contributions of this study are threefold. First, this study extends prior literature analyzing complex multiattribute judgment problems and contributes to examining effects of information processing from a procedural rationality perspective (Simon 1976). Prior studies predominantly base judgment quality on substantive rationality, i.e. they evaluate result accuracy in comparing human judgments to a prespecified optimal alternative. However, frequently no optimal alternative exists for a complex multi-criteria decision problem without further specifying criteria weights. This makes it necessary to base judgment quality on procedural rationality. Thus, in this study, the analytic hierarchy process (AHP) (Saaty 1994, 1990, 1980) is utilized to reveal consistency of participants’ subjective weights for the multiple decision criteria during the judgment process. Second, from a theoretical perspective, this study contributes to prior literature in examining the interplay of presentation format and information aggregation for complex tasks. The analysis of complex tasks is an important aspect of further developing the theory of cognitive fit towards an aspect of a general theory of problem solving (Vessey 2006). Although extensive empirical evidence supports for the theory of cognitive fit for simple tasks, in this study cognitive fit theory fails to predict the interplay of presentation format and information aggregation. Third, from a practical perspective, in examining effects of pure and combined presentation formats, this study contributes to the design of management information and control systems. It contradicts the widespread belief that combined presentation formats should be superior to pure formats. For complex multiattribute judgments, providers and designers of management information and con-

trol systems should keep it short and simple, that is, adequately aggregate information and refrain from adding data values to graphs.

This paper proceeds as follows: The next section develops hypotheses based on cognitive fit theory and information-processing theory. The two subsequent sections describe the experimental design and results. The final section provides a discussion of the results including their limitations and implications.

2 Theory and Hypotheses

2.1 General Theory of Problem Solving

The theory of human problem solving (Newell and Simon 1972) proposes that boundedly rational decision makers use a serial information-processing system with limited short-term memory. They extract information from the task environment to create a mental representation in the internal problem space which, by using a strategy to perform elementary information-processing tasks (EIP's), builds the basis for evaluation and heuristic problem solving. The relative difficulty of problem solving will depend on how effective the decision maker has been in representing task characteristics in his mental representation (Simon 1978a). As a decision maker's ability to process huge amounts of complex data is limited (Newell and Simon 1972, Schroder et al. 1967), task characteristics should be modified to extend the limits of human cognition by decreasing task complexity and hence increase judgment quality (Tuttle and Kershaw 1998).

According to Payne (1982) there are multiple task characteristics pertaining to the task environment that may influence effort required to make a judgment or decision, which in turn can cause decision makers to change strategy. Among those task characteristics examined in problem-solving tasks are: (1) task complexity (e.g. number of alternatives and or dimensions); (2) presentation format (e.g. graphs or tables); and (3) response mode (e.g. judgment or choice).

If the short-term memory constraint is exceeded because of task complexity, i.e. the number of alternatives or criteria exceeds short-term memory, information processing decreases and problem-solving performance is reduced (Schroder et al. 1967). Also, a mismatch of presentation format and task type hinders performance (Vessey 2006, 1994, 1991). Finally, response mode influences demands on information processing. Judgments in accounting can, for example, appear in the form of predictions about a future state of affairs or events (e.g. bankruptcy prediction) or as an evaluation of cur-

rent state of affairs or events (e.g. performance evaluation). Choices are a part of decision making which typically follows judgments (Bonner 2008). Therefore, judgment usually facilitates choice; however, according to common sense it is not necessary and not sufficient for choice. Choice only depends on selection, whereas a requirement for judgment is an evaluation of alternatives. Judgment in general is more cognitively demanding and time-consuming and involves more comprehensive consideration of each problem dimension than choice (Tuttle and Kershaw 1998, Einhorn and Hogarth 1981).

Judgment quality can be evaluated based on (1) substantive rationality, i.e. the extent to which judgment output (e.g. a final judgment), corresponds with some “right judgment”, or based on (2) procedural rationality, i.e. the extent to which a judgment (e.g. the use of information) corresponds with some “right process” (Bonner 2008, Simon 1978b, 1976). These two perspectives can be described as:

(1) Performance view of judgment quality with result accuracy as a criterion, which compares a judgment (e.g. subjective ranking of alternatives) with an optimal judgment (e.g. optimal ranking of alternatives), i.e. a judgment is accurate when it is appropriate to the attainment of given goals within the limits imposed by given conditions and constraints (Bonner 2008, Simon 1978b, 1976).

(2) Process view of judgment quality, with the criteria of (a) procedural consistency (e.g. consistent preferences for attributes of an alternative), i.e. a judgment is consistent when it is the outcome of careful consideration; and (b) procedural time, i.e. a fast judgment process (Bonner 2008, Simon 1978b, 1976).

2.2 Cognitive Fit Theory

Cognitive fit theory is one aspect of a general theory of problem solving (Vessey 2006). It develops the notion of “cognitive fit”: a match of problem representation and problem-solving task, both elements of the task environment, enables the problem solver to use cognitive processes that emphasize the same type of information. Thus, the processes used to act on the problem representation and the task will match. The resultant consistent mental representation will facilitate the judgment process and lead to an accurate and fast problem solution. When problem representation and task do not match, similar processes cannot be used to act both on the problem representation and to solve the task. The mental representation will have to be transformed. The decision strategy will require more effort and will most likely be less accurate and more time-

consuming. Therefore, there is no reason to deviate from cognitive fit (Vessey 2006, 1994, 1991).

Cognitive fit theory can be applied to problem representations in the form of graphical and tabular presentation formats. Main assumptions are that representation format and task characteristics are independent, as data can be represented independently from the problem and solutions can be achieved with different presentation formats. Also, graphs and tables must be derived from equivalent raw data so that all information in one presentation format is inferable from the other (Vessey 1994, 1991).

Graphs are spatial problem representations and present spatially-related information. They emphasize relationships in the data and do not present discrete data values directly. They facilitate viewing the information at a glance without addressing the elements separately or analytically. The data in a graph are accessed using perceptual processes (Vessey 1994, 1991). Perceptual processes include visually perceiving an object and deriving its meaning (Bolles 1991). Tables are symbolic problem representations and present symbolic information. They emphasize discrete data values and facilitate extracting specific data values. They do not show relationships directly and the data in a table are accessed using analytical processes (Vessey 1994, 1991). Analytical processes involve cognitive effort to understand or evaluate a given phenomenon (Simon and Lea 1974). Perceptual processes are employed more effectively on spatial presentation formats, whereas analytical processes are employed more effectively on symbolic presentation formats (Vessey 1994, 1991).

In line with Einhorn and Hogarth (1981) cognitive fit theory differentiates between information acquisition tasks and information evaluation tasks. Information acquisition tasks (i.e. search of the task environment and short-term memory to access information) are solved directly via information acquisition processes, which can either be (1) spatial, i.e. relationships among the data are perceived perceptually, or (2) symbolic, the acquisition of discrete and precise data values is processed analytically (Speier 2006). Information evaluation tasks (i.e. the strategy a problem solver implements to understand or manipulate information) can also either be (1) spatial, i.e. associations among the data are perceived perceptually, or (2) symbolic, i.e. the extraction of given data values followed by a series of calculations is processed analytically. Cognitive fit exists when spatial information acquisition and evaluation tasks are supported with spatial problem representations and when symbolic information acquisition and evaluation tasks are

supported with symbolic problem representations (Vessey 2006, 1994, 1991). There is widespread empirical support for cognitive fit with spatial or symbolic presentation formats for information acquisition tasks and simple information evaluation tasks (for an overview, see Kelton et al. 2010, Speier 2006, Vessey 2006).

2.3 Cognitive Fit Theory in the Context of Complex Multiattribute Judgment Tasks

The theory of cognitive fit is also supported for complex multiattribute judgment tasks, e.g. in accounting (Vessey 2006).¹ Task complexity refers to the amount of information processing a task requires and the amount of structure the task provides. Hence, task complexity increases as the required amount of information processing increases and as the degree of task structure decreases (Campbell 1988, Wood 1986). Multiattribute judgment tasks are considered to be inherently complex, due to the high number of criteria and alternatives underlying multiattribute judgment tasks and the substantial evaluation necessary for problem solving. The information load of complex multiattribute judgments tasks exceeds the cognitive capabilities of a boundedly rational decision maker to employ cognitively demanding analytical processes and he is not capable of solving the task analytically (Umanath and Vessey 1994).

As perceptual processes require less effort than analytical processes, the data in complex multiattribute judgment tasks are accessed using perceptual and holistic processes. Thus, according to the theory of cognitive fit, complex multiattribute judgment tasks are called “holistic” (Vessey 2006, Umanath and Vessey 1994). Holistic processes aid in integrating large amounts of data (e.g. integration of a number of performance measures over a number of years for different business units) (Umanath and Vessey 1994).

In performance evaluation, when information for multiple performance measures is first integrated to produce a “result” over one object of a category (e.g. time period, organizational unit or employee) and then the results for the multiple objects of a category

¹ Types of multiattribute judgment tasks analyzed in accounting are bankruptcy prediction (So and Smith 2004, Umanath and Vessey 1994, Eggleton et al. 1992, MacKay and Villareal 1987, Moriarity 1979), prediction of changes in bond rating (Frownfelter-Lohrke 1998, Wright 1995, Nibbelin et al. 1992, Stock and Watson 1984), and performance evaluation (Banker et al. 2011, Tuttle and Kershaw 1998). These multi-criteria tasks include the presentation of several performance measures (profitability, liquidity, leverage, etc.) over one or multiple categories (time periods, firms, subunits of a firm, employees, etc.) and require a single categorical assessment of a category (e.g. bankruptcy for a firm yes/no; rating change over time yes/no) or a comparison across categories (e.g. performance evaluation across subunits for a given time period).

(e.g. time periods, organizational units or employees) are integrated to produce the required response, a holistic process is used. When data for each performance measure is first integrated to produce a “result” over multiple objects of a category (e.g. time periods) and then the results for each performance measure are integrated to produce the required response, a perceptual process is used. Both holistic and perceptual processes involve making associations among data values (Umanath and Vessey 1994).

A performance evaluation of organizational subunits over multiple time periods based on multiple performance measures is a complex multiattribute judgment task. To evaluate and compare the financial performance of organizational subunits over multiple time periods, a decision maker may first integrate performance measures for one subunit during one time period (holistic process). Then he will repeat this process for all time periods (perceptual process) to generate a response for the subunit and in the next step compare the subunits to generate a performance evaluation judgment. He may as well compare a single performance measure for a subunit across multiple time periods (perceptual process) and then to other subunits to generate a response for the single indicator across subunits. In the next step he will integrate the result for each indicator (holistic process) to produce the performance evaluation judgment across the subunits.²

Both complex spatial tasks and multiattribute judgment tasks require the integration of large amounts of data so that results for complex spatial judgment tasks can also be applied to multiattribute judgment tasks (Umanath & Vessey 1994). Empirical evidence shows that complex spatial judgment tasks are supported by a graphical presentation format when there is no predominant accuracy criterion (Vessey 2006). Speier (2006) supports cognitive fit for a complex multiattribute spatial judgment task with graphs leading to more accurate and faster judgments than tables. Speier et al. (2003) predict and find for both low complex and high complex multiattribute tasks that spatial tasks are best facilitated using spatial formats with respect to accuracy and time. Wilson and

² As Umanath and Vessey (1994) note, these processes are idealized processes that take into account only the relationships among a number of performance measures for one category or the relationships between values of the same performance measure over a number of objects of a category. Also the values or levels of performance measures can be relevant for judgment. To facilitate this type of evaluation, presentation formats must also make available the underlying data. Graphs are a presentation format that supports both relationships among the data and the underlying data values. With respect to referencing underlying values or levels of data graphs both preserve characteristics of the underlying data and provide integrating features. However, to obtain data values in graphs, it takes a large amount of time and errors can be made. Tables emphasize symbolic information, as they present discrete data values. Therefore, tables aid analytic processes but do not support the decision maker in integrating data. Integrating data with tables involves considerable cognitive resources, will probably take up a huge amount of time and be erroneous (Umanath and Vessey 1994).

Addo (1994) analyze complex spatial and symbolic uniattribute judgment tasks and find that both are generally solved faster with graphs compared to tables.

From the performance view of judgment quality, result accuracy with graphs should be higher than with tables, as graphs both aid in integrating large amounts of data and allow referring to underlying data values, whereas tables only show discrete data values and emphasize symbolic information. Umanath and Vessey (1994) hypothesize and find that graphs are more accurate than tables for a complex multiattribute judgment task. Tuttle and Kershaw (1998) employ a complex multiattribute performance evaluation judgment task of plant managers and manipulate judgment strategy. They also find that for a holistic strategy graphs are more accurate than tables.

From the process view of judgment quality, procedural consistency should also be higher with graphs than with tables, as the data in graphs are accessed using perceptual processes and problem solving in multiattribute judgment tasks is facilitated by perceptual processes. This match of problem representation and task reduces task complexity, the short-term memory constraint is extended, and the judgment process is facilitated by a consistent mental representation.

Finally judgment time may not differ between graphs and tables. On one hand graphs facilitate perceptual processes and time consuming mental transformations will be necessary for tabular presentation formats, on the other hand the reference to underlying data values with graphs also takes a huge amount of time and can be erroneous, whereas tables show data values directly. Empirical evidence supports this, as Tuttle and Kershaw (1998) and Umanath and Vessey (1994) find judgment time not to differ between graphs and tables. Therefore, the following hypotheses are stated:

***HI(a):** For a complex multiattribute judgment task graphs result in higher accuracy as tables.*

***HI(b):** For a complex multiattribute judgment task graphs result in higher consistency as tables.*

***HI(c):** For a complex multiattribute judgment task there is no difference in speed between graphs and tables.*

Numerical data at the end (bottom) of horizontal (vertical) bar charts results in a combined graphical and tabular representation and conveys both spatial and symbolic information. An extension of cognitive fit suggests that such combined representations

will outperform tabular formats on spatial tasks and have no performance effects symbolic tasks (Vessey 1991). DeSanctis and Jarvenpaa (1989) employ horizontal bar charts (with numerical values at the end of each bar in the combined format) for a multiattribute spatial forecasting task. After subjects gain practice with the task, both graphical formats are more accurate than a table. Moreover, the combined format marginally improves accuracy after the first and over the following four trials when compared to the graphical format. Also given a reasonable amount of time to solve the task, the combined and graphical group subjects complete the task faster than tabular group subjects. Wright (1995) analyzes a multiattribute judgment task with high task complexity and finds for a combined presentation format with tables and additional graphical representations (bar and line charts) that accuracy is higher compared to a tabular format. However, studies exist where a pure format has higher accuracy than a combined format. Both Nibbelin et al. (1992) and Frownfelter-Lohrke (1998) examine complex multiattribute bond-rating-change decisions. Nibbelin et al. (1992) examine tables and tables with bar charts for a spatial task. Tables lead to higher accuracy than the combined format. Frownfelter-Lohrke (1998) examines tables, bar charts and a combination of these. Neither for a spatial nor for a symbolic task is accuracy significantly different between the presentation formats. For both task types the graphical presentation format required less time than the combined format and both formats lead to significantly faster judgments than the tables. So and Smith (2004) analyze a spatial multiattribute bankruptcy prediction task. When information complexity is low, presentation format has no impact on accuracy. However, when information complexity is high, the tabular-alone format shows the highest accuracy compared to a combined tabular-graphical format. Overall, results from prior research provide mixed evidence for an extension of cognitive fit theory to combined presentation formats.

Multiattribute performance evaluation judgments afford integrating large amounts of data and the ability to reference ranges or levels of performance measures. Graphs with data values support both relationships among the data and directly provide the underlying data values. If data values are presented at the end of bar charts or at data points of line charts they convey symbolic information as they present discrete data values. Symbolic information facilitates analytic processes but complex multiattribute judgment tasks demand holistic and perceptual processes. Therefore, the symbolic information may merely be discarded or even hinders a consistent mental representation. Thus, at best both accuracy and consistency are not expected to change when comparing pure

graphical representations with graphs with data values. However graphs with data values are still expected to lead to more accurate and consistent results than tables. As graphs with data values directly reference numerical values the amount of time to obtain a value is shorter and less prone to error when compared to graphs, though the combined spatial and symbolic presentation may lead to an increased judgment time to form a consistent mental representation. It is therefore expected that graphs with data values do neither differ to graphs nor tables with respect to procedural speed. The following hypotheses are stated for graphs versus graphs with data values:

***H2(a):** For a complex multiattribute judgment task there is no difference in accuracy between graphs and graphs with data values.*

***H2(b):** For a complex multiattribute judgment task there is no difference in consistency between graphs and graphs with data values.*

***H2(c):** For a complex multiattribute judgment task there is no difference in speed between graphs and graphs with data values.*

The following hypotheses are stated for graphs with data values versus tables:

***H3(a):** For a complex multiattribute judgment task graphs with data values result in higher accuracy than tables.*

***H3(b):** For a complex multiattribute judgment task graphs with data values result in higher consistency than tables.*

***H3(c):** For a complex multiattribute judgment task there is no difference in speed between graphs with data values and tables.*

2.4 Information Aggregation

Information-processing theory (Driver and Streufert 1969, Schroder et al. 1967) can be interpreted as an aspect of a general theory of problem solving (Simon 1978a). According to information-processing theory, individuals can be viewed as information-processing systems which respond in a curvilinear fashion to information load: the amount of information integrated into the problem-solving process reaches a maximum, at which a problem solver is expected to achieve maximum complexity in information processing, and then declines as task complexity increases. The problem solver is considered to have sensed information overload at the point where the amount of information integrated into the mental representation begins to decrease. Beyond this point,

the individual's problem solutions reflect a lesser implementation of the information available, i.e. judgment quality decreases (Chewning and Harrell 1990). The relationship between amount of information provided and amount of information integrated by problem-solver information processing is known as an inverted u-curve (Eppler and Mengis 2004).

Aggregation of information is known as a countermeasure against information overload (Eppler and Mengis 2004). Information aggregation may be of two different types: Aggregation of type I reduces the amount of data provided to a problem solver, but does not change the information content of that data with respect to a particular judgment. Aggregation of type II reduces both the amount of data provided to a problem solver and the information content of that data. When a problem solver moves from highly disaggregate information to more aggregate information, judgment quality will improve, provided that information aggregation is of type I, as a constant information content and less information cues reduce task complexity and facilitate information processing. Once information aggregation becomes type II, performance declines due to the lack of relevant information necessary for problem solution. Hence, provided both types of aggregation are included, the judgment and decision-making performance of a problem solver is also expected to behave as an inverted u-shaped function of information aggregation (Otley and Dias 1982).

Prior studies in accounting by Ronen (1971), Chervany and Dickson (1974) and Tiessen (1976) examine information aggregation of type I. Ronen (1971) shows for a production planning task that information aggregation can improve accuracy. Chervany and Dickson (1974) conclude that aggregated information leads to higher accuracy compared to raw data. However, procedural speed is slower with aggregated data. Tiessen (1976) also shows positive effects of an increasing degree of aggregation. Harvey et al. (1979) examine information aggregation of type II and systematically change information content. Investment decision can be interpreted better by financial analysts with disaggregated information as with aggregated information of type II. Otley and Dias (1982) examine both types of information aggregation and find evidence for the inverted u-shaped function of information aggregation for result accuracy. With respect to procedural speed, information aggregation of type I leads to faster decisions. Kaufmann and Weber (2013) analyze aggregation of type I in the context of multi-criteria investment decisions. Results show that information aggregation results in greater risk-taking and increased risk-taking is associated with a lower risk perception and a more

precise estimation of the probability of a loss. Procedural speed however does not change with information aggregation.

To facilitate the judgment process, and ultimately judgment quality, information aggregation of type I is to be employed. Information aggregation of type I reduces the amount of information, whereas the information content stays constant. As the number of information cues is reduced, task complexity decreases and the demands on information processing are relaxed. This in turn extends the short-term memory constraint and leads to an increased accuracy, a more consistent mental representation, and a reduction in judgment time. The following hypotheses can be stated:

H4(a): *For a complex multiattribute judgment task, accuracy is higher with high information aggregation than with low information aggregation.*

H4(b): *For a complex multiattribute judgment task, consistency is higher with high information aggregation than with low information aggregation.*

H4(c): *For a complex multiattribute judgment task, speed is higher with high information aggregation than with low information aggregation.*

Umanath and Vessey (1994) examine the effect of presentation format on judgment quality with a change in information load. They analyze additional information, which has no impact on the judgment outcome. Therefore, they differentiate between (1) nominal information load and (2) increased level of information load. Nominal information load is defined as the least amount of information that makes an objective difference in prediction accuracy. Increased information load is defined as nominal information load plus a number of variables which convey no additional information (i.e. do not change the information content). They find for a multiattribute judgment task that for graphs there is no difference in accuracy with a change in information load. They also find no difference in accuracy with change in information load when tables are used. With respect to time it is shown that judgments with graphs are marginally faster and judgments with tables are significantly faster when information load decreases.

Graphs support both relationships among the data and provide the underlying data values. Given information aggregation of type I, when information is disaggregated without adding information content, information load increases, therefore task complexity rises and additional processing is required to handle more information. This stresses the short-term memory constraint so that the judgment process takes more time, the mental representation is likely to be less consistent and result accuracy will be prone to

more error. However, graphs possess integrating capabilities that mitigate the adverse effects on accuracy and consistency (Umanath and Vessey 1994). Therefore, the following hypotheses are stated:

H5(a): *With graphs, there is no difference in accuracy task with change in information aggregation for a complex multiattribute judgment task.*

H5(b): *With graphs, there is no difference in consistency with change in information aggregation for a complex multiattribute judgment task.*

H5(c): *With graphs, speed is lower at low information aggregation compared to high information aggregation for a complex multiattribute judgment task.*

Graphs with data values possess integrating capabilities and directly provide the underlying data values. Thus, by analogy to graphs, the following hypotheses are stated:

H6(a): *With graphs with data values, there is no difference in accuracy with change in information aggregation for a complex multiattribute judgment task.*

H6(b): *With graphs with data values, there is no difference in consistency with change in information aggregation for a complex multiattribute judgment task.*

H6(c): *With graphs with data values, speed is lower at low information aggregation compared to high information aggregation for a complex multiattribute judgment task.*

With tables, the chance of error will be higher at increased information load as tables possess no integrating capabilities (Umanath and Vessey 1994). By analogy, with disaggregated information, task complexity increases. Hence, the additional information processing will take more time and will be more susceptible to error, leading to less consistency and accuracy. The following hypotheses can be stated:

H7(a): *With tables, accuracy is lower at low information aggregation compared to high information aggregation for a complex multiattribute judgment task.*

H7(b): *With tables, consistency is lower at low information aggregation compared to high information aggregation for a complex multiattribute judgment task.*

H7(c): *With tables, speed is lower at low information aggregation compared to high information aggregation for a complex multiattribute judgment task.*

3 Method and Design

3.1 Experimental Design

This experimental study utilizes a 3×2 between-subjects design to analyze the effects of the independent variables *presentation format* and *information aggregation* on *judgment quality*. Three types of presentation format are examined: *tables*, *graphs* and *graphs with data values*. Information aggregation is manipulated with either *high information aggregation* or *low information aggregation*. To test the hypotheses based on cognitive fit theory and information-processing theory, a between-subjects design allows for a precise assessment of subjects' natural reasoning processes. Also, it minimizes the chance to correct inconsistencies and errors in participant's judgments which provides for an accurate test of judgment quality (Libby et al. 2002, Kahneman and Tversky 1996). Judgment quality is evaluated from a performance perspective with result *accuracy* as a dependent variable and from a process perspective with procedural *consistency* and procedural *time* as dependent variables.

3.2 Participants

Two hundred and six undergraduate students (151 female, 55 male) enrolled in a management accounting course at a public university in Germany participated in the study and each subject was randomly assigned to one of the six between-subjects treatments. Participation in the experiment was voluntary. On average, participants were 22.0 years old and took 5.2 courses in accounting and finance. All students were undergraduates and 93.7 % were in the fourth or a higher semester.³ In line with prior studies (e.g. Cardinaels 2008, Speier 2006, So and Smith 2004, Tuttle and Kershaw 1998) the use of undergraduate students is justified for several reasons: (1) the information-processing behavior of students and managerial decision makers does not differ in a review of prior studies, as reported by Ashton and Kramer (1980); (2) there are no reported differences between subjects without and with work experience for the kind of tasks examined in this study (MacKay and Villarreal 1987, Stock and Watson 1984); (3) the performance evaluation task in this study was drawn from the management accounting domain so that prior exposure to the task could be embedded into the subjects' course work increasing the domain knowledge of the subject pool (Speier 2006).

³ The gender distribution represents the average subject pool of the respective university. There are no significant differences across treatments with respect to gender ($p = .699$), age ($p = .624$), year in school ($p = .812$), and prior accounting and finance knowledge ($p = .116$).

3.3 Experimental Task and Independent Variables

The complex multiattribute judgment task employed in this experiment is to evaluate the performance of three business units of a company based on accounting information. Each subject receives a decision proposal depicting five performance measures over a time period of four years for all three business units. In line with Speier (2006) the decision proposal fits on one page to better control the cognitive processes needed for information acquisition and evaluation. Also, as in So and Smith (2004) all five performance measures are accounting ratios that represent key dimensions of accounting information: (1) liquidity, measured by cash ratio as liquid assets over current liabilities, (2) profitability, measured by operating profit margin as earnings before interest and taxes (EBIT) over sales, (3) profitability, measured by return on invested capital (ROIC) as earnings before interest and taxes over invested capital, (4) leverage, measured by debt to equity ratio as debt (book value) over equity (book value), and (5) risk, measured by weighted average cost of capital (WACC) as weighted cost of debt (market value) plus weighted cost of equity (market value). The numerical values of the performance measures are specified to provide for a quasi-dominant alternative an objectively correct ranking of the business units.

The subject's task is (1) to assess the business development of the three business units based on the five performance measures for the past four years and (2) to determine a ranking of the business units, starting with the business unit, which was financially most successful over time. With 60 information cues, this multiattribute judgment task can be considered as sufficiently complex for a problem solver not being able to solve it with analytical processes. Thus, the task consists of two sub-tasks: (1) holistic and perceptual information acquisition, as the attributes of the various alternatives over time need to be considered and compared; (2) holistic and perceptual information evaluation, as the information needs to be evaluated and the business units have to be ranked based on the development of the performance measures over time.

Introductory statements, task instructions and definitions of the performance measures are identical for all treatments. Also, subjects were not given instructions how to weigh the performance measures. However the decision proposal in each treatment differed with respect to *information aggregation* and *presentation format*.

Information Aggregation

In order to be consistent with the hypotheses, *information aggregation* needs to be of type I, i.e. the amount of information is reduced, but not the content of that information with respect to a specific decision. Hence, the same relevant information should be available for all subjects and based on this information they should all be given the possibility to rank the business units in the correct order.

Therefore, the number of alternatives (3 business units), number of attributes (5 performance measures), and the time period (4 years) were held constant. The *high information aggregation* and *low information aggregation* treatments differed only with respect to the degree of aggregation of the performance measures. The experimental groups with *high information aggregation* only received the ratios of the performance measures (e.g. cash ratio). The groups with *low information aggregation* additionally received information on the two components of the performance measures (e.g. cash ratio as liquid assets over current liabilities). Hence, the groups with *high information aggregation* had to process a total of 60 information units whereas the groups with *low information aggregation* had to process 180 information units. Whereas the difference in the amount of information in the two treatments was chosen to be sufficiently large, the value of information did not differ between the treatments, because the two components of the performance measures in the *low information aggregation* treatments are neither necessary nor sufficient data for the evaluation of the business units.

Presentation Format

The accounting information in the decision proposal is presented in three different *presentation formats* (*graphs, graphs with data values, tables*). Care was taken to balance the information content of *tables* (Appendix 1, 2), of *graphs* (Appendix 3, 4), and of *graphs with data values* (Appendix 5, 6) with respect to the *information aggregation* manipulation (Tuttle and Kershaw 1998). Specifically, the construction of the three *presentation formats* in this experiment is as follows:

(1) For the presentation format with *tables* each business unit is represented by a separate table. The three tables for the respective business units are ordered vertically on one page. For the *tables with high information aggregation* treatment (Appendix 1) for each business unit the numerical values of the five ratios are presented for four years. For the *tables with low information aggregation* treatment (Appendix 2) addi-

tionally the two components of each ratio are added as line items above the respective accounting ratio. The components and the ratio are separated by a horizontal line and the calculation of the ratio from the components is shown in the form of arithmetic symbols. For both the *high information aggregation* and the *low information aggregation* treatments the order from top to bottom is: cash ratio, operating profit margin, return on invested capital, debt to equity ratio, weighted average cost of capital, whereas the order is not varied.

(2) For the presentation format with *graphs* each business unit is represented by five distinct and vertically ordered graphs with one ratio each. The three business units are ordered horizontally on one page. For the *graphs with high information aggregation* treatment (Appendix 3) for each business unit the five ratios are presented as line charts over four years. For the *graphs with low information aggregation* treatment (Appendix 4) for each year the two components of each ratio are added as bar charts with different shades of grey in the respective line chart. Again, for both the *high information aggregation* and the *low information aggregation* treatments the order from top to bottom is: cash ratio, operating profit margin, return on invested capital, debt to equity ratio, weighted average cost of capital, and is not varied.

(3) For the presentation format *graphs with data values*, the presentation is identical to the respective *graphs* treatment for both *high information aggregation* and *low information aggregation*, except for the addition of numerical values for each ratio and each year. The additional data values are depicted directly in the charts above the lines for the ratios in the *high information aggregation* treatment (Appendix 5) and directly above the lines for the ratios and the bars for the components in the *low information aggregation* treatment (Appendix 6).

3.4 Dependent Variables

The goal of this study is to analyze *judgment quality* for a complex multiattribute judgment task when both *information aggregation* and *presentation format* are varied. *Judgment quality* is operationalized through result *accuracy*, procedural *consistency*, and procedural *time*.

Accuracy

Accuracy as a dependent variable compares an objectively correct judgment with a subject's actual judgment. *Accuracy* decreases when the actual judgment deviates from

the correct judgment. This approach can be pursued within the framework of this analysis, as the experimental task was formulated in a way that there was an optimal solution in terms of an objective ranking of alternatives. To evaluate performance, subjects had to “enter a ranking of the business units, starting with the business unit, which was financially most successful over time”. This wording indicates that the profitability and risk indicators are of great relevance and the other key figures largely can be ignored in this context. Furthermore, in each year the value spread for each business unit can be calculated as the difference between the return on invested capital and the weighted average cost of capital. Given the numerical values of the task, on average business unit B has the highest average value spread over time and business unit C has the lowest, resulting in the optimal ranking: B > A > C. To increase domain knowledge, all students enrolled in the management accounting course, in which the experiment was held, were taught the relevant content to solve this problem.

To determine *accuracy*, for each business unit the deviation between the actual rank and the objectively correct rank is determined and the sum of absolute values is calculated. The dependent variable *accuracy* is then determined as follows:

$$Accuracy = \left| Rank_{optimal}^A - Rank_{actual}^A \right| + \left| Rank_{optimal}^B - Rank_{actual}^B \right| + \left| Rank_{optimal}^C - Rank_{actual}^C \right| \quad (1)$$

The highest *accuracy* is achieved, when there is no deviation between the optimal and actual judgment so that the value for accuracy equals 0. A higher value indicates a less accurate judgment. Possible values are 0, 2, and 4. For example, if a subject ranks the business units: A = 3, B = 2, C = 1 while the optimal ranking is A = 2, B = 1, C = 3, then *accuracy* equals $|2-3| + |1-2| + |3-1| = 4$.

Consistency

Consistency as a dependent variable measures the consistency of the specification of the weights of the accounting ratios in the performance evaluation task. *Consistency* is calculated utilizing the analytic hierarchy process (Saaty 1994, 1990, 1980). Based on pairwise comparisons between the ratios, individual ratio weights can be determined for each subject. Pairwise comparisons are utilized to reconstruct participants’ preference relations, which are represented in comparison matrices. A comparison matrix can either be consistent or not. To measure the degree of procedural consistency, an examination of consistency can be performed. For this purpose, an average matrix is generated

as a basis of comparison, whose average eigenvalue λ_{\max} is included in the determination of a consistency index CI (Saaty 1994):

$$CI = \frac{\lambda_{\max} - n}{n - 1}, \quad (2)$$

where n denotes the number of rows respectively columns of the average matrix. Finally the consistency ratio CR is calculated by dividing the consistency index CI by a random average index R , $R = 1.12$ (Saaty 1994):

$$CR = \frac{CI}{R} \quad (3)$$

The consistency ratio measures *Consistency* as a dependent variable. The lower the value of the consistency ratio the more consistent are the pairwise comparisons of a subject and therefore the specified weights of the accounting ratios (Saaty 1994).

Time

Time as a dependent variable is measured as a period of time in seconds. The subjects take notes of the current time in minutes and seconds (1) after reading task instructions and (2) after entering their respective ranking of the business units. The difference between the two values indicates the time of the judgment process.

3.5 Experimental Procedures

The experiment was held during regular class sessions. After the instructions were read out aloud, the experimental materials, including introductory statements, task instructions, decision proposals, and questionnaires were distributed randomly among participants. Then the participants performed the judgment task at own speed and answered the questions on the post-test questionnaire.

In the experiment participants were asked twice to record the current time in minutes and seconds (time was projected in digital letters on the wall of the class room using a video projector): (1) after reading the introductory statements and task instructions, and (2) after the judgment task. Also after the judgment, pairwise comparisons between the performance measures had to be performed to calculate the consistency ratio based on the analytic hierarchy process (Saaty 1994, 1990, 1980). With the aim to limit the total number of pairwise comparisons, reciprocal values were used for reverse

comparisons (Saaty 1994). After all participants completed both the judgment task and the post-test questionnaire, the experimental materials were collected.

4 Results

4.1 Manipulation Checks

To measure a general understanding of the task, participants are asked in the post-experiment questionnaire to rate their own understanding for the multiattribute judgment task. The mean of 204 participants' responses to the item that measures general understanding is 4.16 (standard deviation = .96). This applies for the statement "I have understood the instructions and the task (creation of a ranking based on ratios)", on a five-point Likert scale with 1 = strongly disagree and 5 = strongly agree. Therefore the task can be said to be understood on average.

Also, as in Umanath and Vessey (1994) it is tested whether a participant's result *accuracy* differs significantly from randomly generated responses. Participant's performance evaluations for each combination of *presentation format* and level of *information aggregation* were on average significantly higher than the random *accuracy* of 2.66⁴ (all $t > 4.0$; all $p < .001$; two-tailed). This indicates that individuals put effort into the task and merely random responses can be ruled out on average.

Furthermore, it was analyzed whether subjects used more *holistic* or *perceptual processing* compared to *analytic processing* (Panel A in Table 4), which is an indicator that the task is processed holistically compared to being perceived analytically. Two items measure cognitive processing, both statements on a five-point Likert scale with 1 = strongly disagree and 5 = strongly agree. The mean of participants' responses for *holistic/perceptual processing* is 3.91 (standard deviation = .89) for the statement "In assessing the business units I have tried to capture business development as a whole in order to gain a general impression of the development over time". The mean of participants' responses for *analytic processing* is 2.21 (standard deviation = 1.03) for the statement "In assessing the business units I have analyzed the discrete numerical values of the indicators in order to calculate exact results for the business development". Answers to both items differ significantly (Wilcoxon signed-rank test; $z = 10.84$; $p < .001$; two-tailed). This result is consistent with the task being perceived *holistically* as the participants employ significantly more *holistic* or *perceptual processing*.

⁴ With three business units to evaluate, there are six possible rankings. With absolute deviations possible values are: 0, 2, 2, 4, 4, 4. Hence, the average random response is $(0 + 2 + 2 + 4 + 4 + 4) / 6 = 2.66$.

Table 1: Descriptive Statistics for Accuracy, Consistency and Time

Descriptive Statistics for Accuracy ^a , Consistency ^b , and Time ^c Mean; [Standard Deviation]; (Number of Observations)				
Presentation Format ^d	Dependent Variable	Information Aggregation ^e		Means
		High Aggregation (Ratio)	Low Aggregation (Ratio, Numerator, Denominator)	
Tables	Accuracy	1.45 [1.15] (n = 33)	1.71 [1.38] (n = 35)	1.59 [1.27] (n = 68)
	Consistency	0.223 [0.280] (n = 33)	0.208 [0.193] (n = 35)	0.215 [0.237] (n = 68)
	Time	219.94 [93.05] (n = 33)	232.17 [97.43] (n = 35)	226.24 [94.82] (n = 68)
Graphs	Accuracy	0.79 [1.11] (n = 33)	1.45 [1.52] (n = 33)	1.12 [1.36] (n = 66)
	Consistency	0.161 [0.127] (n = 33)	0.248 [0.212] (n = 33)	0.205 [0.179] (n = 66)
	Time	204.64 [91.18] (n = 33)	219.65 [92.35] (n = 34)	212.25 [91.39] (n = 67)
Graphs with Data Values	Accuracy	0.78 [1.29] (n = 36)	1.47 [1.58] (n = 34)	1.11 [1.47] (n = 70)
	Consistency	0.226 [0.228] (n = 36)	0.274 [0.260] (n = 34)	0.249 [0.243] (n = 70)
	Time	213.56 [106.81] (n = 36)	255.74 [120.09] (n = 34)	234.04 [114.61] (n = 70)
Means	Accuracy	1.00 [1.22] (n = 102)	1.55 [1.49] (n = 102)	1.27 [1.38] (n = 204)
	Consistency	0.204 [0.221] (n = 102)	0.243 [0.222] (n = 102)	0.223 [0.222] (n = 204)
	Time	212.74 [96.79] (n = 102)	235.82 [103.99] (n = 103)	224.33 [100.89] (n = 205)

^a *Accuracy* as a dependent variable compares an objectively correct judgment with a subject's actual decision. *Accuracy* decreases when the actual decision deviates from the correct judgment. To determine *accuracy*, for each alternative the deviation between the actual rank and the objectively correct rank is determined and the absolute sum is calculated. The highest *accuracy* is achieved, when there is no deviation between the optimal and actual judgment so that the value for *accuracy* equals 0. A higher value indicates a less accurate judgment. Possible values are 0, 2, and 4.

^b *Consistency* as a dependent variable measures the procedural consistency of the specification of the weights of the accounting ratios in the performance evaluation task. *Consistency* is calculated utilizing the analytic hierarchy process (Saaty 1980, 1990, 1994). Based on pairwise comparisons between the ratios, a consistency ratio CR is calculated. The lower the value of the consistency ratio the more consistent are the pairwise comparisons of the subjects and therefore the specified weights of the accounting ratios.

^c *Time* as a dependent variable is measured as a period of time in seconds. The subjects write down the current time in minutes and seconds (1) after reading the introductory statements and after task instructions and (2) after entering their respective actual ranking of the business units. The difference between the two values indicates the *time* of the judgment process.

^d *Presentation Format* as an independent variable was manipulated between-subjects with the three factor levels *tables*, *graphs*, *graphs with data values*. For the presentation format with *tables* each business unit is represented by a separate table. For the *tables with high information aggregation* treatment for each business unit the numerical values of the five ratios are presented for four years. For the *tables with low information aggregation* treatment additionally the two components of each ratio are added as line item above the respective accounting ratio. For the presentation format with *graphs* each business unit is represented by five separate graphs with one ratio each. For the *graphs with high information aggregation* treatment for each business unit the five ratios are presented as line charts over four years. For the *graphs with low information aggregation* treatment additionally for each year the two components of each ratio are added as bar charts with different shades of gray in the respective line chart. For the presentation format with *graphs with data values*, the presentation is identical to the respective graphs treatment for both *high information aggregation* and *low information aggregation*, except for the addition of numerical values for each ratio and each year. The additional data values are depicted directly in the charts above the lines for the ratios in the *high information aggregation* treatment and directly above the lines for the ratios and the bars for the components in the *low information aggregation* treatment.

^e *Information aggregation* as an independent variable was manipulated between-subjects with the two factor levels *high information aggregation* and *low information aggregation*. The number of alternatives, number of attributes, and the time period were chosen as a constant. The treatments differed only with respect to the degree of aggregation of the key indicators. The experimental groups with *high information aggregation* were given only the ratios of the key indicators. The groups with *low information aggregation* additionally received information on the two components of the key indicators. The groups with *high information aggregation* had to process a total of 60 information units whereas the groups with *low information aggregation* had to process 180 information units.

4.2 Hypothesis Tests

As in Umanath and Vessey (1994), first it is tested, whether the average *accuracy* of the performance evaluations differs significantly from randomly generated responses. Subject's evaluations for each *presentation format* and each level of *information aggregation* are on average significantly higher than random evaluation scores (see Section 4.1).

The second step is to perform multivariate (MANOVA) analyses when dependent variables are correlated. Pearson product-moment correlation coefficient indicates a small relationship between $\ln(\textit{consistency})^5$ and *time* ($r = -.125$; $p = .075$). Results indicate that *information aggregation* simultaneously affects $\ln(\textit{consistency})$ and *time* (MANOVA; not tabulated; Wilk's Lambda = .969; $F = 3.194$; $p = .043$) whereas *presentation format* does not influence $\ln(\textit{consistency})$ and *time* simultaneously (MANOVA; not tabulated; Wilk's Lambda = .987; $F = .623$; $p = .623$).

The next steps involve the follow-up analyses. The SPSS generalized linear model procedure (GENLIN) is used to analyze the effects of the independent between-subjects variables: *presentation format* and *information aggregation* on *accuracy*, *consistency* and *time*. Specifically, six unique cell means can be created – *high aggregation* and *low aggregation*, and *presentation format* with *graphs*, *graphs with data values*, and *tables*.

Because generalized linear models are overall tests of the effects of the independent variables and because the *presentation format* contains three levels (*graphs*, *graphs with data values*, *tables*), additional follow up tests are performed to better understand the effects of *information aggregation* and *presentation format*. The descriptive statistics are depicted in Table 1. The generalized linear models as overall tests of the effects are shown in Table 2 and the detailed mean by mean comparisons are reported in Table 3.

Accuracy

A generalized linear model analysis using *accuracy* as the dependent variable is performed (Table 2). *Accuracy* has a multinomial distribution; therefore, a cumulative logit link-function is used. *Information aggregation* and *presentation format* are specified as independent variables.

⁵ Consistency as a dependent variable is skewed to the right. If variables are right-skewed a measure such as correlation can be influenced by one or a few cases at the high end on one or both variables. Taking the log eliminates skew for consistency. Therefore, consistent with Cardinaels (2008), Waller et al. (1999) and Gupta and King (1997), either a log-linear relation specification of consistency or $\ln(\textit{consistency})$ is employed.

Table 2: Generalized Linear Models (GENLIN)^a with Accuracy, Consistency, and Time as Dependent Variables

Dependent Variable	Accuracy ^b			Consistency ^c			Time ^d		
Distribution	Multinomial			Gamma			Normal		
Link Function	Logit (cumulative)			Log			Identity		
Source of Variation	df	Wald-Chi-Square	p-Value ^e	df	Wald-Chi-Square	p-Value ^e	df	Wald-Chi-Square	p-Value ^e
<i>Factors</i>									
Information Aggregation (A)	1	7.48	.006***	1	4.26	.039**	1	3.80	.051*
Presentation Format (P)	2	7.06	.029**	2	3.29	.193	2	1.90	.387
A x P Interaction	2	1.62	.446	2	3.39	.184	2	1.00	.607
<i>Covariates</i>									
Time	–	–	–	1	7.17	.007***	–	–	–
LN(Consistency)	–	–	–	–	–	–	1	4.45	.035**
	<i>N = 204</i>			<i>N = 204</i>			<i>N = 204</i>		

^a The generalized linear model extends the general linear model in order that the dependent variable is linearly related to the independent factors and covariates via a specified link function. The model allows for the dependent variable to have a non-normal distribution. The dependent variable can be scale, counts, binary, or events-in-trials. Independent factors are assumed to be categorical, covariates are assumed to be scale. The generalized linear model covers widely used statistical models, such as linear regression for normally distributed responses, logistic models for binary data, loglinear models for count data, complementary log-log models for interval-censored survival data, plus many other statistical models through its very general model formulation (IBM Corporation 2012).

^b Accuracy as a dependent variable has a multinomial distribution. This distribution is appropriate for variables that represent ordinal data. The link function is specified as cumulative logit, so that $f(x) = \ln(x / (1-x))$ is applied to the cumulative probability of each category of the dependent variable (IBM Corporation 2012).

^c Consistency as a dependent variable has a gamma distribution. This distribution is appropriate for variables with positive scale values that are skewed toward larger positive values. The link function is specified as log, so that $f(x) = \log(x)$ (IBM Corporation 2012).

^d Time as a dependent variable has a normal distribution. This is appropriate for scale variables whose values take a symmetric, bell-shaped distribution about a central (mean) value. The link function is specified as identity, so that $f(x) = x$. The dependent variable is not transformed. (IBM Corporation 2012).

^e * Significance level of 10% (two-tailed).

** Significance level of 5% (two-tailed).

*** Significance level of 1% (two-tailed).

There is a significant main effect of *information aggregation* on *accuracy* (Wald- $\chi^2 = 7.48$; $p = .006$), a significant main effect of *presentation format* on *accuracy* (Wald- $\chi^2 = 7.06$; $p = .029$), and no significant interaction effect between *information aggregation* and *presentation format* (Wald- $\chi^2 = 1.62$; $p = .446$).

To further examine and better understand the effects of *presentation format* on *accuracy*, mean by mean comparisons are performed with non-parametric tests (Table 3). A mean by mean comparison shows *graphs* (Table 1; mean deviation from optimal solution = 1.12) to give significantly higher *accuracy* than *tables* (1.59) (Mann-Whitney U; $z = -2.24$; $p = .025$). This result supports H1(a). Comparing *graphs* (1.12) to *graphs with data values* (1.11) there is no significant difference in *accuracy* (Mann-Whitney U; $z = -.22$; $p = .827$), which supports H2(a). *Graphs with data values* (1.11) lead to a significantly higher *accuracy* than *tables* (1.59) (Mann-Whitney U; $z = -2.38$; $p = .017$). This result supports H3(a).

Next, the effect of *information aggregation* on *accuracy* is assessed. The mean by mean comparison shows *accuracy* to be significantly higher for *high information aggregation* (1.00) compared to *low information aggregation* (1.55) (Table 3; Mann-Whitney U; $z = -2.65$; $p = .008$). In line with the overall result of the generalized linear model for *information aggregation*, this result provides support for H4(a).

Finally, the effect of *information aggregation* on *accuracy* for a given *presentation format* is examined (Table 3). With *graphs* there is a marginally significant difference in *accuracy*, with *high information aggregation* (.79) leading to more accurate judgments than *low information aggregation* (1.45) (Mann-Whitney U; $z = -1.81$; $p = .070$). This result contradicts H5(a). Also *accuracy* with *graphs with data values* is significantly higher with *high information aggregation* (.78) compared to *low information aggregation* (1.47) (Mann-Whitney U; $z = -1.98$; $p = .048$), which contradicts H6(a). *Tables* do not differ in *accuracy* with *high information aggregation* (1.45) or *low information aggregation* (1.71) (Mann-Whitney U; $z = -.73$; $p = .464$), a contradiction of H7(a).

Consistency

A generalized linear model analysis using *consistency* as the dependent variable is performed (Table 2). *Consistency* has a gamma distribution, thus a log link function is employed. *Information aggregation* and *presentation format* are specified as independent variables, while accounting for the effect of *time* as a covariate.

Table 3: Mean by Mean Comparisons for Accuracy, Consistency, and Time

Accuracy ^a						Consistency ^b						Time ^c					
H _A	Hypothesis ^d	Treat-ment 1	Treat-ment 2	Z-stat	p-Value ^e	H _A	Hypothesis ^d	Treat-ment 1	Treat-ment 2	Wald-Chi-Square	p-Value ^e	H _A	Hypothesis ^d	Treat-ment 1	Treat-ment 2	Wald-Chi-Square	p-Value ^e
Presentation Format						Presentation Format						Presentation Format					
H1(a)	G > T <input checked="" type="checkbox"/>	1.12	1.59	-2.24	.025**	H1(b)	G > T	.194	.216	.62	.431	H1(c)	G = T <input checked="" type="checkbox"/>	212.60	225.17	.55	.460
H2(a)	G = G _D <input checked="" type="checkbox"/>	1.12	1.11	-.22	.827	H2(b)	G = G _D	.194	.248	3.20	.073*	H2(c)	G = G _D <input checked="" type="checkbox"/>	212.60	235.90	1.90	.168
H3(a)	G _D > T <input checked="" type="checkbox"/>	1.11	1.59	-2.38	.017**	H3(b)	G _D > T	.248	.216	1.04	.308	H3(c)	G _D = T <input checked="" type="checkbox"/>	235.90	225.17	.41	.523
Information Aggregation						Information Aggregation						Information Aggregation					
H4(a)	IA _H >IA _L <input checked="" type="checkbox"/>	1.00	1.55	-2.65	.008***	H4(b)	IA _H >IA _L <input checked="" type="checkbox"/>	.194	.245	4.17	.041**	H4(c)	IA _H >IA _L <input checked="" type="checkbox"/>	211.02	238.09	3.80	.051*
Presentation Formats with varying Information Aggregation						Presentation Formats with varying Information Aggregation						Presentation Formats with varying Information Aggregation					
H5(a)	G _H = G _L	.79	1.45	-1.81	.070*	H5(b)	G _H = G _L	.153	.245	5.23	.022**	H5(c)	G _H > G _L	200.97	224.23	.91	.341
H6(a)	G _{DH} = G _{DL}	.78	1.47	-1.98	.048**	H6(b)	G _{DH} =G _{DL} <input checked="" type="checkbox"/>	.217	.283	1.85	.174	H6(c)	G _{DH} >G _{DL} <input checked="" type="checkbox"/>	213.20	258.59	3.70	.054*
H7(a)	T _H > T _L	1.45	1.71	-.73	.464	H7(b)	T _H > T _L	.220	.212	.03	.861	H7(c)	T _H > T _L	218.88	231.45	.28	.599

^a Mean by mean comparisons for accuracy are performed using non-parametric Mann-Whitney tests.

^b Mean by mean comparisons for consistency are performed using simple contrasts based on a generalized linear model for consistency as a dependent variable with time as a covariate as specified in Table 2. Tabulated means for consistency are estimated marginal means. The covariate for time is fixed at the mean of 224.62.

^c Mean by mean comparisons for time are performed using simple contrasts based on a generalized linear model for time as a dependent variable with ln(consistency) as a covariate as specified in Table 2. Tabulated means for time are estimated marginal means. The covariate for ln(consistency) is fixed at the mean of 3.13.

^d Hypotheses as specified in Chapter 2. Treatment 1 refers to the left hand side of the hypothesis equation. Treatment 2 refers to the right hand side. Abbreviations are as follows: G = Graphs, T = Tables; G_D = Graphs with Data Values; IA_H = High Information Aggregation; IA_L = Low Information Aggregation. G_H = Graphs with High Information Aggregation, G_L = Graphs with Low Information Aggregation, G_{DH} = Graphs with Data Values with High Information Aggregation, G_{DL} = Graphs with Data Values with Low Information Aggregation, T_H = Tables with High Information Aggregation, T_L = Tables with Low Information Aggregation. A symbol indicates that the respective hypothesis is supported.

^e p-values are two-tailed; ***, **, * represents significance at the 1%, 5% or 10% level.

There is a significant main effect of *information aggregation* on *consistency* (Wald- $\chi^2 = 4.26$; $p = .039$), no significant main effect of *presentation format* on *consistency* (Wald- $\chi^2 = 3.29$; $p = .193$), and no significant interaction effect between *information aggregation* and *presentation format* (Wald- $\chi^2 = 3.39$; $p = .184$).

To further analyze the effects of *presentation format* on *consistency*, mean by mean comparisons are performed with contrast analysis (Table 3) (Buckless and Ravenscroft 1990). A mean by mean comparison shows no difference between *graphs* (Table 1; mean consistency ratio = .194) and *tables* (.216) with respect to *consistency* (Wald- $\chi^2 = .62$; $p = .431$). This result contradicts H1(b). Comparing *graphs* (.194) to *graphs with data values* (.248) there is marginally significant difference in *consistency* (Wald- $\chi^2 = 3.20$; $p = .073$), judgments with *graphs* are more consistent which provides weak support for H2(b). There is no significant difference in *consistency* between *graphs with data values* (.248) and *tables* (.216) (Wald- $\chi^2 = 1.04$; $p = .308$). This result contradicts H3(b).

Next, the effect of *information aggregation* on *consistency* is analyzed. The mean by mean comparison shows *consistency* to be significantly higher for *high information aggregation* (.194) compared to *low information aggregation* (.245) (Table 3; Wald- $\chi^2 = 4.17$; $p = .041$). In line with the overall result of the generalized linear model for *information aggregation*, this result supports H4(b).

Lastly, the effect of *information aggregation* on *consistency* for a given *presentation format* is examined (Table 3). With *graphs* there is a significant difference in *consistency*, with *high information aggregation* (.153) leading to more consistent judgments than *low information aggregation* (.245) (Wald- $\chi^2 = 5.23$; $p = .022$). This result contradicts H5(b). *Consistency* with *graphs with data values* does not differ between *high information aggregation* (.217) and *low information aggregation* (.283) (Wald- $\chi^2 = 1.85$; $p = .174$), which supports H6(b). *Tables* do not differ in *consistency* with *high information aggregation* (.220) or *low information aggregation* (.212) (Wald- $\chi^2 = .03$; $p = .861$), a contradiction of H7(b).

Time

Also, a generalized linear model analysis using *time* as the dependent variable is performed (Table 2). *Time* has a normal distribution; hence, an identity link-function is utilized. *Information aggregation* and *presentation format* are specified as independent variables, while accounting for the effect of *consistency* as a covariate.

There is a marginally significant main effect of *information aggregation* on *time* (Wald- $\chi^2 = 3.80$; $p = .051$), no significant main effect of *presentation format* on *time* (Wald- $\chi^2 = 1.90$; $p = .387$) and no significant interaction effect between *information aggregation* and *presentation format* (Wald- $\chi^2 = 1.00$; $p = .607$).

To further examine the effects of *presentation format* on *time*, mean by mean comparisons are performed with contrast analysis (Table 3). A mean by mean comparison shows no difference in *time* between *graphs* (Table 1; mean judgment speed in seconds = 212.60) and *tables* (225.17) (Wald- $\chi^2 = .55$; $p = .460$). This result supports H1(c). Comparing *graphs* (212.60) to *graphs with data values* (235.90) there is no significant difference in *time* (Wald- $\chi^2 = 1.90$; $p = .168$), which provides support for H2(c). Also, there is no significant difference between *graphs with data values* (235.90) and *tables* (225.17) with respect to *time* (Wald- $\chi^2 = .41$; $p = .523$). This result supports H3(c).

Next, the effect of *information aggregation* on *time* is examined. The mean by mean comparison shows *time* to be marginally significant higher for *high information aggregation* (211.02) compared to *low information aggregation* (238.09) (Table 3; Wald- $\chi^2 = 3.80$; $p = .051$). In line with the overall result of the generalized linear model for *information aggregation*, this result provides weak support for H4(c).

Lastly, the effect of *information aggregation* on *time* for a given *presentation format* is examined (Table 3). With *graphs* there is no significant difference in *time* for *high information aggregation* (200.97) compared to *low information aggregation* (224.23) (Wald- $\chi^2 = .91$; $p = .341$). This contradicts H5(c). There is a marginally significant difference in *time* with *graphs with data values* for *high information aggregation* (213.20) compared to *low information aggregation* (258.59) (Wald- $\chi^2 = 3.70$; $p = .054$). *Graphs with data values* lead to faster judgments with *high information aggregation* compared to *low information aggregation*. This weakly supports H6(c). *Tables* do not differ in *time* with *high information aggregation* (218.88) or *low information aggregation* (231.45) (Wald- $\chi^2 = .28$; $p = .599$), a contradiction of H7(c).

4.3 Supplemental Analysis

Additional analyses were performed to examine the *perceived information load*, *perceived task complexity* and *perceived task realism* (Panel B in Table 4) of the participants for each combination of *presentation format* and *information aggregation* as well as the use of the data values in the *graphs with data values* treatments for both *high* and *low information aggregation* (Panel C in Table 4). The answer categories to the post-

experiment questionnaire are on a five-point Likert scale with 1 = strongly disagree and 5 = strongly agree.

An analysis of variance (ANOVA; not tabulated) shows that *information aggregation* significantly influences *perceived information load* ($F = 4.81$; $p = .029$). *Presentation format* has no significant effect ($F = .49$; $p = .613$). The mean of participants' responses for the statement „I felt overwhelmed by the amount of information provided” is 2.70 (standard deviation = 1.18) for *high information aggregation* and 3.04 (standard deviation = 1.06) for *low information aggregation* (Panel B in Table 4). The item being measured on a five-point Likert scale with 1 = strongly disagree and 5 = strongly agree. This result indicates that *information aggregation* decreases *perceived information load*.

Also being analyzed with an ANOVA (not tabulated), task complexity did not differ significantly with respect to both *information aggregation* ($F = 2.57$; $p = .110$) and *presentation format* ($F = .61$; $p = .559$). The overall mean of participants' responses for the statement „I found it difficult to solve the problem” is 3.49 (Panel B in Table 4; standard deviation = 1.03) on a five-point Likert scale with 1 = strongly disagree and 5 = strongly agree. Hence, the task can be interpreted as being of medium complexity.

Perceived task realism is measured with the statement “I appreciate the case study and the task as being realistic”, on a on a five-point Likert scale with 1 = strongly disagree and 5 = strongly agree. *Information aggregation* has no effect on *perceived task realism* (ANOVA; not tabulated; $F = .45$; $p = .503$), whereas *presentation format* significantly influences *perceived task realism* (ANOVA; not tabulated; $F = .3.51$; $p = .032$). Post-hoc comparisons with Bonferroni correction (not tabulated) show the difference between *tables* (mean = 3.25; Panel B in Table 4) and *graphs* (3.64) ($p = .059$) as well as the difference between (2) *tables* (3.25) and *graphs with data values* (3.61) ($p = .085$), to be marginally significant. These results indicate that both graphical formats are being perceived more realistic, which supports the notion of cognitive fit that multiattribute judgment tasks are facilitated by spatial *presentation formats*.

Panel C in Table 4 shows the descriptive statistics for the use of data values in the *graphs with data values* treatment. All items are measured on a five-point Likert scale with 1 = strongly disagree and 5 = strongly agree. Descriptive results show that on average participants use the data values as well as the graphic elements. This indicates that the *graphs with data values* treatment and the *graphs* treatment are perceived differently.

Table 4: Post-Experiment Questionnaire

Panel A: Descriptive Statistics for Perceived Holistic Processing^a and Perceived Analytic Processing^b with Mean; [Standard Deviation]; (Number of Observations)				
Presentation Format	Item/ Variable	Information Aggregation		Means
		High Aggregation (Ratio)	Low Aggregation (Ratio, Numerator, Denominator)	
Tables	Holistic Processing	3.79 [1.04] (n = 34)	4.17 [.66] (n = 35)	3.99 [.88] (n = 69)
	Analytic Processing	2.21 [1.01] (n = 34)	2.23 [1.03] (n = 35)	2.22 [1.01] (n = 69)
Graphs	Holistic Processing	4.09 [.58] (n = 33)	3.85 [.91] (n = 33)	3.97 [.76] (n = 66)
	Analytic Processing	2.21 [1.05] (n = 33)	1.97 [.90] (n = 34)	2.09 [.98] (n = 67)
Graphs with Data Values	Holistic Processing	3.89 [1.12] (n = 36)	3.68 [.84] (n = 34)	3.79 [.99] (n = 70)
	Analytic Processing	2.31 [1.22] (n = 36)	2.35 [.98] (n = 34)	2.33 [1.10] (n = 70)
Means	Holistic Processing	3.92 [.95] (n = 103)	3.90 [.83] (n = 102)	3.91 [.89] (n = 205)
	Analytic Processing	2.24 [1.09] (n = 103)	2.18 [.98] (n = 103)	2.21 [1.03] (n = 206)

Panel B: Descriptive Statistics for Perceived Information Load^c, Perceived Task Complexity^d and Perceived Task Realism^e with Mean; [Standard Deviation]; (Number of Observations)				
Presentation Format	Item/ Variable	Information Aggregation		Means
		High Aggregation	Low Aggregation	
Tables	Information Load	2.79 [1.27] (n = 34)	3.03 [.82] (n = 35)	2.91 [1.07] (n = 69)
	Task Complexity	3.32 [1.04] (n = 34)	3.51 [.78] (n = 35)	3.42 [.91] (n = 69)
	Task Realism	3.21 [1.15] (n = 34)	3.29 [.96] (n = 33)	3.25 [1.05] (n = 69)
Graphs	Information Load	2.48 [1.09] (n = 33)	3.03 [1.22] (n = 34)	2.76 [1.18] (n = 67)
	Task Complexity	3.58 [1.03] (n = 33)	3.62 [1.10] (n = 34)	3.60 [1.06] (n = 67)
	Task Realism	3.64 [.90] (n = 33)	3.64 [.93] (n = 33)	3.64 [.91] (n = 66)
Graphs with Data Values	Information Load	2.81 [1.17] (n = 36)	3.06 [1.13] (n = 34)	2.93 [1.15] (n = 70)
	Task Complexity	3.22 [1.15] (n = 36)	3.68 [1.01] (n = 34)	3.44 [1.10] (n = 70)
	Task Realism	3.51 [.85] (n = 35)	3.71 [.97] (n = 34)	3.61 [.91] (n = 69)
Means	Information Load	2.70 [1.18] (n = 103)	3.04 [1.06] (n = 103)	2.87 [1.13] (n = 206)
	Task Complexity	3.37 [1.08] (n = 103)	3.60 [.96] (n = 103)	3.49 [1.03] (n = 206)
	Task Realism	3.45 [.98] (n = 102)	3.54 [.96] (n = 102)	3.50 [.97] (n = 204)

Panel C: Descriptive Statistics for the Use of Data Values^f in the Graphs with Data Values Treatments with Mean; [Standard Deviation]; (Number of Observations)			
While analyzing the indicators...	Information Aggregation		Means
	High Aggregation	Low Aggregation	
... I have paid only attention to the specific numerical values	1.80 [.80] (n = 35)	2.21 [.98] (n = 34)	2.00 [.91] (n = 69)
... I have paid more attention to the specific numerical values	2.49 [.92] (n = 35)	2.56 [.96] (n = 34)	2.52 [.93] (n = 69)
... I have paid equal attention to the specific numerical values and the graphic elements	3.49 [.98] (n = 35)	3.18 [1.06] (n = 34)	3.33 [1.02] (n = 69)
... I have paid more attention to the graphic elements	3.26 [1.12] (n = 35)	3.50 [1.02] (n = 34)	3.38 [1.07] (n = 69)
... I have paid only attention to the graphic elements	2.43 [1.22] (n = 35)	2.38 [1.26] (n = 34)	2.41 [1.23] (n = 69)

Table 4: Post-Experiment Questionnaire

- ^a *Perceived Holistic Processing* is measured on a 5-point Likert scale with the statement: "In assessing the business units I have tried to capture business development as a whole in order to gain a general impression of the development over time", with 1 = strongly disagree and 5 = strongly agree.
- ^b *Perceived Analytic Processing* is measured on a 5-point Likert scale with the statement: "In assessing the business units I have analyzed the discrete numerical values of the indicators in order to calculate exact results for the business development", with 1 = strongly disagree and 5 = strongly agree.
- ^c *Perceived Information Load* is measured on a 5-point Likert scale with the statement: „I felt overwhelmed by the amount of information provided”, with 1 = strongly disagree and 5 = strongly agree.
- ^d *Perceived Task Complexity* is measured on a 5-point Likert scale with the statement: „I found it difficult to solve the problem”, with 1 = strongly disagree and 5 = strongly agree.
- ^e *Perceived Task Realism* is measured on a 5-point Likert scale with the statement, "I appreciate the case study and the task as being realistic", with 1 = strongly disagree and 5 = strongly agree.
- ^f *Use of Data Values* is measured on a 5-point Likert scale with 1 = strongly disagree and 5 = strongly agree.

5 Discussion and Conclusion

This study analyzes the effects of information aggregation and presentation format on judgment quality for a complex multiattribute judgment task in performance evaluation. From a performance view of judgment quality, graphs as a presentation format, either without or with data values, lead to judgments with the highest accuracy compared to tables. These results support cognitive fit theory and extend the empirical findings of Tuttle and Kershaw (1998) and Umanath and Vessey (1994), who also find graphs to be more accurate than tables for multiattribute judgment tasks. Also, in line with Frownfelter-Lohrke (1998), accuracy does not differ between the pure and combined graphical format.

From a process view of judgment quality, graphs without and with data values do not increase procedural consistency compared to tables, whereas graphs lead to more consistent criteria weights than graphs with data values. These results contradict cognitive fit theory, as graphs should lead to a more consistent mental representation for a complex multiattribute judgment task than tables. In contrast, Tuttle and Kershaw (1998) analyze consistency from a performance view and find graphs to be significantly more consistent than tables, when participants employ a holistic strategy.

Results of this study show that presentation format does not influence the speed of the judgment process. This finding is in line with cognitive fit theory and conforms to results from Tuttle and Kershaw (1998) and Umanath and Vessey (1994). Complex multiattribute judgment tasks require the problem solver to identify relationships among the data and to reference to the underlying data values. The mental transformations necessary for tabular presentation formats and the high amount of time necessary for graphs

to refer to underlying data values tend to offset each other. Furthermore, graphs with data values do not lead to a faster judgment process than graphs. Numerical values facilitate symbolic information; hence the corresponding analytic processes do not support a complex multiattribute judgment task and hinder a fast development of a consistent mental representation.

In this study, consistent with information-processing theory, information aggregation with constant information content leads to more accurate, consistent and fast judgments. From a performance view of judgment quality, results for accuracy support existing empirical evidence (Kaufman and Weber 2013, Tiessen 1976, Chervany and Dickson 1974, Ronen 1971). With respect to procedural rationality results show that positive effects of information aggregation can be extended to procedural consistency. With regard to procedural speed, results are in line with Otley and Dias (1982).

Also, the interplay of information aggregation and presentation format is analyzed. When information aggregation is high, graphs with data values lead to more accurate judgments and graphs lead to more accurate and consistent judgments compared to low information aggregation. Accuracy and consistency with tables does not differ significantly with a change in information aggregation. These results contradict cognitive fit theory, which asserts graphs (tables) to have (have not) integrating abilities and therefore not to be (to be) impaired from higher information load through decreased information aggregation. In contrast, Umanath and Vessey (1994) find accuracy with graphs not to differ with a change in information load. However, for tables they also do not find a significant effect on accuracy.

Descriptive statistics for all presentation formats show that judgment time for aggregated information is lower compared to disaggregated information. However, results are only significant for graphs with data values and therefore cognitive fit theory is not supported. In contrast, Umanath and Vessey (1994) find judgments with both, graphs and tables, to be significantly faster with a reduction in information load.

These results have several implications for researchers and practitioners. From a theoretical viewpoint this study supports the fundamentals of cognitive fit theory, as result accuracy is higher with graphs compared to tables. However, extensions of the notion of cognitive fit to procedural rationality, combined formats, and a change in information aggregation are not supported for the complex multiattribute judgment task being examined. This implies that further research is necessary to develop cognitive fit

theory towards an aspect of a general theory of problem solving. For example future studies could analyze whether the extensions of cognitive fit apply for complex symbolic choice tasks.

Furthermore, findings support the application of information-processing theory to procedural rationality, as information aggregation significantly increases procedural consistency, i.e. the consistent specification of criteria weights. Moreover, this finding is important for practitioners, as performance evaluation is a complex task to human judgment whereby the accurate response to a task is often not instantly obvious. In such a situation a consistent specification of criteria weights is especially important, when they serve as input for analytical models.

From a practical viewpoint, this study can offer guidelines for graphical representations of complex multiattribute judgment tasks. The widespread belief of practitioners that combined graphical formats are superior to pure graphical formats is not supported. Graphs with data values lead to less consistent performance measure weight specifications than graphs, which can pose a threat for performance evaluation. Therefore, for the present problem class an aggregated pure graphical format should be chosen to support accurate and consistent judgments.

Finally, this study is subject to limitations that restrict the extent to which the findings can be generalized. Participants were students with domain knowledge on the experimental task due to prior course work. Studies employing practicing professionals in problem-solving tasks of a similar kind report no differences between the more and less experienced groups (Eggleton et al. 1992, Stock and Watson 1984, Moriarity 1979). In addition, this experiment uses only performance measures which were taught to participants during class.

Also, this experiment did not employ monetary incentives. Participation was voluntary and no course credit was given for a correct judgment. Therefore it could not be completely prevented that some participants lacked motivation to provide effort for the task. However, on average results were significantly different from completely random judgments, indicating intrinsic motivation to solve the task.

Furthermore, despite careful considerations, it cannot be completely ruled out that criteria have been overlooked with respect to the design of the graphical presentation formats. Consistent with respect to existing studies, performance measures were presented as line graphs, and components of performance measures were depicted as bar

graphs in the respective treatments. In the treatments with graphs the vertical axis of the respective performance measure was aligned across alternatives in order to make them directly comparable and to avoid optical illusions. In both treatments, graphs and graphs with data values, the design was identical except for the added numerical values. For the tabular format, the treatments with low information aggregation differed from those with high information aggregation only by additional descriptions and numerical values of the performance measure components.

Future studies could address these limitations and use professionals, provide monetary incentives, and analyze whether the results also apply for other presentation formats. As information in this study was aggregated with information content held constant, future studies could assess information aggregation with a loss in information content.

Appendix

Appendix 1: Tables with High Aggregation

business unit A				
year	2009	2010	2011	2012
cash ratio [%]	81.35%	105.14%	101.4%	96.09%
operating profit margin [%]	7.76%	6.42%	14.07%	15.17%
ROIC [%]	18.86%	15.42%	21.58%	31.27%
debt to equity ratio [%]	162.95%	169.57%	138.44%	142.08%
WACC [%]	10.12%	12.13%	17.4%	19.65%
business unit B				
year	2009	2010	2011	2012
cash ratio [%]	108,90%	92,69%	111,46%	100,39%
operating profit margin [%]	11,51%	9,85%	16,92%	13,75%
ROIC [%]	25,52%	18,52%	39,52%	34,29%
debt to equity ratio [%]	91,47%	102,65%	103,27%	106,60%
WACC [%]	6,37%	4,31%	8,20%	4,19%
business unit C				
year	2009	2010	2011	2012
cash ratio [%]	110,11%	116,86%	93,54%	97,87%
operating profit margin [%]	26,06%	28,57%	16,60%	24,68%
ROIC [%]	35,22%	30,10%	24,66%	20,00%
debt to equity ratio [%]	84,00%	122,64%	158,00%	159,96%
WACC [%]	26,48%	26,10%	21,37%	21,00%

Appendix 2: Tables with Low Aggregation

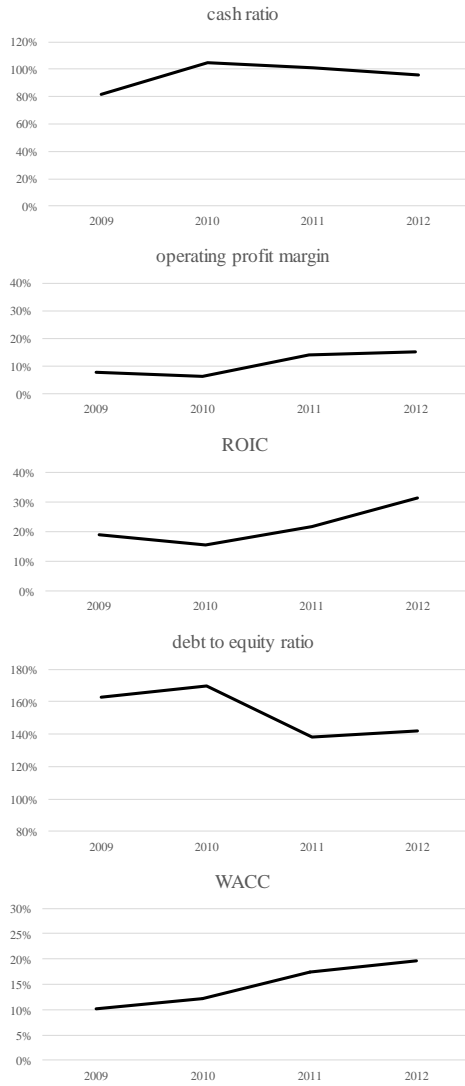
business unit A				
year	2009	2010	2011	2012
liquid assets [€]	2430	2660	2830	2210
÷ current liabilities [€]	2987	2530	2791	2300
= cash ratio [%]	81,35%	105,14%	101,40%	96,09%
EBIT [€]	198	183	410	455
÷ sales [€]	2551	2850	2915	3000
= operating profit margin [%]	7,76%	6,42%	14,07%	15,17%
EBIT [€]	198	183	410	455
÷ invested capital [€]	1050	1187	1900	1455
= ROIC [%]	18,86%	15,42%	21,58%	31,27%
debt (book value) [€]	3422	3561	3400	3599
÷ equity (book value) [€]	2100	2100	2456	2533
= debt to equity ratio [%]	162,95%	169,57%	138,44%	142,08%
weighted cost of debt (market value) [%]	4,12%	4,00%	8,50%	8,50%
+ weighted cost of equity (market value) [%]	6,00%	8,13%	8,90%	11,15%
= weighted average costs of capital (WACC) [%]	10,12%	12,13%	17,40%	19,65%

business unit B				
year	2009	2010	2011	2012
liquid assets [€]	1150	1001	1362	1300
÷ current liabilities [€]	1056	1080	1222	1295
= cash ratio [%]	108,90%	92,69%	111,46%	100,39%
EBIT [€]	536	389	830	720
÷ sales [€]	4655	3950	4905	5235
= operating profit margin [%]	11,51%	9,85%	16,92%	13,75%
EBIT [€]	536	389	830	720
÷ invested capital [€]	2100	2100	2100	2100
= ROIC [%]	25,52%	18,52%	39,52%	34,29%
debt (book value) [€]	3110	3295	3315	3422
÷ equity (book value) [€]	3400	3210	3210	3210
= debt to equity ratio [%]	91,47%	102,65%	103,27%	106,60%
weighted cost of debt (market value) [%]	2,94%	1,04%	3,35%	1,22%
+ weighted cost of equity (market value) [%]	3,43%	3,27%	4,85%	2,97%
= weighted average costs of capital (WACC) [%]	6,37%	4,31%	8,20%	4,19%

business unit C				
year	2009	2010	2011	2012
liquid assets [€]	1100	2530	2666	2936
÷ current liabilities [€]	999	2165	2850	3000
= cash ratio [%]	110,11%	116,86%	93,54%	97,87%
EBIT [€]	615	1025	699	1423
÷ sales [€]	2360	3588	4210	5766
= operating profit margin [%]	26,06%	28,57%	16,60%	24,68%
EBIT [€]	615	1025	699	1423
÷ invested capital [€]	1746	3405	2835	7115
= ROIC [%]	35,22%	30,10%	24,66%	20,00%
debt (book value) [€]	2100	3066	3950	3999
÷ equity (book value) [€]	2500	2500	2500	2500
= debt to equity ratio [%]	84,00%	122,64%	158,00%	159,96%
weighted cost of debt (market value) [%]	10,32%	15,20%	9,75%	10,20%
+ weighted cost of equity (market value) [%]	16,16%	10,90%	11,62%	10,80%
= weighted average costs of capital (WACC) [%]	26,48%	26,10%	21,37%	21,00%

Appendix 3: Graphs with High Aggregation

business unit A



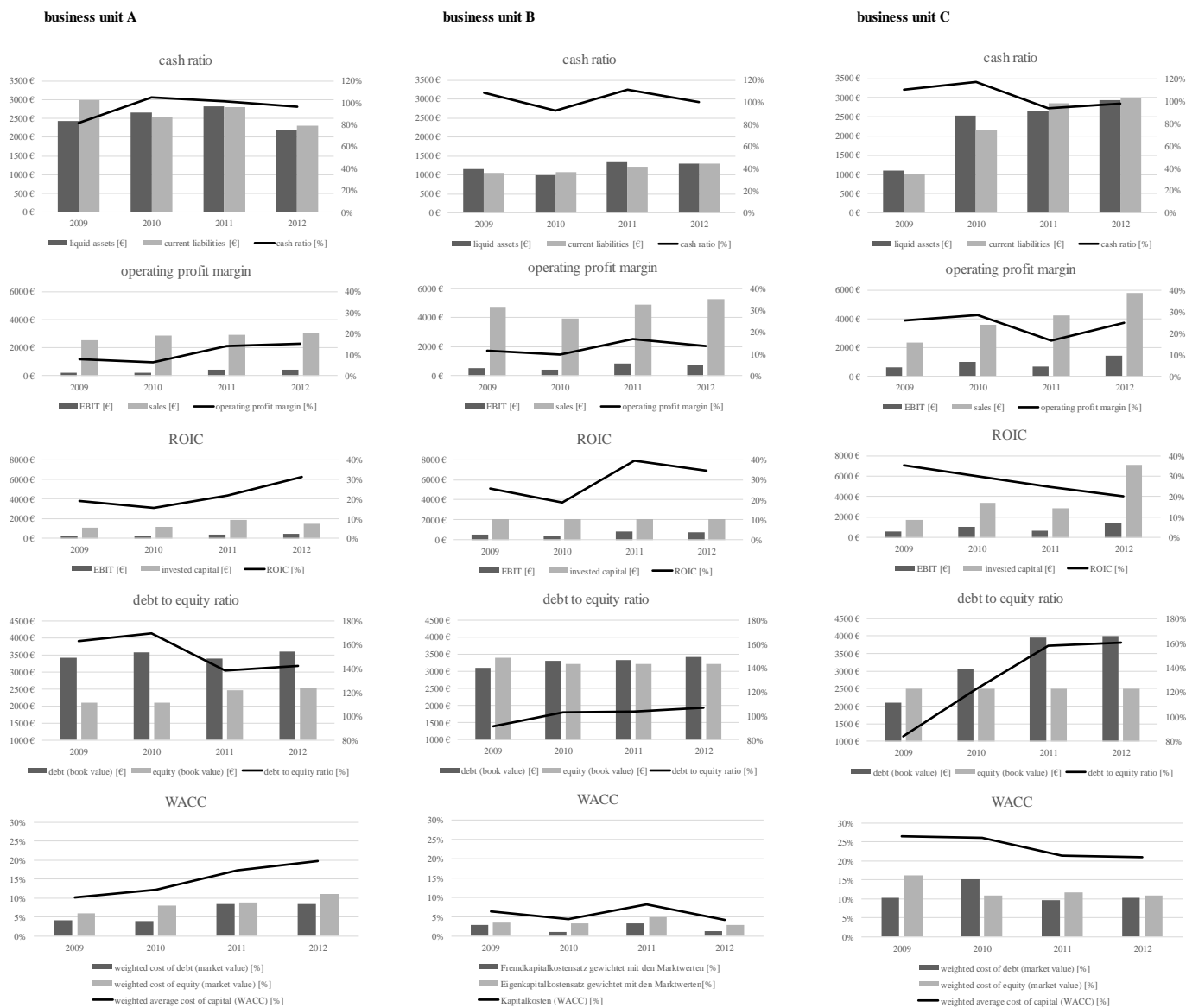
business unit B



business unit C



Appendix 4: Graphs with Low Aggregation

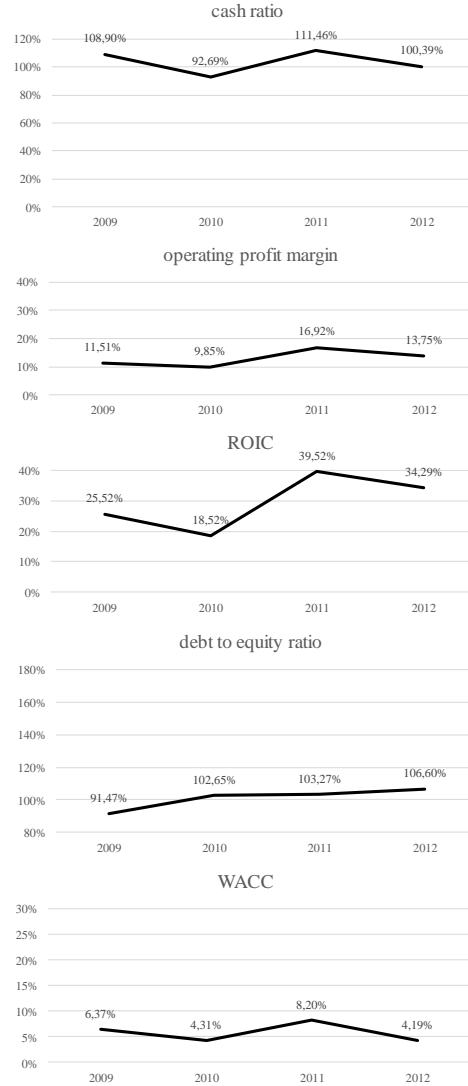


Appendix 5: Graphs with Data Values and High Aggregation

business unit A



business unit B

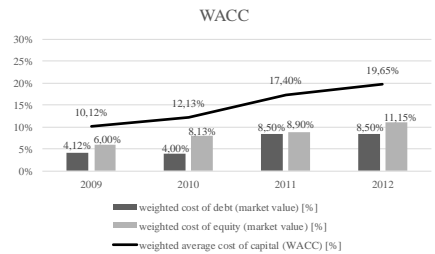
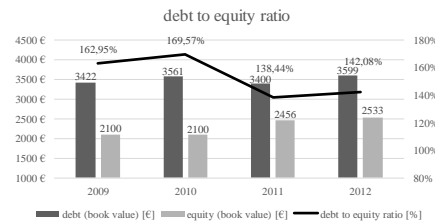
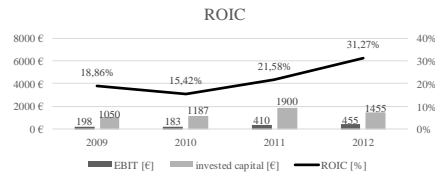
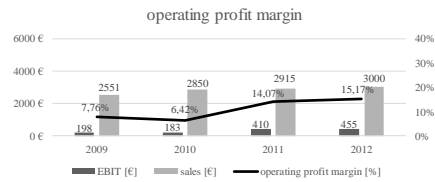
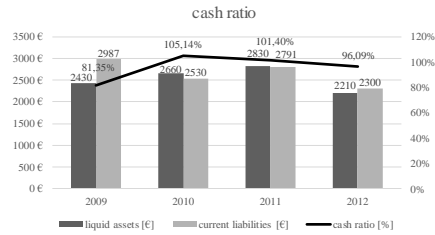


business unit C

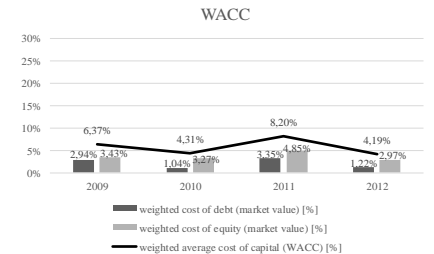
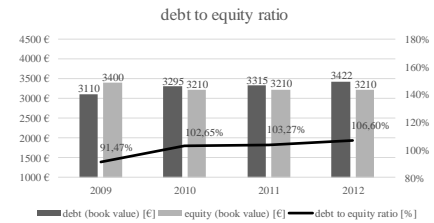
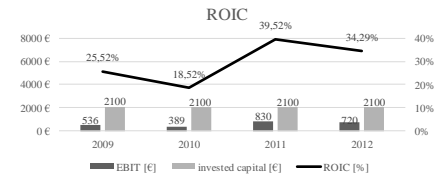
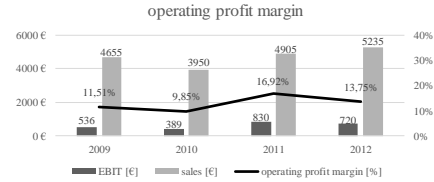
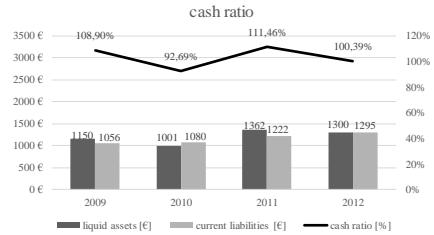


Appendix 6: Graphs with Data Values and Low Aggregation

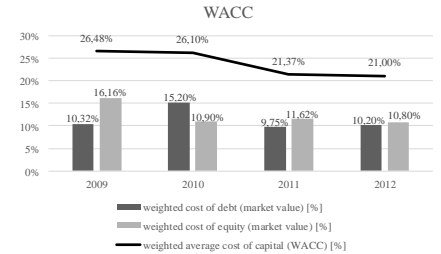
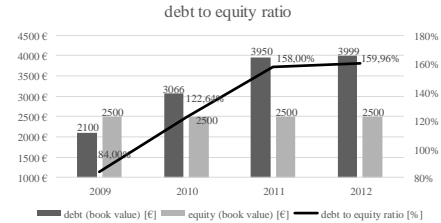
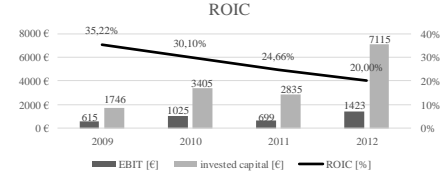
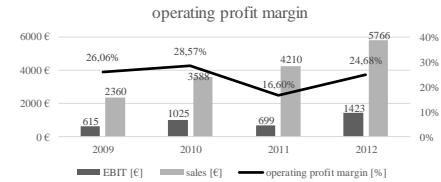
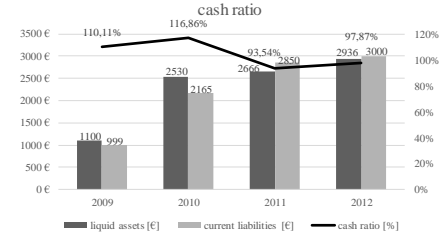
business unit A



business unit B



business unit C



References

- Ashton, R. H., and S. S. Kramer. 1980. Students as surrogates in behavioral accounting research: Some evidence. *Journal of Accounting Research* 18 (1): 1-15.
- Banker, R. D., H. Chang, and M. Pizzini. 2011. The judgmental effects of strategy maps in balanced scorecard performance evaluations. *International Journal of Accounting Information Systems* 12 (4): 259-279.
- Bolles, E. B. 1991. *A Second way of knowing: The riddle of human perception*. New York, NY: Prentice Hall Press.
- Bonner, S. E. 2008. *Judgment And Decision Making in Accounting*. Upper Saddle River, NJ: Pearson/ Prentice Hall.
- Buckless, F. A., and S. P. Ravenscroft. 1990. Contrast Coding: A Refinement of ANOVA in Behavioral Analysis. *The Accounting Review* 65 (4): 933-945.
- Campbell, D. J. 1988. Task Complexity: A Review and Analysis. *The Academy of Management Review* 13 (1): 40-52.
- Cardinaels, E. 2008. The interplay between cost accounting knowledge and presentation formats in cost-based decision-making. *Accounting, Organizations and Society* 33 (6): 582-602.
- Chervany, N. L., and G. W. Dickson. 1974. An Experimental Evaluation of Information Overload in a Production Environment. *Management Science* 20 (10): 1335-1344.
- Chewing Jr., E. G., and A. M. Harrell. 1990. The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. *Accounting, Organizations and Society* 15 (6): 527-542.
- DeSanctis, G. 1984. Computer Graphics as Decision Aids: Directions for Research. *Decision Sciences* 15 (4): 463-487.
- DeSanctis, G., and S. L. Jarvenpaa. 1989. Graphical presentation of accounting data for financial forecasting: An experimental investigation. *Accounting, Organizations and Society* 14 (5-6): 509-525.
- Driver, M. J., and S. Streufert. 1969. Integrative complexity: An approach to individuals and groups as information-processing systems. *Administrative Science Quarterly* 14 (2): 272-285.
- Dyer, J. S., P. C. Fishburn, R. E. Steuer, J. Wallenius, and S. Zionts. 1992. Multiple criteria decision making, multiattribute utility theory: the next ten years. *Management Science* 38 (5): 645-654.
- Eggleton, I. R. C., P. Knipping, and B. McCulloch. 1992. *Schematic faces and the communication of multivariate financial information in a bankruptcy prediction task*. Unpublished Working Paper. University of Western Australia.
- Einhorn, H. J., and R. M. Hogarth. 1981. Behavioral Decision Theory: Processes of Judgment and Choice. *Journal of Accounting Research* 19 (1): 1-31.
- Eppler, M. J., and J. Mengis. 2004. The Concept of Information Overload - A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society: An International Journal* 20 (5): 1-20.

- Frownfelter-Lohrke, C. 1998. The effects of differing information presentations of general purpose financial statements on users' decisions. *Journal of Information Systems* 12 (2): 99-107.
- Gupta, M., and R. R. King. 1997. An Experimental Investigation of the Effect of Cost Information and Feedback on Product Cost Decisions. *Contemporary Accounting Research* 14 (1): 99-127.
- Harvey, D. W., J. G. Rhode, and K. A. Merchant. 1979. Accounting aggregation: User Preferences and Decision Making. *Accounting, Organizations and Society* 4 (3): 187-210.
- Hülle, J., R. Kaspar, and K. Möller. 2011. Multiple Criteria Decision-Making in Management Accounting and Control-State of the Art and Research Perspectives Based on a Bibliometric Study. *Journal of Multi-Criteria Decision Analysis* 18 (5-6): 253-265.
- IBM Corporation. 2012. IBM SPSS Advanced Statistics 21. Copyright IBM Corporation 1989.
- Kahneman, D., and A. Tversky. 1996. On the reality of cognitive illusions. *Psychological Review* 103 (3): 582-591.
- Kaufmann, C., and M. Weber. 2013. Sometimes less is more – The influence of information aggregation on investment decisions. *Journal of Economic Behavior & Organization* 95: 20-33.
- Kelton, A. S., R. R. Pennington, and B. M. Tuttle. 2010. The effects of information presentation format on judgment and decision making: A review of the information systems research. *Journal of Information Systems* 24 (2): 79-105.
- Libby, R. 1981. *Accounting and Human Information Processing: Theory and Applications*. Englewood Cliffs, NJ: Prentice Hall.
- Libby, R., R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27 (8): 775-810.
- MacKay, D. B., and A. Villarreal. 1987. Performance differences in the use of graphic and tabular displays of multivariate data. *Decision Sciences* 18 (4): 535-546.
- Merchant, K. A., and W. A. van der Stede. 2012. *Management control systems. Performance measurement, evaluation and incentives*. 3. Edition. Harlow, U.K.: Pearson, Financial Times/ Prentice Hall.
- Moriarty, S. 1979. Communicating financial information through multidimensional graphics. *Journal of Accounting Research* 17 (1): 205-224.
- Newell, A., and H. A. Simon. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Nibbelin, M. C., C. D. Bailey, and R. W. Zmud. 1992. The effects of mode of information presentation and cognitive style on bond rating change decisions. *Advances in Accounting* 10: 159-174.
- Otley, D. T., and F. J. B. Dias. 1982. Accounting Aggregation and Decision-Making Performance: An Experimental Investigation. *Journal of Accounting Research* 20 (1): 171-188.
- Payne, J. W. 1982. Contingent decision behavior. *Psychological Bulletin* 92 (2): 382-402.

- Ronen, J. 1971. Some Effects of Sequential Aggregation in Accounting on Decision-Making. *Journal of Accounting Research* 9 (2): 307-332.
- Saaty, T. L. 1980. *The analytic hierarchy process: Planning, priority setting, resource allocation*. New York, NY: McGraw-Hill.
- Saaty, T. L. 1990. How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research* 48 (1): 9-26.
- Saaty, T. L. 1994. How to Make a Decision: The Analytic Hierarchy Process. *Interfaces* 24 (6): 19-43.
- Schroder, H. M., M. J. Driver, and S. Streufert. 1967. *Human Information Processing*. New York, NY: Holt, Rinehart and Winston.
- Simon, H. A. 1976. From substantive to procedural rationality. In: S. J. Latsis (Ed.): *Method and Appraisal in Economics*. Cambridge, U.K.: Cambridge University Press.
- Simon, H. A. 1978a. Information-processing theory of human problem solving. *Handbook of learning and cognitive processes* 5: 271-295.
- Simon, H. A. 1978b. Rationality as Process and as Product of Thought. *American Economic Review* 68 (2): 1-16.
- Simon, H. A., and G. Lea. 1974. Problem solving and rule induction: A unified view. In: L. W. Gregg (Ed.): *Knowledge and Cognition*. Pocomac, MD: Lawrence Erlbaum Associates: 105-128.
- So, S., and M. Smith. 2004. Multivariate decision accuracy and the presentation of accounting information. Corporate financial communication and voluntary disclosure. *Accounting Forum* 28 (3): 283-305.
- Speier, C. 2006. The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies* 64 (11): 1115-1131.
- Speier, C., I. Vessey, and J. S. Valacich. 2003. The Effects of Interruptions, Task Complexity, and Information Presentation on Computer-Supported Decision-Making Performance. *Decision Sciences* 34 (4): 771-797.
- Stock, D., and C. J. Watson. 1984. Human judgment accuracy, multidimensional graphics, and humans versus models. *Journal of Accounting Research* 22 (1): 192-206.
- Tiessen, P. 1976. *An investigation of some effects of data aggregation and cognitive style on decision making performance*. Dissertation. Minnesota.
- Tufte, E. R. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.
- Tuttle, B., and R. Kershaw. 1998. Information presentation and judgment strategy from a cognitive fit perspective. *Journal of Information Systems* 12 (1): 1-17.
- Umanath, N. S., and I. Vessey. 1994. Multiattribute Data Presentation and Human Judgment: A Cognitive Fit Perspective. *Decision Sciences* 25 (5-6): 795-824.
- Vessey, I. 1991. Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature. *Decision Sciences* 22 (2): 219-240.

- Vessey, I. 1994. The effect of information presentation on decision making: A cost-benefit analysis. *Information and Management* 27 (2): 103-119.
- Vessey, I. 2006. The Theory of Cognitive Fit. In: P. Zhang, and D. Galletta (Ed.): *Human-Computer Interaction and Management Information Systems: Foundations* 5. Armonk, NY: M. E. Sharpe: 141-183.
- Wallenius, J., J. S. Dyer, P. C. Fishburn, R. E. Steuer, S. Zionts, and K. Deb. 2008. Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science* 54 (7): 1336-1349.
- Waller, W. S., B. Shapiro, and G. Sevcik. 1999. Do cost-based pricing biases persist in laboratory markets? *Accounting, Organizations and Society* 24 (8): 717-739.
- Wilson, E. V., and T. Addo. 1994. An investigation of the relative presentation efficiency of computer-displayed graphs. *Information and Management* 26 (2): 105-115.
- Wood, R. E. 1986. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes* 37 (1): 60-82.
- Wright, W. F. 1995. Superior loan collectibility judgments given graphical displays. *Auditing. A Journal of Practice & Theory* 14 (2): 144-154.
- Yigitbasioglu, O. M., and O. Velcu. 2012. A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems* 13 (1): 41-59.

Effort Allocation in Multi-Task Environments – The Interplay of Performance Measure Characteristics and Task Difficulty

Christian Meier[✉]

Abstract

To motivate employees to allocate effort with the goal of creating firm value, a common approach in multi-task environments is to evaluate performance with multiple performance measures. However, effects of interactions between performance measure characteristics and task characteristics on effort allocation remain unclear in a multi-task environment. Therefore, this study examines the interplay of performance measure characteristics, that is, sensitivity and precision, as well as task difficulty on effort allocation in a multi-task laboratory experiment. Results show that, given a performance-based compensation system sensitivity influences effort allocation, irrespective of task difficulty. Under a fixed-wage compensation system, sensitivity has no effect on effort allocation. Hence, task difficulty does not influence the effect of sensitivity on effort allocation. Under a performance-based compensation system, when precision on the easy task decreases, significantly more effort is allocated to the difficult task. However, when precision on the difficult task decreases, effort allocation is not influenced. In contrast, given a fixed-wage compensation system, when precision on the difficult task decreases more effort is shifted towards the difficult task, whereas a change in precision on the easy task does not influence effort allocation. Thus, task difficulty moderates the effect of precision on effort allocation for both compensation systems. The results of this study are of interest for designers of management control, information, and compensation systems.

Keywords: effort allocation, incentive systems, multi-task environment, performance measure characteristics, task characteristics.

[✉] Christian Meier: University of Passau, Chair of Accounting and Control, Innstraße 27, 94032 Passau, Germany, E-Mail: controlling@uni-passau.de; Phone: 0851-509-3271.

1 Introduction

In multi-task environments, employees exert effort on multiple dimensions of a single task, on multiple tasks, or on a combination thereof (e.g. Holmstrom and Milgrom 1991, Baker 1992, 2002, Feltham and Xie 1994, Hemmer 1996). They repeatedly have to decide how to allocate effort between task dimensions or tasks (Sprinkle and Williamson 2007).¹ For example, production workers are often simultaneously responsible for the quantity and quality of a single product, whereas salesmen have sequential duties such as customer care, stocktaking, ordering and cashing-up.

To motivate employees to allocate effort on multiple tasks with the goal of creating firm value, a common approach is to evaluate their performance with multiple performance measures (Feltham and Xie 1994, Sprinkle and Williamson 2007). These performance measures are often linked with performance-based monetary incentives in employees' compensation contracts in order to reward them for generating good results (Merchant and van der Stede 2012). The examination of this performance-evaluation use of management accounting information in multi-task environments has become a focus of a number of experimental studies (Brüggen and Moers 2007, Kachelmeier et al. 2008, Sloof and van Praag 2008, Brüggen 2011, Hecht et al. 2012, Hannan et al. 2013).

Effort allocation can be influenced by performance measure characteristics and task characteristics (Bonner and Sprinkle 2002). Two important performance measure characteristics that influence effort allocation are sensitivity and precision. They have been extensively focused on by agency theoretic studies in the performance measurement literature (e.g. Holmstrom 1979, Banker and Datar 1989, Feltham and Xie 1994, Feltham and Wu 2000, Datar et al. 2001, Christensen and Feltham 2008, Demski 2008, Demski et al. 2009). This literature, which typically focuses on the linear agency model (LEN model², Holmstrom and Milgrom 1987), shows that, other things equal, performance measures with high precision and large sensitivity to an agent's effort are preferred by a risk-averse agent, when choosing a performance-based compensation contract (Lambert 2007). Thus, imprecise and ill-calibrated performance measures can distort effort allocation in a multi-task environment (Prendergast 1999, 2002). Along with these essential theoretical contributions in identifying performance measure characteris-

¹ Effort allocation refers to the tasks which an individual chooses to exert, and how to distribute effort between tasks. In this study tasks are substitutes and time is limited. Thus, more effort allocated to one task leads to less effort allocated to the other task (Bonner and Sprinkle 2002).

² LEN stands for Linear compensation, negative Exponential utility, and Normally distributed performance measures.

tics that are important when deciding upon including a performance measure into a compensation system, several empirical studies emerged, which examine behavioral effects of performance measure characteristics (Moers 2000, Brügger 2005, Krishnan et al. 2005, Brügger and Moers 2007, Sloof and van Praag 2008, 2010, Brügger 2011). Recent studies indicate that precision can influence effort intensity (Sloof and van Praag 2010) and distort effort allocation (Brügger 2005, Brügger and Moers 2007, Brügger 2011), for a given performance-based compensation system.

Despite the growing empirical and broad analytical research on performance measure characteristics, Bai et al. (2010) observe that empirical accounting research has paid relatively little attention to task characteristics. In line with this, the effect of performance measure precision on effort allocation in a multi-task environment has only been examined in studies where the same task is employed twice, neglecting effects of task characteristics (Brügger 2005, Brügger and Moers 2007, Sloof and van Praag 2008, Brügger 2011). Agency theoretic performance measurement models neglect task characteristics as well and implicitly assume that all tasks are basically the same with respect to task difficulty, i.e. have equal marginal costs of effort (e.g. Feltham and Wu 2000). However, Schnedler (2006) analytically shows that performance measure characteristics, including sensitivity and precision, are not sufficient to predict an agent's effort allocation when tasks vary in difficulty (i.e. marginal costs of effort differ across tasks). Moreover, Bailey and Fessler (2011) indicate that task characteristics, that is, task complexity and task attractiveness can moderate the effect of performance-based monetary incentives on task performance. Hence, for a performance-based compensation system it needs to be examined, whether task difficulty influences the effect of precision and sensitivity on effort allocation.

With respect to precision, economic theory predicts that, when a performance measure becomes too imprecise, linking performance-based monetary incentives to that performance measure becomes too expensive, as a risk-averse agent needs to be compensated for the higher variance in compensation (Prendergast 1999, 2002). Holmstrom and Milgrom (1991) address this issue and posit that a fixed-wage may thus be optimal in a multi-task environment. A fixed-wage compensation system does not link performance measures to compensation and hence agents should be indifferent between a range of effort allocations and follow a principal's preferred effort allocation. Therefore, it needs to be analyzed, if potential distortions of effort allocation caused by an interaction of

performance measure characteristics and task characteristics are restricted to performance-based compensation systems or if they also appear for a fixed-wage.

To address these issues, this study examines the effects of performance measure characteristics, that is, sensitivity and precision as well as task difficulty on effort allocation in a multi-task environment. A real-effort laboratory experiment is conducted, where precision, sensitivity and task difficulty are varied within-subjects. Compensation system is varied between-subjects with either a fixed-wage or performance-based monetary incentives. The following research question is stated: Does task difficulty influence the effects of sensitivity and precision on effort allocation in a multi-task environment? To derive the hypotheses, economic agency theory and psychological expectancy theory (Vroom 1964) are utilized.

Results show that, given a performance-based compensation system, sensitivity influences effort allocation, irrespective of task difficulty. Under a fixed-wage, sensitivity has no effect on effort allocation. Hence, task difficulty does not influence the effect of sensitivity on effort allocation. Under a performance-based compensation system, when precision on the easy task decreases, significantly more effort is allocated to the difficult task. However, when precision on the difficult task decreases, effort allocation is not influenced. In contrast, given a fixed-wage, when precision on the difficult task decreases more effort is shifted towards the difficult task, whereas a change in precision on the easy task does not influence effort allocation. Thus, task difficulty moderates the effect of precision on effort allocation for both compensation systems.

In analyzing the interplay of performance measure characteristics and task characteristics, this study contributes to the design of optimal management control, information, and compensation systems. Particularly, results have implications for how sensitive and precise performance measures should be, when underlying tasks have different difficulties. Results imply that the effect of sensitivity on effort allocation is reliably predictable. Higher sensitivity on a task increases effort allocated to that task, irrespective of task difficulty, and only for performance-based monetary incentives. Moreover, results show that the effect of precision on effort allocation is difficult to predict. When a task is measured imprecisely, depending on task difficulty and compensation system, effort allocation is shifted towards the respective task, does not change, or is allocated to another task. Hence, to avoid effort distortions tasks should be measured as precisely as it is economically feasible.

Also, this study contributes to the request of Bonner and Sprinkle (2002), as it is the first to employ a multi-task environment with two separate and distinct tasks which differ in difficulty. Subjects can select which task to work on, and choose their effort allocation without obstruction. Bonner and Sprinkle (2002) call for accounting researchers to employ two separate and distinct tasks in a laboratory experiment, whereby subjects need to explicitly decide which task to work on, and how much effort to devote to each task.

The remainder of this paper is structured as follows: The next section develops hypotheses based on economic agency theory and psychological expectancy theory. The two subsequent sections describe the experimental design and results. The final section provides a discussion of the results including their limitations and implications.

2 Theory and Hypotheses

The hypotheses in the following sections are stated for a two-task environment, with each task being measured with a separate performance measure. Performance measures are assumed to be uncorrelated. Tasks are substitutes, differ in difficulty, and time is limited. Thus, more effort allocated to one task leads to less effort allocated to the other task.

Kachelmeier (1996) proposes to move from experiments that demonstrate results from economic models to experiments that investigate deviations from economic models. Moser (1998) recommends designing experiments with greater tension between competing theories. Therefore, in the following sections, hypotheses are stated from an economic agency theory perspective as an agency theoretic prediction (ATP) and from a psychological expectancy theory perspective as a behavioral theory prediction (BTP).

2.1 Behavioral Assumptions of Agency Theory

Standard assumptions in economic agency theory are that individuals are risk-averse expected utility maximizers (e.g. Baiman 1982, 1990, Christensen and Feltham 2008). They are rational and have well-defined preferences that conform to the axioms of expected utility theory. Economic agents are motivated only by self-interest and their utility function consists of wealth and leisure. They prefer increases in wealth and increases in leisure, i.e. reductions in effort. Hence, they only exert effort on a task, if it contributes to their own economic well-being. Furthermore, incentives that are not contingent on performance are generally not sufficient to induce agent effort (Bonner and Sprinkle

2002). In a multi-task environment, if compensation systems provide performance-based pay only for a subset of tasks, agents do not provide effort on tasks for which monetary incentives are not provided. This can lead to effort distortions, if important tasks are not linked to monetary incentives, as they are expensive or difficult to measure (Hecht et al. 2012). Holmstrom and Milgrom (1991) relax the assumption that non-contingent pay is not sufficient to induce agent effort. They posit that even without financial incentives agents put more than minimum effort into tasks for fixed wages because of some intrinsic motivation to work. Moreover, given a fixed-wage compensation system, an agent is indifferent between a certain range of effort allocations in a multi-task environment and follows the principal's preferences.

2.2 Behavioral Assumptions of Expectancy Theory

Psychological expectancy theory states that individuals are rational and act to maximize expected satisfaction with desired outcomes (Vroom 1964, Bonner and Sprinkle 2002). According to expectancy theory, an individual's motivation in a given situation is a function of two factors: (1) effort-outcome expectancy, which is defined as a current belief concerning the likelihood that a particular effort will be followed by a particular outcome, i.e. reward (e.g. higher pay for higher effort), and (2) valence, which refers to affective orientations toward particular outcomes (Vroom 1964). The effort-outcome expectancy can be further differentiated into the (1) effort-performance expectancy and the (2) performance-outcome expectancy (Bonner and Sprinkle 2002, Sloof and van Praag 2008). Effort-performance expectancy is a current belief that effort positively influences performance (Vroom 1964). The stronger the perceived relationship between effort and performance, the higher is the individual's motivation to allocate effort to a task (Sloof and van Praag 2008, 2010). The performance-outcome expectancy, which is also referred to as instrumentality, is a subject's expectation that better performance results in a higher reward. An increase in instrumentality also has a positive effect on motivation to exert effort on a task. Valence captures the attractiveness, desirability and anticipated satisfaction of different rewards for an individual (Vroom 1964). Generally, monetary payoff has a higher valence than no pay. However, the valence of a fixed-wage relative to performance-based monetary incentives depends on their relative payoff magnitude (Bonner and Sprinkle 2002).

The motivation created by effort-performance expectancy, instrumentality and valence induces individuals in a multi-task environment to choose an effort allocation that leads to a desired outcome (Bonner and Sprinkle 2002).

2.3 Sensitivity and Precision

Performance measure sensitivity and performance measure precision originate in the agency theoretic literature in accounting (Banker and Datar 1989, Lambert 2007). Sensitivity is defined as the degree with which a performance measure changes (in expected value), if an agent's effort changes.³ Precision is defined as the inverse of the variance of a performance measure's additive noise term (Banker and Datar 1989, Lambert 2007).⁴ It is a common assumption in the performance measurement literature, which regularly employs the linear agency model (LEN model), to utilize precision as an additive noise term, which is normally distributed with mean zero and is independent from an agent's effort (e. g. Holmstrom and Milgrom 1991, Feltham and Xie 1994, Feltham and Wu 2000, Datar et al. 2001, Baker 2002, Christensen and Feltham 2008, Demski et al. 2009).

It is important to differentiate between two fundamental situations:⁵ (1) an agent has a set of compensations contracts to choose from, and (2) the compensation system is given. In the second situation, which this study focuses on, the problem of contract choice is either excluded, or the agent has chosen an optimal contract from the available set of contracts. In this situation, an agent's problem in a multi-task environment is to choose an optimal allocation of effort. For a given performance-based compensation system, if two tasks differ with respect to performance measure sensitivity, other things equal, an agent will allocate more effort to the task measured with higher sensitivity, as he will receive a higher pay for a unit of effort. However, an agent cannot influence

³ Performance measure sensitivity as an effort-performance sensitivity needs to be differentiated from pay-performance sensitivity as a performance-outcome sensitivity as specified by Jensen and Murphy (1990). Jensen and Murphy (1990, p. 227) characterize pay-performance sensitivity as: "(...) the dollar change in the CEO's wealth associated with a dollar change in the wealth of shareholders." Sensitivity in this sense is not a performance measure characteristic but an incentive scheme variable, as it assumes variable pay and is focused on executive compensation and the relation between agent wealth and principal wealth.

⁴ The performance measure noise term is also known as uncontrollable component, observation error, or error term (e.g. Feltham and Xie 1994, Prendergast 2002).

⁵ The first situation is an agent's problem of contract choice. In this situation, a risk averse agent's utility function does directly recognize precision (i.e. the risk measured by the noise term in a performance-based compensation contract), as it is reflected in the risk premium (Sloof and van Praag 2008). Thus, other things equal, performance measures with large sensitivity to an agent's actions and high precision (low noise) are preferred by a risk-averse agent, when choosing a performance-based compensation contract (Lambert 2007). When choosing a fixed-wage compensation contract, neither sensitivity nor precision influence an agent's compensation and are thus unimportant for his contract choice.

performance measure precision (measured with an additive noise term) and hence the risk premium does not change with his choice of effort allocation. Thus, there is no trade-off of risk and incentives (Prendergast 1999, 2002), as less effort allocated to a task does not decrease risk (Sloof and van Praag 2008). Hence, an agent's choice of effort allocation is independent from precision. Given a fixed-wage compensation system, neither sensitivity nor precision influence an agent's compensation and therefore do not influence his effort allocation choice.

From an expectancy theory perspective, under a performance-based compensation system, performance measure sensitivity and performance measure precision influence an individual's effort-performance expectancy. Other things being equal, by definition, an increase in sensitivity increases the relationship between effort and performance for a task, i.e. the effort-performance expectancy. Therefore, similar to agency theory, an individual's motivation to allocate effort to the respective task increases and he shifts effort towards the task measured with higher sensitivity. Performance measure precision also influences an individual's effort-performance expectancy. Other things being equal, a decrease in precision for a task, that is, an increase in performance measure variance, negatively influences the perceived relationship between effort and performance for that task. Therefore, in contrast to agency theory, an individual's motivation to allocate effort to the imprecise task decreases and he shifts effort away from that task (Sloof and van Praag 2008, 2010). By definition, a fixed-wage as an outcome has no relation to performance. Thus, under a fixed-wage, sensitivity and precision cannot influence the overall link between effort and rewards, that is, the effort-outcome expectancy, and they become irrelevant for an individual's effort allocation choice.

Overall, both theories offer the same predictions, except for the effect of performance measure precision on effort allocation under a performance-based compensation system. Agency theory in its linear form posits that precision has no effect on effort allocation, whereas expectancy theory predicts a shift away from tasks measured with lower precision.

2.4 Task Difficulty

From an agency theory perspective, task difficulty⁶ attenuates the relationship between effort and performance, that is, effort becomes less sensitive to performance.

⁶ Task difficulty is generally seen as an aspect of 'task complexity' besides constructs such as 'task structure' or the algorithmic/ heuristic problem solving dimension of tasks (Bonner and Sprinkle 2002).

Task difficulty also increases the marginal costs of effort (Bonner and Sprinkle 2002). With respect to the relationship between effort and performance, March and Simon (1958) emphasize the skill requirements of difficult tasks. As task difficulty increases, the gap between subjects' skill and tasks' skill requirements increases, making it less likely that effort will positively influence performance and, *ceteris paribus*, an agent will allocate effort to an easier task, as he will receive a higher pay for a unit of effort (Bonner and Sprinkle 2002). Furthermore, an increase in task difficulty leads to an increase in effort requirements for a task (Wood 1986, Campbell 1988), that is, marginal costs of effort increase with task difficulty (Schnedler 2006). Agents consider the costs of effort as well as rewards related to a task before performing that task. Thus, they weigh the rewards associated with increasing performance against the marginal costs of effort required to realize higher performance. Assuming that the rewards of performance are approximately equal under easy and difficult tasks, the higher marginal costs of effort are more likely to outweigh the rewards in difficult tasks. If marginal costs of effort outweigh the rewards, then agents will trade off a reduction in performance for reductions in effort. Therefore, given a performance-based compensation system, with tasks being substitutes and time being limited, agents will allocate relatively more effort to an easier task.

Both effects, the attenuated effort-performance sensitivity and the higher marginal costs of effort on a difficult task, lead to effort allocated away from difficult tasks to easier tasks for a performance-based compensation system. However, for a fixed-wage compensation system, the adverse effects of task difficulty do not influence an agent's compensation. Accordingly, an agent is indifferent between a certain range of effort allocations in a multi-task environment and follows the principal's preferences.

From an expectancy theory perspective, task difficulty decreases the effort-performance expectancy, as task difficulty makes it less likely that effort will positively influence performance (Smith et al. 1976, Bonner and Sprinkle 2002). Smith et al. (1976) show in an experimental study that effort-performance expectancy declines rapidly as a function of task difficulty, when performance is measured as an absolute criterion. Therefore, under a performance-based compensation system, an individual's motivation to allocate

In a very general sense, task complexity refers to the amount of attention or processing a task requires as well as the amount of structure and clarity the task provides. Thus, task complexity increases, as the level of structure decreases, and as the required amount of task processing increases (Wood 1986, Campbell 1988). Complex tasks are, by their nature, difficult. However, a difficult task must not be complex but can still require high effort (Campbell 1988).

effort to a difficult task decreases and, *ceteris paribus*, he shifts effort away from difficult tasks to easier tasks. Under a fixed-wage compensation system, task difficulty does not influence the overall link between effort and rewards, that is, the effort-outcome expectancy, and becomes irrelevant for an individual's effort allocation choice.

2.5 Effects of Performance Measure Characteristics and Task Difficulty on Effort Allocation for a Fixed-Wage Compensation System

From both, an agency theory perspective and an expectancy theory perspective, under a fixed-wage compensation system neither performance measure characteristics, i.e. sensitivity and precision, nor task difficulty influence an individual's effort allocation choice. As stated in agency theory, economic agents do not react to performance measure characteristics and task difficulty, as they do not influence an agent's compensation under a fixed-wage. According to expectancy theory, under a fixed-wage, performance measure characteristics and task difficulty do not influence the overall link between effort and a fixed-wage as an outcome and they become irrelevant for an individual's effort allocation choice. Therefore, the following hypotheses can be stated:

***H1:** Under a fixed-wage compensation system, sensitivity has no effect on effort allocation (ATP, BTP).*

***H2:** Under a fixed-wage compensation system, precision has no effect on effort allocation (ATP, BTP).*

Experimental results from Brüggén (2011) support the second hypothesis. He employs a fixed-wage in a multi-task environment with two tasks being measured with varying precision. Participants do not significantly deviate from the principal's preferred effort allocation. However, task difficulty plays no role in his experiment, as he employs the same task twice.

2.6 Effects of Performance Measure Characteristics and Task Difficulty on Effort Allocation for a Performance-based Compensation System

Under a performance-based compensation system, both, performance measure characteristics and task difficulty, can influence effort allocation. If sensitivity on a task increases, from an agency theory perspective more effort is allocated to that task, because of cost-benefit considerations. From an expectancy theory perspective more effort is allocated to a task measured with higher sensitivity, as an increase in sensitivity in-

creases the effort-performance expectancy and hence an individual's motivation to exert effort on that task.

Task difficulty attenuates the effect of sensitivity on effort allocation. From an agency theory perspective, difficult tasks are less sensitive to effort and have higher marginal costs of effort than easy tasks. Hence, comparing increased benefits when sensitivity on a task increases, which are approximately equal under easy and difficult tasks, agents will allocate relatively less effort to a difficult task than to an easy task, due to difficult tasks' inherent lower relationship between effort and performance and their higher marginal costs. From an expectancy theory perspective, effort-performance expectancy is inherently lower for difficult tasks than for easy tasks. Thus, an increase in sensitivity increases the effort-performance expectancy relatively lower for difficult tasks than for easy tasks and less effort is allocated to difficult tasks than to easy tasks. The following hypothesis can be stated:

H3: *Under a performance-based compensation system, the effect of an increase in sensitivity on effort allocation will be less positive, when the task is difficult than when the task is easy (ATP, BTP).*

With respect to precision, agency theory and expectancy theory offer different predictions for a performance-based compensation system. From an agency theory perspective, an additive noise term is independent from an agent's effort. Hence, an agent cannot influence the increase in risk when performance measure precision decreases and his effort allocation is unaffected by precision. From an expectancy theory perspective, less effort is allocated to a task measured with lower precision, as a decrease in precision decreases the effort-performance expectancy and hence an individual's motivation to exert effort on that task. Furthermore, effort-performance expectancy and thus an individual's motivation is inherently lower for difficult tasks than for easy tasks. Consequently, a decrease in precision decreases the effort-performance expectancy relatively higher for easy tasks than for difficult tasks and hence more effort is allocated to difficult tasks than to easy tasks. Thus, two conflicting hypotheses can be stated:

H4a: *Under a performance-based compensation system, precision has no effect on effort allocation (ATP).*

H4b: *Under a performance-based compensation system, the effect of a decrease in precision on effort allocation will be less negative, when the task is difficult than when the task is easy (BTP).*

Experimental evidence for the effect of precision on effort allocation is indicated in Brügger (2005) for two tasks of same difficulty. Although not being explicitly tested, the descriptive statistics from Brügger (2005) show a highly distorted effort allocation for a performance-based compensation system, with more effort being allocated to the task measured more precisely.

3 Method and Design

3.1 Experimental Design

This study utilizes a $3 \times 3 \times 2$ (within-subjects) $\times 2$ (between-subjects) mixed experimental design to analyze the effects of performance measure characteristics on effort allocation. Precision, sensitivity, and task difficulty are manipulated within-subjects. Compensation system is manipulated between-subjects.

Within-subjects designs enhance statistical power by allowing control of between-subjects differences, which has the advantage of using fewer subjects, and using subjects' time very efficiently (Libby et al. 2002). According to Libby et al. (2002) a within subject designs is '(...) most effective when increased salience of manipulated variables is desirable from the standpoint of the experiment's goals (...)' (Libby et al. 2002, p. 804).

In this study increased salience of the performance measure characteristics as independent variables is desirable, as the goal of the study is to analyze subjects' reactions to changes in sensitivity and precision. From an expectancy theory perspective, Van Eerde and Thierry (1996) suggest a within-subjects manipulation of variations in effort-performance expectancy in order to capture Vroom's (1964) notion of them as forces which act relative to other forces within the individual. In line with this, Kahneman and Tversky (1996) state that a within-subjects design puts a focus on the independent variables of interest, and gives subjects the possibility to identify and adjust errors and inconsistencies in their decisions.

As Libby et al. (2002) mention, a within-subjects design also requires balancing the order of treatments, to ensure that treatment effects are not confounded with order effects. Therefore, in this study the treatments with respect to sensitivity and precision are counterbalanced using a Williams design (Williams 1949).

3.2 Multi-Task Environment, Task Difficulty, Performance Measure Characteristics and Compensation System

This study employs a sequential real-effort multi-task environment with an easy task and a difficult task. The two tasks fall into the category of simple production tasks as specified by Bonner et al. (2000). Therefore, they are equivalent with respect to complexity, but vary in difficulty. As in other studies (e.g. Chow 1983, Waller and Chow 1988, Shields and Waller 1988, Dillard and Fisher 1990, Bailey et al. 1998), the production tasks are used to simulate a firm setting with an employer and employee, respective principal and agent. The objective of the participants is to achieve a high performance on both tasks with respect to an even (1:1) performance allocation.

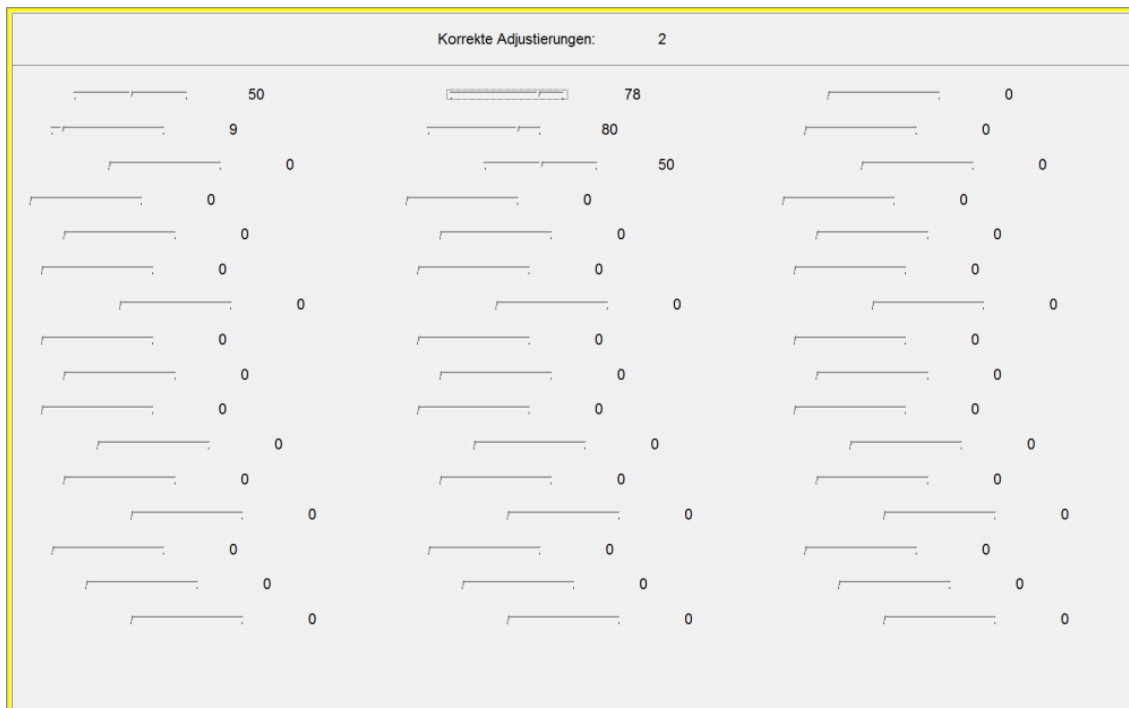
Both tasks are tied to separate performance measures, which are independent from remuneration for the fixed-wage compensation system, and connected to compensation for the performance-based compensation system. The tasks are performed sequentially to allow for a precise separation of effort allocation between tasks, and to avoid the possibility that subjects aggregate the noise of the different performance measures mentally (Sloof and van Praag 2010). Moreover, it would be difficult to achieve true ‘simultaneity’ among the tasks, as presenting two tasks simultaneously does not ensure that they will be performed simultaneously (Hecht et al. 2014).

Figure 1: Screenshot of the easy task (decoding task)

Korrekte Dekodierungen: 0											
Ea	<input type="text"/>	OK	Da	<input type="text"/>	OK	Ef	<input type="text"/>	OK	Bc	<input type="text"/>	OK
Fb	<input type="text"/>	OK	Hc	<input type="text"/>	OK	Ae	<input type="text"/>	OK	Bf	<input type="text"/>	OK
Ac	<input type="text"/>	OK	Ha	<input type="text"/>	OK	Ga	<input type="text"/>	OK	Fc	<input type="text"/>	OK
Dd	<input type="text"/>	OK	Ed	<input type="text"/>	OK	Dc	<input type="text"/>	OK	Ce	<input type="text"/>	OK
Cc	<input type="text"/>	OK	Cd	<input type="text"/>	OK	Ec	<input type="text"/>	OK	Ae	<input type="text"/>	OK
Gb	<input type="text"/>	OK	Gd	<input type="text"/>	OK	Ac	<input type="text"/>	OK	Hc	<input type="text"/>	OK
Dd	<input type="text"/>	OK	Fa	<input type="text"/>	OK	Hf	<input type="text"/>	OK	Eb	<input type="text"/>	OK
Hd	<input type="text"/>	OK	Eb	<input type="text"/>	OK	Ba	<input type="text"/>	OK	Ff	<input type="text"/>	OK
Ce	<input type="text"/>	OK	Gc	<input type="text"/>	OK	Ae	<input type="text"/>	OK	Bb	<input type="text"/>	OK
Fe	<input type="text"/>	OK	Hf	<input type="text"/>	OK	Be	<input type="text"/>	OK	Ba	<input type="text"/>	OK
De	<input type="text"/>	OK	Ga	<input type="text"/>	OK	Ae	<input type="text"/>	OK	Ac	<input type="text"/>	OK
Dd	<input type="text"/>	OK	Ac	<input type="text"/>	OK	Ca	<input type="text"/>	OK	Hd	<input type="text"/>	OK
Cd	<input type="text"/>	OK	Fd	<input type="text"/>	OK	Ce	<input type="text"/>	OK	Db	<input type="text"/>	OK
Cc	<input type="text"/>	OK	Bc	<input type="text"/>	OK	Ba	<input type="text"/>	OK	Ad	<input type="text"/>	OK
Gc	<input type="text"/>	OK	Bf	<input type="text"/>	OK	Be	<input type="text"/>	OK	Ac	<input type="text"/>	OK
Fd	<input type="text"/>	OK	Ad	<input type="text"/>	OK	He	<input type="text"/>	OK	Be	<input type="text"/>	OK

The easy task is a decoding task, which is similar to studies by Chow (1983), Shields and Waller (1988), Chow et al. (1988), Sillamaa (1999), Fisher et al. (2002), Stevens (2002), and Brügger (2011). The decoding task on a single screen is handled by decoding a combination of two letters into a three-digit number. The objective is to perform as many decodings as possible (with respect to an even performance allocation as specified by the employer). Figure 1 shows the decoding task.

Figure 2: Screenshot of the difficult task (slider task)



The difficult task is a slider task displaying a number of sliders on a single screen and is adopted from Gill and Prowse (2012). The number and position of the sliders on the screen do not vary across repetitions of the task or across experimental subjects. The objective is to perform as many slider adjustments as possible (with respect to an even performance allocation, too). Figure 2 shows the slider task.

Due to their simplicity, both tasks give a finely gradated measure of performance and involve little randomness. Thus, as in Brügger (2011) and in Gill and Prowse (2012), a single correct decoding, and a single correct slider adjustment are interpreted as a single unit of effort on the respective task. With respect to task difficulty, experience from two pre-tests has shown that compared to a single correct decoding, on average subjects need more attempts on the slider task to perform a single correct slider adjustment. Furthermore, a single unit of performance on the decoding task was on aver-

age quicker achieved than a single unit of performance on the slider task. The tasks were chosen because of their simplicity, as they do not require former knowledge of the participants, are easy to communicate, and simple to understand. They are identical across periods, and leave no room for speculating (Gill and Prowse 2012).

Performance on the two tasks is measured with two sales figures:

$$sales_{decoding} = correct_{decodings} \times price_{decoding} + noise_{decoding} \quad (1)$$

$$sales_{slider} = correct_{slider\ adjustments} \times price_{slider} + noise_{slider} \quad (2)$$

All monetary amounts are denoted in an experimental currency called Laboreuro, L€. One Laboreuro has a value of 0.01 Euro. Based on two pre-tests, the parameters for sensitivity and precision are determined.

In this study, because of the simplicity of the tasks, sensitivity is interpreted as the price for a single correct decoding, respective correct slider adjustment. Sensitivity is manipulated by altering the price for a single correct decoding and a single correct slider adjustment. The baseline setting is 1 L€ (low sensitivity easy task) for a single correct decoding, and 1 L€ (low sensitivity difficult task) for a single correct slider adjustment. The treatments are 2 L€ (high sensitivity easy task) for a single correct decoding, and 1 L€ (low sensitivity difficult task) for a single correct slider adjustment and vice versa. The three respective factor levels are *low sensitivity easy task / low sensitivity difficult task* (1 L€ / 1 L€) as a baseline, and *high sensitivity easy task / low sensitivity difficult task* (2 L€ / 1 L€) and *low sensitivity easy task / high sensitivity difficult task* (1 L€ / 2 L€) as the treatments.

Precision is utilized as the additive noise term of the respective performance measure. The noise term of the respective performance measure for each task is an individually generated random number for each participant in each period. Precision is manipulated by varying the standard deviation of the noise term of each performance measure. In line with most agency models, the noise term of each performance measure is normally distributed with a mean of zero. The standard deviation for performance measure noise is either 2 L€ (high precision) or 8 L€ (low precision). The three respective factor levels are *high precision easy task / high precision difficult task* (2 L€ / 2 L€) as a baseline setting and *low precision easy task / high precision difficult task* (8 L€ / 2 L€), and *high precision easy task / low precision difficult task* (2 L€ / 8 L€) as treatments. The

parameters for precision were calibrated to ensure that overall compensation literally always stays positive, whereas a performance measure itself can become negative.

The compensation system as between-subjects factor is either a fixed-wage compensation system, or a performance-based compensation system. In the fixed-wage compensation system treatment, the fixed-wage is 80 L€ for each round.⁷ In the performance-based compensation system treatment the following incentive scheme is used for each round:⁸

$$compensation_{performance} = 55 L€ + 0.5 \times sales_{decoding} + 0.5 \times sales_{slider} \quad (3)$$

The incentive weights (0.5 / 0.5) are equal for both tasks and for all rounds. Therefore, in each period a participant receives fifty percent of the sales measured for each task.

3.3 Dependent Variable

Although a single correct decoding, and a single correct slider adjustment are a sufficient proxy for a single unit of effort on the respective task, it is inappropriate for this study to utilize performance allocation between both tasks as a proxy for effort allocation as used by Brüggem (2011), because both tasks are not of the same difficulty, and therefore the translation of effort into performance differs between both tasks (Hannan et al. 2013).

Thus, the dependent variable for effort allocation is the amount of time allocated to the easy task relative to the difficult task in percentage terms (with a fixed time limit of two minutes):

$$effort_{allocation} = \frac{time_{decoding}}{time_{decoding} + time_{slider}} = \frac{time_{decoding}}{120_{seconds}} \% \quad (4)$$

Hannan et al. (2013) use a similar measure of effort allocation, and argue that this measure has the advantage of being a continuous measure. Furthermore, effort allocation as measured by the amount of time subjects allocate between tasks has the benefit that different effort-performance relationships on both tasks do not confound the measure.

⁷ Except for the first round.

⁸ Except for the first round.

3.4 Participants and Experimental Procedures

In total, 56 students from a public university in Germany participate in the study ($n = 504$ observations). The mean age of participants is 22.75 years, with the youngest participant being 18 and the oldest 28. There are 44 (78.6 percent) female participants and 12 (21.4 percent) male participants.⁹ As in other studies, which examine effort allocation in multi-task environments (Kachelmeier et al. 2008, Sloof and van Praag 2008, Hecht et al. 2012), the participants are undergraduates and graduate students.

Participants take part in a computer-based laboratory experiment.¹⁰ The laboratory consists of 20 cubicles, each with a computer, and an experimenter cubicle. Upon arrival the students enter a waiting room. An instructor explains the basic procedures. Next, the participants are randomly assigned to a cubicle, and start reading the instructions.¹¹ The instructions explain the task, and describe the procedures for determination of a participants' payoff. To ensure that participants perceive themselves as economic agents, the instructions label the roles of participants as employees. During the experiment, participants have the opportunity to refer to written materials for the instructions and the experimental task.

The experimental task consists of three subtasks: (1) time allocation, (2) easy task and (3) difficult task. The participants have to allocate the available time (120 seconds in total) between the two tasks. They can allocate the time without obstruction. They are told that an even (1:1) performance allocation is most desirable for the employer because it maximizes firm profit. Participants are told that their employer can never observe their performance directly, but only the performance measures at the end of each round.

Subjects have to answer a set of questions about the experimental task to ensure that they understand the task. They cannot continue without having given the correct answer to all questions. After a participant successfully completes this check, the first round begins.

Participants perform the experimental task consecutively for twelve rounds. The first three rounds are trial rounds.¹² The remaining nine rounds comprise the experi-

⁹ There is no effect of gender in this study ($p = .127$).

¹⁰ The experiment is programmed and conducted with the software z-Tree (Fischbacher 2007).

¹¹ The computer task starts automatically, after the participants enter the cubicles.

¹² In the second round participants receive information on precision. They are instructed that the performance measures for the decoding task and the slider task are both influenced by factors subjects cannot control, and can therefore, independently from their effort, vary within a certain bandwidth. They are told

mental treatments with respect to sensitivity and precision. In each round, a participant first receives information on precision and sensitivity for both tasks for the respective round, and allocates the available time between the tasks. Afterwards, participants perform the decoding task and slider task consecutively.¹³ Having finished the two tasks, participants receive information on the performance measures and their compensation for the respective round¹⁴, and continue with the next round.¹⁵ As noise terms in each round are newly generated random numbers, this design ensures that participants do not receive information on the actual noise terms prior to finishing a round, as the actual noise term for the easy task could influence the behavior for the difficult task.

After the last round, participants have to complete a postexperimental questionnaire. Finally, after completing the questionnaire, participants receive their payoff and leave. Under the fixed-wage compensation system participants earn 8.80 €, under the performance-based compensation system the average amount earned is 7.64 € with a minimum of 7.09 € and a maximum of 8.00 €. On average, one session takes 75 minutes.

4 Results

4.1 Manipulation Checks

Manipulation Check of General Task Understanding

Before the first and after the last trial round, participants are asked several questions regarding a general understanding of the experimental task and payoff. No participant can proceed to the first treatment unless all questions are answered correctly. If an answer is wrong, a help box for the respective question appears on the screen explaining the correct answer. All participants succeed to proceed during the experiment.

Manipulation Check of Task Difficulty

Analyses are performed to check, whether both objective and perceived difficulty of the easy task (decoding) varied from the difficult task (slider). As an objective measure of task difficulty, performance on both tasks is measured as correct decodings per mi-

the standard deviation for each performance measure. This is done with the 68-95-99.7 (three-sigma) rule. Participants can then simulate possible realizations of the noise terms with realizations of the performance measure shown separately.

¹³ In the first two rounds the amount of time is fixed to 60 seconds for each task to ensure that a participant cannot skip a task at will.

¹⁴ The first trial round is not compensated. For the remaining eleven rounds participants receive a payoff for each round.

¹⁵ In the first round both performance measures are measured precisely with an error term of zero in order to have an unbiased start for all participants.

nute for the easy task, and correct slider adjustments per minute for the difficult task. For all participants, correct decodings per minute (median = 12.00) are significantly higher than correct slider adjustments per minute (median = 8.88) (Wilcoxon signed-rank test; $z = -17.22$; $p < .001$, two-tailed). This indicates that objective difficulty for the decoding task is lower than for the slider task.

To measure the perceived difficulty of both tasks, participants are asked in the post-experimental questionnaire to rate their perceived own ability for the respective task (Brüggen 2011). A higher perceived own ability for a task relative to another task indicates a lower task difficulty compared to another task.

The mean of participants' responses to the two items that measure perceived own ability is 5.23 (standard deviation = 1.28) for the statement 'I am pretty skilled at the decoding task', and 5.20 (standard deviation = 1.33) for the statement 'I think I am pretty good at the decoding task', both statements on a seven-point Likert scale with 1 = strongly disagree and 7 = strongly agree. Answers to both items strongly correlate with each other ($\rho = .68$; $p < .001$; two-tailed), therefore the combined measure, i.e. the mean responses of participants to both statements, is used to compare perceived own ability between the easy and the difficult task. However, results are not different in quality with each of the responses to the two statements individually.

The mean of participants' responses to the two items that measure perceived own ability is 3.71 (standard deviation = 1.41) for the statement 'I am pretty skilled at the slider task', and 3.45 (standard deviation = 1.43) for the statement 'I think I am pretty good at the slider task', both statements on a seven-point Likert scale with 1 = strongly disagree and 7 = strongly agree. Again, answers to both items strongly correlate with each other ($\rho = .90$; $p < .001$; two-tailed) and concordantly the combined measure of the mean responses to both statements is used. However, results are not different in quality with each of the responses to the two statements individually.

The combined measure of perceived own ability for the easy task (mean = 5.21, median = 5.5, standard deviation = 1.22) is significantly higher, compared to the combined measure of perceived own ability for the difficult task (mean = 3.58, median = 3.50, standard deviation = 1.38) (Wilcoxon signed-rank test; $z = -5.64$; $p < .001$; two-tailed). Therefore, subjective difficulty for the decoding task is significantly lower compared to the slider task.

4.2 Hypothesis Tests

This study uses a mixed-design analysis of variance (ANOVA, Panel B in Table 1) on *effort_{allocation}*, where *compensation system* is a between-subjects factor that equals 1 for *performance-based compensation system*, and 0 for *fixed-wage compensation system*. *Sensitivity*, and *precision* are manipulated within-subjects, each on three factor levels as specified earlier. Descriptive statistics (Panel A), ANOVA (Panel B) and simple contrasts (Panel C) are shown in Table 1.

There is a significant main effect of *sensitivity* on *effort_{allocation}* ($F = 14.25$; $p < .001$; Greenhouse-Geisser; Panel B in Table 1), and a significant interaction effect between *sensitivity* and *compensation system* ($F = 6.06$; $p = .011$; Greenhouse-Geisser). This indicates that the effect of *sensitivity* on *effort_{allocation}* differs between the *fixed-wage compensation system*, and the *performance-based compensation system*. To break down this interaction, contrast coding (Buckless and Ravenscroft 1990) is performed comparing the *high sensitivity easy task / low sensitivity difficult task* level, and the *low sensitivity easy task / high sensitivity difficult task* level to the *low sensitivity easy task / low sensitivity difficult task* level for the *fixed-wage compensation system*, and *performance-based compensation system*. Panel A in Figure 3 shows the interaction effect and the respective contrasts.

There is a significant main effect of *precision* on *effort_{allocation}* ($F = 5.48$; $p = .007$; Greenhouse-Geisser; Panel B in Table 1), and a significant interaction effect between *precision* and *compensation system* ($F = 3.86$; $p = .027$; Greenhouse-Geisser). This indicates that the effect of *precision* on *effort_{allocation}* differs between the *fixed-wage* and the *performance-based compensation system*.

To break down this interaction, again contrast coding is performed comparing the *low precision easy task / high precision difficult task* level and the *high precision easy task / low precision difficult task* level to the *high precision easy task / high precision difficult task* level for the *fixed-wage compensation system* and *performance-based compensation system*. Panel B in Figure 3 shows the interaction effect and the respective contrasts.

Table 1: Effects of Precision, Sensitivity, Task Difficulty and Compensation System on Effort Allocation

Panel A: Descriptive Statistics for <i>effort</i>_{allocation}^a				
<i>Fixed-wage compensation system</i>	High Sensitivity Easy / Low Sensitivity Difficult	Low Sensitivity Easy / Low Sensitivity Difficult	Low Sensitivity Easy / High Sensitivity Difficult	Marginal Means
Low Precision Easy / High Precision Difficult	48.60 % [10.48 %] (n = 28)	48.75 % [12.54 %] (n = 28)	43.60 % [11.54 %] (n = 28)	46.98 % [11.66 %] (n = 84)
High Precision Easy / High Precision Difficult	50.18 % [15.68 %] (n = 28)	47.23 % [7.00 %] (n = 28)	48.30 % [9.37 %] (n = 28)	48.57 % [11.22 %] (n = 84)
High Precision Easy / Low Precision Difficult	47.65 % [13.27 %] (n = 28)	46.07 % [12.11 %] (n = 28)	45.30 % [8.40 %] (n = 28)	46.34 % [11.35 %] (n = 84)
Marginal Means	48.81 % [13.20 %] (n = 84)	47.35 % [10.77 %] (n = 84)	45.73 % [9.93 %] (n = 84)	47.30 % [11.41 %] (n = 252)
<i>Performance-based compensation system</i>	High Sensitivity Easy / Low Sensitivity Difficult	Low Sensitivity Easy / Low Sensitivity Difficult	Low Sensitivity Easy / High Sensitivity Difficult	Marginal Means
Low Precision Easy / High Precision Difficult	55.92 % [20.22 %] (n = 28)	49.52 % [13.42 %] (n = 28)	42.65 % [6.84 %] (n = 28)	49.37 % [15.38 %] (n = 84)
High Precision Easy / High Precision Difficult	62.62 % [21.41 %] (n = 28)	54.61 % [15.09 %] (n = 28)	45.09 % [16.91 %] (n = 28)	54.11 % [19.19 %] (n = 84)
High Precision Easy / Low Precision Difficult	60.12 % [15.40 %] (n = 28)	54.94 % [15.81 %] (n = 28)	47.17 % [16.37 %] (n = 28)	54.08 % [16.56 %] (n = 84)
Marginal Means	59.55 % [19.16 %] (n = 84)	53.03 % [14.84 %] (n = 84)	44.96 % [14.10 %] (n = 84)	52.52 % [17.19 %] (n = 252)

Panel B: Analysis of Variance for <i>effort</i>_{allocation}^b					
Within-Subjects Factor	df	SS	MS	F-Value	p-Value
Precision	1.848	.086	.047	5.475	.007***
Precision × Compensation System	1.848	.061	.033	3.859	.027**
Error Term (Precision)	99.807	.852	.009		
Sensitivity	1.269	.657	.518	14.248	<.001***
Sensitivity × Compensation System	1.269	.279	.137	6.061	.011**
Error Term (Sensitivity)	68.532	2.489	.220		
Precision × Sensitivity	3.322	.014	.036	.283	.702
Precision × Sensitivity × Compensation System	3.322	.034	.004	.953	.318
Error Term (Precision × Sensitivity)	179.403	1.543	.010		
Between-Subjects Factor	df	SS	MS	F-Value	p-Value
Constant Term	1	125.534	125.534	1,451.870	<.001***
Compensation System	1	.343	.343	3.968	.051*
Error Term	54	4.669	.086		

Panel C: Simple Contrasts for <i>effort</i>_{allocation}^c						
Compensation System	Sensitivity	df	SS	MS	F-Value	p-Value
(1) Fixed-wage	Easy Low / Diff Low vs. Easy High / Diff Low	1	.006	.006	1.084	.302
(2) Fixed-wage	Easy Low / Diff Low vs. Easy Low / Diff High	1	.014	.014	.490	.487
(3) Performance-based	Easy Low / Diff Low vs. Easy High / Diff Low	1	.063	.063	21.725	<.001*** ^d
(4) Performance-based	Easy Low / Diff Low vs. Easy Low / Diff High	1	.046	.046	12.168	<.001*** ^d
Compensation System	Precision	df	SS	MS	F-Value	p-Value
(5) Fixed-wage	Easy High / Diff High vs. Easy Low / Diff High	1	.007	.007	1.702	.198
(6) Fixed-wage	Easy High / Diff High vs. Easy High / Diff Low	1	.014	.014	2.823	.099*
(7) Performance-based	Easy High / Diff High vs. Easy Low / Diff High	1	.063	.063	15.189	<.001*** ^d
(8) Performance-based	Easy High / Diff High vs. Easy High / Diff Low	1	.000	.000	.001	.982

Table 1: Effects of Precision, Sensitivity, Task Difficulty and Compensation System on Effort Allocation cont.

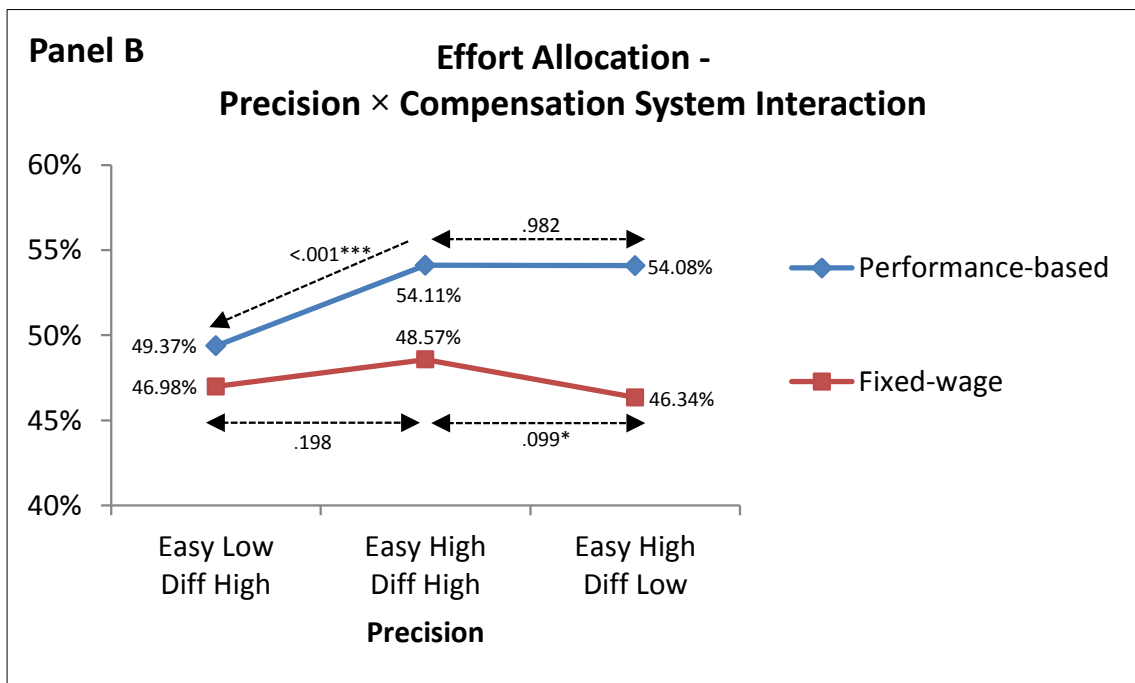
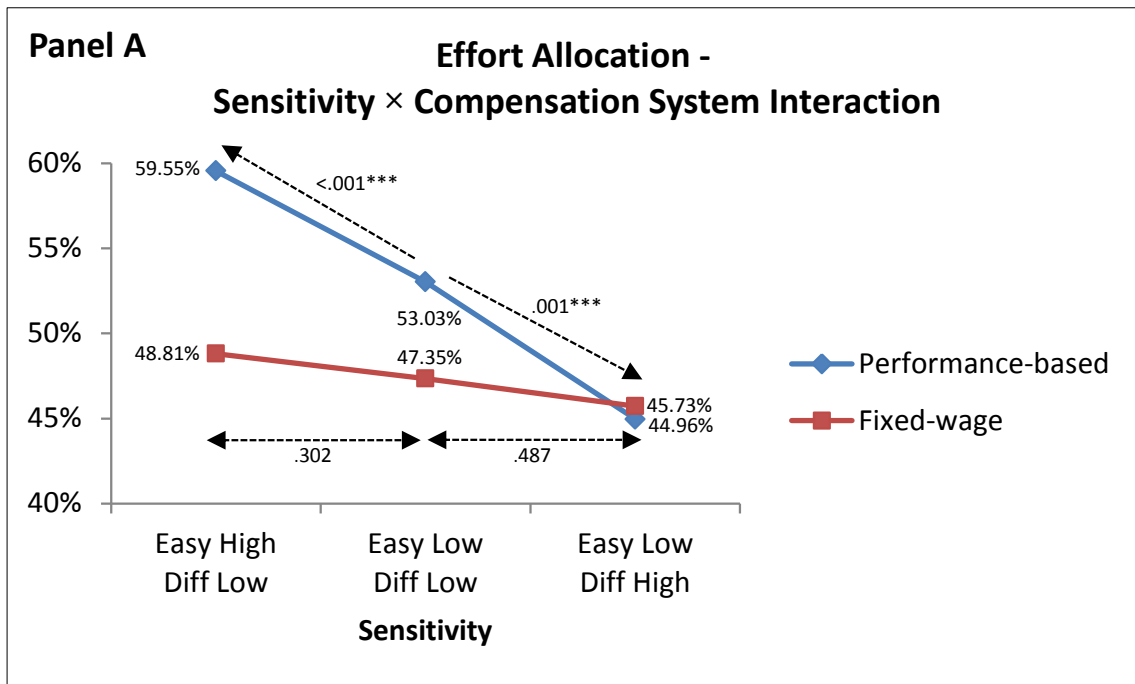
* Significance levels of 10 % (two-tailed).
** Significance levels of 5 % (two-tailed).
*** Significance levels of 1 % (two-tailed).
* ^d Significance levels of 10 % (one-tailed).
** ^d Significance levels of 1 % (one-tailed).
^a Means of <i>effortallocation</i> (respectively, [standard deviation] and (number of observations)) are shown per cell. Furthermore, the marginal means of performance measure precision and performance measure sensitivity are shown, as tests focus on this distinction. <i>effortallocation</i> used the following formula: $\text{time}_{\text{easy}} / (\text{time}_{\text{easy}} + \text{time}_{\text{difficult}}) = \text{time}_{\text{easy}} / (120 \text{ seconds})$ with $\text{time}_{\text{easy}}$ as the seconds allocated to the easy task and $\text{time}_{\text{difficult}}$ as the seconds allocated to the difficult task; time limit was fixed to 120 seconds. 50 % represents an even effort allocation and higher (lower) percentages indicate more effort allocated to the easy task (difficult task).
^b Analysis of variance for <i>effortallocation</i> with performance measure precision and performance measure sensitivity as within-subjects factors and compensation system as between-subjects factor. Degrees of freedom and p-values with Greenhouse-Geisser correction.
^c Contrast analyses to break down the precision \times compensation system and sensitivity \times compensation system interactions. The respective contrast codes are: (1) {1 1 0; -1/3 1/3 0 -1/3 1/3 0 -1/3 1/3 0}, (2) {1 1 0; 0 1/3 -1/3 0 1/3 -1/3 0 1/3 -1/3}, (3) {1 0 1; -1/3 1/3 0 -1/3 1/3 0 -1/3 1/3 0}, (4) {1 0 1; 0 1/3 -1/3 0 1/3 -1/3 0 1/3 -1/3}, (5) {1 1 0; -1/3 -1/3 -1/3 1/3 1/3 1/3 0 0 0}, (6) {1 1 0; 0 0 0 1/3 1/3 1/3 -1/3 -1/3 -1/3}, (7) {1 0 1; -1/3 -1/3 -1/3 1/3 1/3 1/3 0 0 0}, (8) {1 0 1; 0 0 0 1/3 1/3 1/3 -1/3 -1/3 -1/3}.

Effects of Sensitivity and Task Difficulty on Effort Allocation given a Fixed-wage Compensation System

In line with H1, contrast codes are specified as {1 1 0; -1/3 1/3 0 -1/3 1/3 0 -1/3 1/3 0} for the marginal means *fixed-wage compensation system / low sensitivity easy task / low sensitivity difficult task*, and *fixed-wage compensation system / high sensitivity easy task / low sensitivity difficult task* as well as {1 1 0; 0 1/3 -1/3 0 1/3 -1/3 0 1/3 -1/3} for the marginal means *fixed-wage / low sensitivity easy task / low sensitivity difficult task*, and *fixed-wage / low sensitivity easy task / high sensitivity difficult task*, respectively.

The result of contrast analysis for the *fixed-wage compensation system* comparing *effortallocation* for the *low sensitivity easy task / low sensitivity difficult task* level (mean effort allocation = 47.35 %) to the *high sensitivity easy task / low sensitivity difficult task* level (48.81 %), there is no significant difference (F = 1.08; p = .302; two-tailed; Panel C in Table 1). Furthermore, for the *fixed-wage compensation system* comparing *effortallocation* for the *low sensitivity easy task / low sensitivity difficult task* level (47.35 %) to the *low sensitivity easy task / high sensitivity difficult task* level (45.73 %), there is no significant difference (F = .49; p = .487; two-tailed).

Figure 3: Effort Allocation – Sensitivity × Compensation System Interaction and Precision × Compensation System Interaction^a



^a Variable definitions in Table 1. Detailed analyses of the sensitivity × compensation system interaction (Panel A) and the precision × compensation system interaction (Panel B). The horizontal axes represent the performance measure sensitivity on the easy task, respective difficult task (Panel A) and the performance measure precision on the easy task, respective difficult task (Panel B). The full lines represent the compensation system (fixed-wage vs. performance-based). The vertical axes display the mean effort allocation based on marginal means. 50 % represents an even effort allocation and higher (lower) percentages indicate more effort allocated to the easy task (difficult task). One-sided dotted arrows and associated figures present the direction and the p-value of a one-sided mean-by-mean comparison. Two-sided dotted arrows and associated figures present the p-value of a two-sided mean-by-mean comparison. ***, * indicate significance levels of 1 %, or 10 %.

The results indicate that given the fixed-wage compensation system there is no significant effect of sensitivity on *effort_{allocation}*. H1 is therefore supported.

Effects of Precision and Task Difficulty on Effort Allocation given a Fixed-wage Compensation System

In line with H2, contrast codes are specified as {1 1 0; -1/3 -1/3 -1/3 1/3 1/3 1/3 0 0} for the marginal means *fixed-wage compensation system / high precision easy task / high precision difficult task*, and *fixed-wage compensation system / low precision easy task / high precision difficult task* as well as {1 1 0; 0 0 0 1/3 1/3 1/3 -1/3 -1/3 -1/3}, for the marginal means *fixed-wage compensation system / high precision easy task / high precision difficult task*, and *fixed-wage compensation system / high precision easy task / low precision difficult task*, respectively.

The result of contrast analysis for the *fixed-wage compensation system* comparing *effort_{allocation}* for the *high precision easy task / high precision difficult task* level (mean effort allocation = 48.57 %; Panel A in Table 1) to the *low precision easy task / high precision difficult task* level (46.98 %) shows that there is no significant difference (F = 1.70; p = .198; two-tailed; Panel C in Table 1). Also, for the *fixed-wage compensation system* comparing *effort_{allocation}* for the *high precision easy task / high precision difficult task* level (48.57 %) to the *high precision easy task / low precision difficult task* (46.34 %) there is a marginally significant difference (F = 2.82; p = .099; two-tailed).

The results indicate that given the *fixed-wage compensation system*, there is a marginally significant effect of precision on effort allocation for the difficult task. H2 is therefore not supported.

Effects of Sensitivity and Task Difficulty on Effort Allocation given a Performance-based Compensation System

In line with H3, contrast codes are specified as {1 0 1; -1/3 1/3 0 -1/3 1/3 0 -1/3 1/3 0} for the marginal means *performance-based compensation system / low sensitivity easy task / low sensitivity difficult task* and *performance-based compensation system / high sensitivity easy task / low sensitivity difficult task* as well as {1 0 1; 0 1/3 -1/3 0 1/3 -1/3 0 1/3 -1/3} for the marginal means *performance-based compensation system / low sensitivity easy task / low sensitivity difficult task*, and *performance-based compensation system / low sensitivity easy task / high sensitivity difficult task*, respectively.

The result of contrast analysis for the *performance-based compensation system* comparing *effort_{allocation}* for the *low sensitivity easy task / low sensitivity difficult task* level (mean effort allocation = 53.03 %; Panel A in Table 1) to the *high sensitivity easy task / low sensitivity difficult task* level (59.55 %) shows that significantly more effort is allocated to the easy task, when sensitivity on the easy task increases ($F = 21.73$; $p < .001$; one tailed; Panel C in Table 1). For the *performance-based compensation system* comparing *effort_{allocation}* for the *low sensitivity easy task / low sensitivity difficult task* level (53.03 %) to the *low sensitivity easy task / high sensitivity difficult task* level (44.96 %), significantly more effort is allocated to the difficult task, when sensitivity on the difficult task increases ($F = 12.17$; $p < .001$; one tailed).

The results indicate that given the *performance-based compensation system* an increase in *sensitivity* for a task leads to significantly more effort allocated to the respective task.

Finally, an (untabulated) repeated measures ANOVA with *change in effort_{allocation}* as the dependent variable is conducted.¹⁶ Given the *performance-based compensation system* the *change in effort_{allocation}* from the *low sensitivity easy task / low sensitivity difficult task* level to the *high sensitivity easy task / low sensitivity difficult task* level (mean difference = 6.53 %) is smaller than *the change in effort_{allocation}* from the *low sensitivity easy task / low sensitivity difficult task* level to the *low sensitivity easy task / high sensitivity difficult task* level (mean difference = 8.06 %). The *change in effort_{allocation}* is not significantly larger for a change in sensitivity for the easy task ($F = .45$; $p = .254$; one-tailed). This result contradicts H3.

Effects of Precision and Task Difficulty on Effort Allocation given a Performance-based Compensation System

In line with H4a and H4b, contrast codes are specified as {1 0 1; -1/3 -1/3 -1/3 1/3 1/3 1/3 0 0 0} for the marginal means *performance-based compensation system / high precision easy task / high precision difficult task*, and *performance-based compensation system / low precision easy task / high precision difficult task* as well as {1 0 1; 0 0 0 1/3

¹⁶ *Change in effort_{allocation}* is the difference between *effort_{allocation}* for the *low sensitivity easy task / low sensitivity difficult task* level and the *high sensitivity easy task / low sensitivity difficult task* level, as well as the difference between the *low sensitivity easy task / low sensitivity difficult task* level and the *low sensitivity easy task / high sensitivity difficult task* level. One within-subjects variable is included with two factor levels (1) change in *effort_{allocation}* from the *low sensitivity easy task / low sensitivity difficult task* level to the *high sensitivity easy task / low sensitivity difficult task* level and (2) change in *effort_{allocation}* from the *low sensitivity easy task / low sensitivity difficult task* level to the *low sensitivity easy task / high sensitivity difficult task* level.

1/3 1/3 -1/3 -1/3 -1/3}, for the marginal means *performance-based compensation system / high precision easy task / high precision difficult task*, and *performance-based compensation system / high precision easy task / low precision difficult task*, respectively.

The result of contrast analysis for the *performance-based compensation system* comparing *effort_{allocation}* for the *high precision easy task / high precision difficult task* level (mean effort allocation = 54.11 %; Panel A in Table 1) to the *low precision easy task / high precision difficult task* level (49.37 %,) significantly more effort is allocated to the difficult task, when precision on the easy task decreases ($F = 15.19$; $p < .001$; one tailed; Panel C in Table 1). For the *performance-based compensation system* comparing *effort_{allocation}* for the *high precision easy task / high precision difficult task* level (54.11 %) to the *high precision easy task / low precision difficult task* level (54.08 %) there is no significant difference ($F = .00$; $p = .982$; two-tailed).

The results indicate that under the *performance-based compensation system* a decrease in *precision* for the easy task leads to significantly more effort allocated to the difficult task, whereas for a decrease in precision for the difficult task, there is no significant effect of *precision* on *effort_{allocation}*. These results support the expectancy theory prediction. H4b is supported and H4a is not supported.

4.3. Further Analysis for Performance on the Easy Task and Performance on the Difficult Task

To further analyze how a change in *effort_{allocation}* affects *performance* on the easy task and the difficult task, mean by mean comparisons (Panel A and B in Table 3) are performed using simple contrasts based on a (untabulated) generalized linear model (IBM Corporation 2012) with *performance*, that is, *performance_{easy}* (*performance_{diff}*), i.e. the number of correct decodings (slider adjustments) performed in the decode task (slider task) in one period, as the dependent variable. An identity link-function, a robust covariance matrix estimator, and Satterthwaite (1946) degrees of freedom are specified. *Compensation system* (between-subjects), *task difficulty* (within-subjects), *sensitivity* (within-subjects) and *precision* (within-subjects) are specified as independent variables. Descriptive statistics are shown in Table 2 and mean by mean comparisons are shown in Table 3 and Figure 4.

Table 2: Descriptive Statistics for Performance on the Easy Task and Performance on the Difficult Task^a

<i>Fixed-wage compensation system</i>	<u>High Sensitivity Easy / Low Sensitivity Difficult</u>		<u>Low Sensitivity Easy / Low Sensitivity Difficult</u>		<u>Low Sensitivity Easy / High Sensitivity Difficult</u>		<u>Marginal Means</u>	
	<i>performance_{easy}</i>	<i>performance_{diff}</i>	<i>performance_{easy}</i>	<i>performance_{diff}</i>	<i>performance_{easy}</i>	<i>performance_{diff}</i>	<i>performance_{easy}</i>	<i>performance_{diff}</i>
<u>Low Precision Easy/ High Precision Difficult</u>	11.14 [2.61] (n = 28)	9.04 [2.50] (n = 28)	11.29 [3.89] (n = 28)	8.79 [2.73] (n = 28)	9.89 [3.20] (n = 28)	9.82 [2.23] (n = 28)	10.77 [3.29] (n = 84)	9.21 [2.50] (n = 84)
<u>High Precision Easy/ High Precision Difficult</u>	11.57 [3.81] (n = 28)	8.54 [3.36] (n = 28)	11.07 [2.36] (n = 28)	9.29 [2.09] (n = 28)	11.36 [2.70] (n = 28)	9.18 [2.00] (n = 28)	11.33 [2.99] (n = 84)	9.00 [2.55] (n = 84)
<u>High Precision Easy/ Low Precision Difficult</u>	10.54 [5.17] (n = 28)	8.57 [2.92] (n = 28)	10.00 [3.91] (n = 28)	9.14 [3.44] (n = 28)	10.50 [2.66] (n = 28)	9.89 [1.81] (n = 28)	10.35 [4.00] (n = 84)	9.20 [2.83] (n = 84)
<u>Marginal Means</u>	11.08 [3.98] (n = 84)	8.71 [2.92] (n = 84)	10.79 [3.46] (n = 84)	9.07 [2.78] (n = 84)	10.58 [2.89] (n = 84)	9.63 [2.02] (n = 84)	10.82 [3.47] (n = 252)	9.14 [2.62] (n = 252)
<i>Performance-based compensation system</i>	<u>High Sensitivity Easy / Low Sensitivity Difficult</u>		<u>Low Sensitivity Easy / Low Sensitivity Difficult</u>		<u>Low Sensitivity Easy / High Sensitivity Difficult</u>		<u>Marginal Means</u>	
	<i>performance_{easy}</i>	<i>performance_{diff}</i>	<i>performance_{easy}</i>	<i>performance_{diff}</i>	<i>performance_{easy}</i>	<i>performance_{diff}</i>	<i>performance_{easy}</i>	<i>performance_{diff}</i>
<u>Low Precision Easy / High Precision Difficult</u>	13.29 [6.05] (n = 28)	7.71 [3.79] (n = 28)	12.04 [4.65] (n = 28)	9.18 [3.14] (n = 28)	10.18 [2.74] (n = 28)	7.21 [5.45] (n = 28)	11.83 [4.80] (n = 84)	8.04 [4.27] (n = 84)
<u>High Precision Easy / High Precision Difficult</u>	15.21 [6.39] (n = 28)	4.29 [4.20] (n = 28)	12.29 [3.69] (n = 28)	7.39 [3.49] (n = 28)	10.32 [4.91] (n = 28)	9.50 [3.79] (n = 28)	12.61 [5.45] (n = 84)	7.06 [4.36] (n = 84)
<u>High Precision Easy / Low Precision Difficult</u>	14.36 [6.24] (n = 28)	7.11 [2.70] (n = 28)	13.29 [4.67] (n = 28)	8.07 [3.30] (n = 28)	11.00 [4.47] (n = 28)	9.57 [4.47] (n = 28)	12.88 [5.31] (n = 84)	8.25 [3.67] (n = 84)
<u>Marginal Means</u>	14.29 [6.20] (n = 84)	6.37 [3.88] (n = 84)	12.54 [4.34] (n = 84)	8.21 [3.36] (n = 84)	10.50 [4.11] (n = 84)	8.76 [4.69] (n = 84)	12.44 [5.19] (n = 252)	7.78 [4.13] (n = 252)

^a Means of *performance_{easy}* and *performance_{diff}* (respectively, [standard deviation] and (number of observations)) are shown per cell. Furthermore the marginal means of performance measure precision and performance measure sensitivity are shown, as tests focus on this distinction. *performance_{easy}* is the number of correct decodings performed in the decode task in one period. *performance_{diff}* is the number of correct slider adjustments performed in the slider task in one period.

Effects of Sensitivity and Task Difficulty on Performance given a Fixed-wage Compensation System

Given the *fixed-wage compensation system* comparing *performance_{easy}* for the *low sensitivity easy task / low sensitivity difficult task* level (mean task performance = 10.79; Table 2) to the *high sensitivity easy task / low sensitivity difficult task* level (11.08), there is no significant difference ($t = -.84$; $p = .401$; two-tailed; Panel A in Table 3). Also, for the *fixed-wage compensation system* comparing *performance_{easy}* for the *low sensitivity easy task / low sensitivity difficult task* level (mean task performance = 10.79; Table 2) to the *low sensitivity easy task / high sensitivity difficult task* level (10.58), there is no significant difference ($t = .31$; $p = .756$; two-tailed; Panel A in Table 3).

The results indicate that given the *fixed-wage compensation system* there is no significant effect of *sensitivity* on *performance_{easy}*.

Given the *fixed-wage compensation system* comparing *performance_{diff}* for the *low sensitivity easy task / low sensitivity difficult task* level (mean task performance = 9.07; Table 2) to the *high sensitivity easy task / low sensitivity difficult task* level (8.71), there is no significant difference ($t = 1.14$; $p = .253$; two-tailed; Panel B in Table 3). Also, for the *fixed-wage compensation system* comparing *performance_{diff}* for the *low sensitivity easy task / low sensitivity difficult task* level (9.07) to the *low sensitivity easy task / high sensitivity difficult task* level (9.63), there is no significant difference ($t = -1.38$; $p = .172$; two-tailed).

The results show that given the *fixed-wage compensation system* there is no significant effect of *sensitivity* on *performance_{diff}*.

Effects of Precision and Task Difficulty on Performance given a Fixed-wage Compensation System

Given the *fixed-wage compensation system* comparing *performance_{easy}* for the *high precision easy task / high precision difficult task* level (mean task performance = 11.33; Table 2) to the *low precision easy task / high precision difficult task* level (10.77) shows a marginally significant higher *performance*, when *precision* on the easy task is high ($t = 1.83$; $p = .069$; two-tailed; Panel A in Table 3). Furthermore, for the *fixed-wage compensation system* comparing *performance_{easy}* for the *high precision easy task / high precision difficult task* level (11.33) to the *high precision easy task / low precision difficult*

task (10.35) there is a significant decrease in performance, when *precision* on the difficult task is low ($t = 2.53$; $p = .012$; two-tailed).

The results indicate that given the *fixed-wage compensation system performance* on the easy task decreases significantly, when *precision* decreases on the easy task or the difficult task.

Given the *fixed-wage compensation system* comparing *performance_{diff}* for the *high precision easy task / high precision difficult task* level (mean task performance = 9.00; Table 2) to the *low precision easy task / high precision difficult task* level (9.21) shows no significant effect of *precision* on *performance* ($t = -.80$; $p = .425$; two-tailed; Panel B in Table 3). Also, for the *fixed-wage compensation system* comparing *performance_{diff}* for the *high precision easy task / high precision difficult task* level (9.00) to the *high precision easy task / low precision difficult task* (9.20) there is no effect on *performance*, when *precision* on the difficult task is low ($t = -.62$; $p = .536$; two-tailed).

The results show that given the *fixed-wage compensation system* there is no significant effect of *precision* on *performance_{diff}*.

Effects of Sensitivity and Task Difficulty on Performance given a Performance-based Compensation System

Given the *performance-based compensation system* comparing *performance_{easy}* for the *low sensitivity easy task / low sensitivity difficult task* level (mean task performance = 12.54; Table 2) to the *high sensitivity easy task / low sensitivity difficult task* level (14.29) shows that performance on the easy task is significantly higher, when sensitivity on the easy task increases ($t = -3.26$; $p = .003$; two-tailed; Panel A in Table 3) For the *performance-based compensation system* comparing *performance_{easy}* for the *low sensitivity easy task / low sensitivity difficult task* level (12.54) to the *low sensitivity easy task / high sensitivity difficult task* level (10.50), performance on the easy task decreases significantly, when sensitivity on the difficult task increases ($t = 3.39$; $p = .001$; two-tailed).

The results indicate that given the *performance-based compensation system* an increase in *sensitivity* on the easy task increases *performance* for the easy task, whereas an increase in *sensitivity* on the difficult task decreases *performance* on the easy task.

Table 3: Mean by Mean Comparisons for Task Performance^a

Panel A: Mean by Mean Comparisons for $performance_{easy}$^b					
Compensation System	Sensitivity	Treatment 1	Treatment 2	t-Value	p-Value
(1) Fixed-wage	Easy Low / Diff Low vs. Easy High / Diff Low	10.79	11.08	-.841	.401
(2) Fixed-wage	Easy Low / Diff Low vs. Easy Low / Diff High	10.79	10.58	.311	.756
(3) Performance-based	Easy Low / Diff Low vs. Easy High / Diff Low	12.54	14.29	-3.263	.003***
(4) Performance-based	Easy Low / Diff Low vs. Easy Low / Diff High	12.54	10.50	3.385	.001***
Compensation System	Precision	df	MS	t-Value	p-Value
(5) Fixed-wage	Easy High / Diff High vs. Easy Low / Diff High	11.33	10.77	1.829	.069*
(6) Fixed-wage	Easy High / Diff High vs. Easy High / Diff Low	11.33	10.35	2.532	.012**
(7) Performance-based	Easy High / Diff High vs. Easy Low / Diff High	12.61	11.83	2.004	.046**
(8) Performance-based	Easy High / Diff High vs. Easy High / Diff Low	12.61	12.88	-.679	.497
Panel B: Mean by Mean Comparisons for $performance_{diff}$^c					
Compensation System	Sensitivity	Treatment 1	Treatment 2	t-Value	p-Value
(1) Fixed-wage	Easy Low / Diff Low vs. Easy High / Diff Low	9.07	8.71	1.144	.253
(2) Fixed-wage	Easy Low / Diff Low vs. Easy Low / Diff High	9.07	9.63	-1.376	.172
(3) Performance-based	Easy Low / Diff Low vs. Easy High / Diff Low	8.21	6.37	4.781	<.001***
(4) Performance-based	Easy Low / Diff Low vs. Easy Low / Diff High	8.21	8.76	-.826	.411
Compensation System	Precision	Treatment 1	Treatment 2	t-Value	p-Value
(5) Fixed-wage	Easy High / Diff High vs. Easy Low / Diff High	9.00	9.21	-.798	.425
(6) Fixed-wage	Easy High / Diff High vs. Easy High / Diff Low	9.00	9.20	-.619	.536
(7) Performance-based	Easy High / Diff High vs. Easy Low / Diff High	7.06	8.04	-3.001	.003***
(8) Performance-based	Easy High / Diff High vs. Easy High / Diff Low	7.06	8.25	-3.772	<.001***

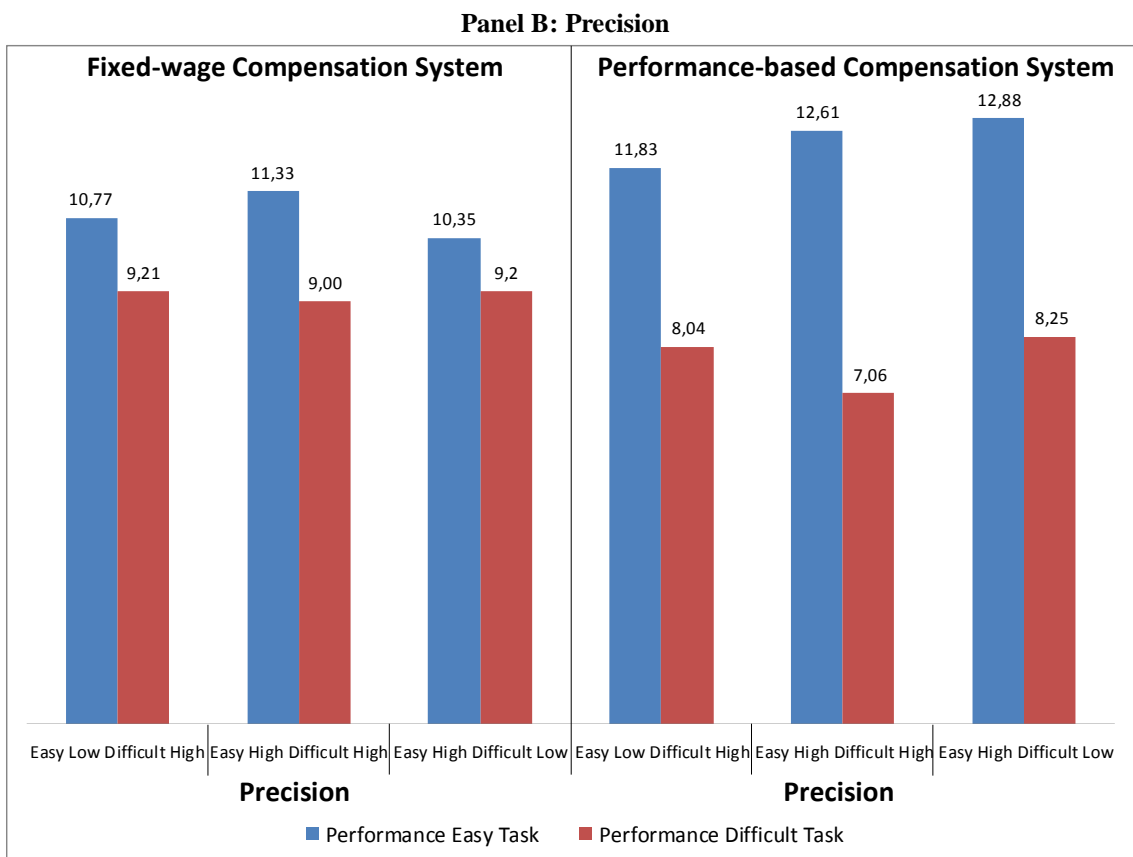
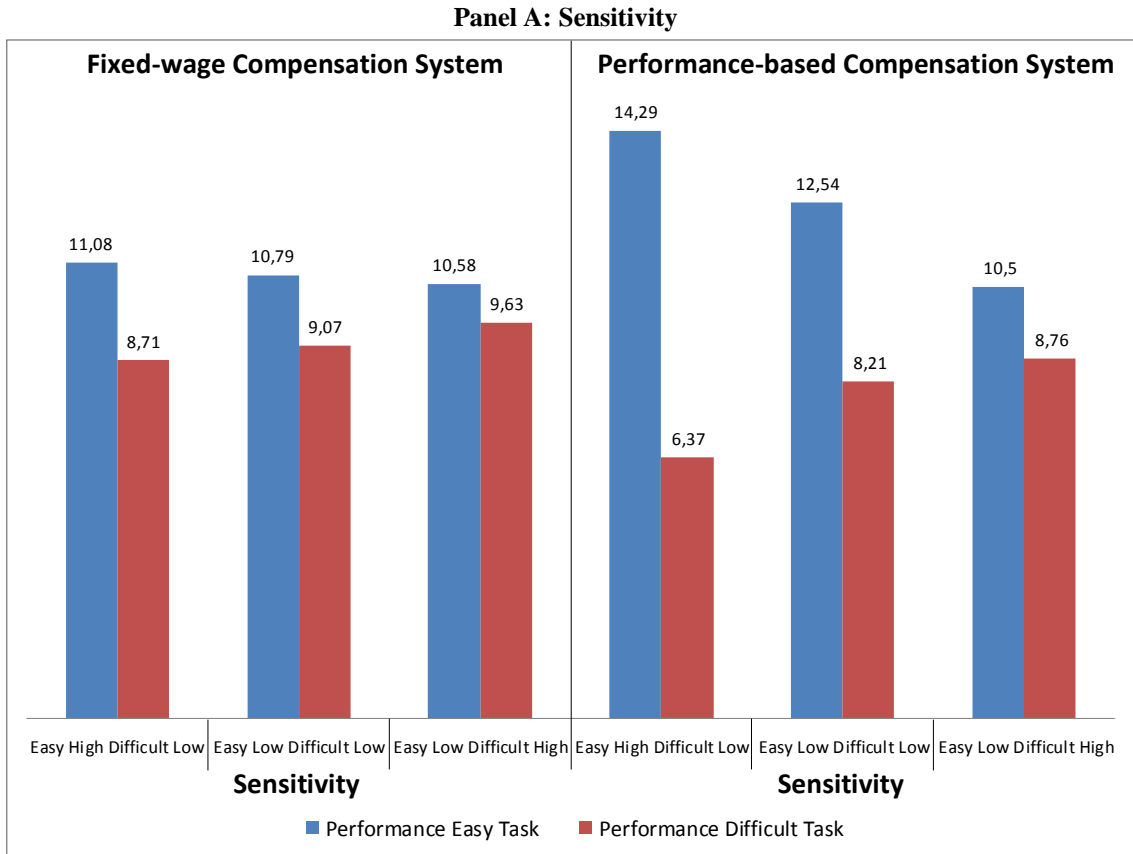
^a Mean by mean comparisons are performed using simple contrasts based on a generalized linear model (IBM Corporation 2012) with $performance_{easy}$ ($performance_{diff}$), i.e. the number of correct decodings (slider adjustments) performed in the decode task (slider task) in one period, as the dependent variable. An identity link-function, a robust covariance matrix estimator, and Satterthwaite (1946) degrees of freedom are specified. *Compensation System* (between-subjects), *task Difficulty* (within-subjects), *precision* (within-subjects) and *sensitivity* (within-subjects) are specified as independent variables.

^b $performance_{easy}$ is the number of correct decodings performed in the decode task in one period.

^c $performance_{diff}$ is the number of correct slider adjustments performed in the slider task in one period.

Given the *performance-based compensation system* comparing $performance_{diff}$ for the *low sensitivity easy task / low sensitivity difficult task* level (mean task performance = 8.21; Table 2) to the *high sensitivity easy task / low sensitivity difficult task* level (6.37) shows that performance on the difficult task is significantly lower, when sensitivity on the easy task increases ($t = 4.78$; $p < .001$; two-tailed; Panel B in Table 3).

Figure 4: Performance on the Easy Task and Performance on the Difficult Task



^a Variable definitions in Table 1 and 3. Columns and data values indicate performance for the easy task and performance for the difficult task for sensitivity (Panel A) and precision (Panel B). The blue column and the respective numerical value on the left side show *performance_{easy}*. The red column and the respective numerical value on the right side show *performance_{diff.}*

For the *performance-based compensation system* comparing *performance_{diff}* for the *low sensitivity easy task / low sensitivity difficult task* level (8.21) to the *low sensitivity easy task / high sensitivity difficult task* level (8.76), performance on the difficult task increases, when sensitivity on the difficult task increases but the effect is not significant ($t = -.83$; $p = .411$; two-tailed; Panel B in Table 3).

The results show that given the *performance-based compensation system* an increase in *sensitivity* on the difficult task increases *performance* for the difficult task but the effect is not significant, whereas an increase in *sensitivity* on the easy task significantly decreases *performance* on the difficult task.

Effects of Precision and Task Difficulty on Performance given a Performance-based Compensation System

Given the *performance-based compensation system* comparing *performance_{easy}* for the *high precision easy task / high precision difficult task* level (mean task performance = 12.61; Table 2) to the *low precision easy task / high precision difficult task* level (11.83) performance on the easy task decreases significantly, when *precision* on the easy task decreases ($t = 2.00$; $p = .046$; two-tailed; Panel A in Table 3). Also, for the *performance-based compensation system* comparing *performance_{easy}* for the *high precision easy task / high precision difficult task* level (12.61) to the *high precision easy task / low precision difficult task* level (12.88) there is no significant difference ($t = -.68$; $p = .497$; two-tailed).

The results indicate that given the *performance-based compensation system* a decrease in *precision* on the easy task significantly decreases *performance* for the easy task, whereas a decrease in *precision* on the difficult task has no significant effect on *performance* for the easy task.

Given the *performance-based compensation system* comparing *performance_{diff}* for the *high precision easy task / high precision difficult task* level (mean task performance = 7.06; Table 2) to the *low precision easy task / high precision difficult task* level (8.04) performance on the difficult task increases significantly, when *precision* on the easy task decreases ($t = -3.00$; $p = .003$; two-tailed; Panel B in Table 3). For the *performance-based compensation system* comparing *performance_{diff}* for the *high precision easy task / high precision difficult task* level (7.06) to the *high precision easy task / low*

precision difficult task level (8.25) performance on the difficult task increases significantly, when *precision* on the difficult task decreases ($t = -3.77$; $p < .001$; two-tailed).

The results indicate that given the *performance-based compensation system*, both a decrease in *precision* on the easy task and the difficult tasks significantly increases *performance* for the difficult task.

4.4 Supplemental Analysis

Additional analyses (Table 4) were performed to examine the intrinsic and extrinsic motivation of the participants as well as the subjectively perceived effects of precision and sensitivity for the fixed-wage and performance-based compensation system.

The answer categories to the post-experiment questionnaire are on a seven-point Likert scale with 1 = strongly disagree and 7 = strongly agree.

The strength of intrinsic motivation was analyzed with the postexperimental statement: 'I was motivated in every month to meet the expectations of my employer.' Results do not differ significantly (Mann-Whitney test; $z = -1.33$; $p = .183$; two-tailed; Table 4) as the responses for both, the *fixed-wage compensation system* (mean = 5.25; median = 6.00) and the *performance-based compensation system* (mean = 4.57; median = 5.00) are high on average. Therefore, on average all participants were intrinsically motivated.

The strength of extrinsic motivation was evaluated with a pair of postexperimental statements (1) 'My fixed-wage has motivated me in every month to provide high performance.' for the *fixed-wage compensation system* and (2) 'Variable pay in each month was an incentive for me to provide high performance.' for the *performance based compensation system*. Results for the questions differ significantly (Mann-Whitney test; $z = -4.92$; $p < .001$; two-tailed; Table 4) between the *fixed-wage compensation system* (mean = 3.71; median = 4) and *performance-based compensation system* (mean = 6.11; median = 6). The results indicate that in accordance with expectancy theory the valence of variable pay is higher than the valence of a fixed-wage.

Table 4: Responses to Postexperimental Questions

<u>Descriptive Statistics^a</u>		<u>Mann-Whitney Test^b</u>			
<u>Construct</u>	<u>Item</u>	<u>Fixed-wage compensation system</u>	<u>Performance-based compensation system</u>	<u>z-Value</u>	<u>p-Value</u>
Intrinsic Motivation	I was motivated in every month to meet the expectations of my employer.	5.25 {6.00} [1.92] (n = 28)	4.57 {5.00} [2.10] (n = 28)	-1.330	.183
Extrinsic Motivation	Fixed-wage compensation system: My fixed-wage has motivated me in every month to provide high performance. Performance-based compensation system: Variable pay in each month was an incentive for me to provide high performance.	3.71 {4.00} [1.84] (n = 28)	6.11 {6.00} [1.17] (n = 28)	4.919	<.001** *
Sensitivity	The price of a task has affected my behavior during the experiment.	3.86 {4.50} [2.05] (n = 28)	4.43 {6.00} [2.35] (n = 28)	1.407	.159
Precision	The non-controllable variations in the performance measures have influenced my behavior during the experiment.	3.14 {3.00} [1.82] (n = 28)	2.82 {2.50} [1.81] (n = 28)	-.712	.477

* Significance levels of 10 % (two-tailed).

** Significance levels of 5 % (two-tailed).

*** Significance levels of 1 % (two-tailed).

^a Means of the respective item (respectively, {median}, [standard deviation] and (number of observations)) are shown for the fixed-wage compensation system and performance-based compensation system. The answer categories to the postexperimental questionnaire are on a seven-point Likert scale with 1 = strongly disagree and 7 = strongly agree.

^b The p-values result from Mann-Whitney tests comparing the two types of compensation system.

With respect to performance measure *precision*, the perceived effect of *precision* was examined with the statement: ‘The non-controllable variations in the performance measures have influenced my behavior during the experiment’ Results do not differ significantly as the responses for the *fixed-wage compensation system* (mean = 3.14; median = 3.00; Table 4) are not significantly higher compared to the *performance-based compensation system* (mean = 2.82; median = 2.5) (Mann-Whitney test; $z = -.71$; $p = .477$; two-tailed). All participants feel moderately influenced by *precision*.

5 Discussion and Conclusion

The main purpose of this study is to analyze the moderating effect of task difficulty on the effects of performance measure sensitivity and performance measure precision on effort allocation in a multi-task environment.

First of all, results are discussed for the fixed-wage compensation system. As results from the postexperimental questionnaire indicate, participants receiving a fixed-wage were highly motivated to meet the expectations of the employer, i.e. provide high performance with respect to an even (1:1) performance allocation. In line with this, results from the experiment show that sensitivity does not influence effort allocation under a fixed-wage. Moreover, as has been examined in additional analyses, a change in sensitivity on either task does not influence performance on either task. Thus, results for sensitivity are as stated by agency theory and expectancy theory. Moreover, given a fixed-wage, although participants do not adjust effort allocation when sensitivity on a task changes, they answer in the postexperimental questionnaire that they feel perceivably influenced by sensitivity. This apparent contradiction indicates that results for sensitivity support the assumption of Holmstrom and Milgrom (1991) that for a range of effort allocations an agent is indifferent and willing to follow a principal's preferences, despite feeling perceivably influenced by sensitivity.

With respect to precision, results for the fixed-wage compensation system indicate that, when precision on the easy task decreases, effort allocation does not change. Performance on the difficult task doesn't change either, whereas performance on the easy task decreases. The effect of precision on performance on the easy task is an unexpected result, as for a given fixed-wage there is no overall link between effort and outcome and hence precision should have no effect on performance. However, it can be assumed that participants derive positive valence from accomplishing the experimental task irrespective of performance-based monetary incentives. Consequently, a decrease in precision on the easy task, which has inherently high effort-performance expectancy, mitigates the link between effort and performance. Hence, the effort-performance expectancy and in turn the motivation of participants decreases and their performance decreases as well.

Under a fixed-wage compensation system, when precision on the difficult task decreases, more effort is shifted to the difficult task, contradicting both agency theory and expectancy theory. Consequently, performance on the easy task decreases. However performance on the difficult task does not increase, which can be attributed to task diffi-

culty. As Bonner and Sprinkle (2002) state, in settings like laboratory experiments, subjects are less likely to have the skills needed for difficult tasks than for easy tasks. If subjects are less likely to have the skills necessary for good performance in difficult tasks, then increases in effort allocation may not translate into performance increases. Furthermore, the results for effort allocation contradict Brügger (2011). He finds for a multi-task environment with two identical tasks, with precision being low on one task, and with precision being high on the other task, that under a fixed-wage (and in absence of career concerns), agents allocate effort evenly between tasks. Prospect theory offers a possible explanation for effort allocated towards the difficult task in this study, when precision on the difficult task decreases. The experimental task is to achieve an even (1:1) performance allocation. Assuming participants have reference-dependent preferences and are loss averse (Kahneman and Tversky 1979, Sloof and van Praag 2010), if subjects have a performance target in mind for the difficult task, they consider performance below this target as a loss. Low precision on the difficult task, that is, a high variance of the performance measure noise term, substantially increases the chance of the performance measure's value being below this target or even becoming negative. To increase the chance of a gain, that is, the performance measures' value being positive or above the reference target, participants allocate slightly more effort towards the difficult task. However, the change in effort allocation is too small for participants to translate it into increased performance on the difficult task.

Thus, with respect to precision, findings do not support the assertion of Holmstrom and Milgrom (1991) that agents are indifferent between a range of effort allocations under a fixed-wage and follow a principal's preferred effort allocation. In line with these results, participants answer in the postexperimental questionnaire that they feel perceivably influenced by precision. Overall, although participants seem to resist the temptation and do not react to performance measure sensitivity given the fixed-wage, they adjust their behavior when precision on a task changes.

Results for the performance-based compensation system are now discussed. As can be shown from the postexperimental questionnaire, participants receiving variable pay are also motivated to meet the expectations of the employer, i.e. provide high performance with respect to an even (1:1) performance allocation.

With respect to sensitivity, results show that participants allocate substantially more effort to a task measured with higher sensitivity. However, contradicting both agency

theory and expectancy theory, participants do not allocate relatively more effort to the easy task compared to the difficult task. With respect to task performance, when sensitivity on the easy task increases, performance on the easy task increases and performance on the difficult task decreases. Also, when sensitivity on the difficult task increases, performance on the difficult task increases (although not significantly) and performance on the easy task decreases. These results show that the predicted directions of the effects of sensitivity on effort allocation are in line with agency theory and expectancy theory. However, task difficulty has no moderating effect on the effect of sensitivity on effort allocation.

Under a performance-based compensation system, when precision on the easy task decreases more effort is allocated to the difficult task. When precision on the difficult task decreases, effort allocation does not change, that is, the change for the difficult task is less than the change for the easy task. These results are in line with expectancy theory and contradict agency theory. Consequently, when precision on the easy task decreases, performance on the easy task decreases and performance on the difficult task increases. When precision on the difficult task decreases, effort allocation does not change and accordingly performance on the easy task also does not change. However, performance on the difficult task increases, indicating an increase in effort intensity for the difficult task. This result contradicts expectancy theory and can also be attributed to loss aversion, as an increase in effort intensity reduces the chance of the performance measure of the difficult task becoming negative. In line with this, Sloof and van Praag (2010) examine the effect of performance measure precision on effort intensity in a single-task environment. They predict and find effort intensity to increase on a task, when it is measured imprecisely.

Results indicate that task difficulty moderates the effect of precision on effort allocation under a performance-based compensation system, as the effect of precision on effort allocation is significantly larger for a change in precision on the easy task compared to a change in precision on the difficult task. The result for the easy task is in line with Brügger (2005), who indicates for performance-based monetary incentives that more effort is allocated to the task that is measured more precisely. Brügger (2005) employs a two-measure performance-based compensation system with two identical tasks having equal difficulty, equal incentive weights, and equal sensitivity on each respective performance measure (both tasks are similar to the easy task employed in this study). On the other hand, Sloof and van Praag (2008) could not observe a change in

effort allocation in a simultaneous multi-task environment with two equal tasks, when precision of a performance measure changed.

Overall, results of this study show the complex interactions between performance measure characteristics and task characteristics on effort allocation. The implications for economic theory are twofold. First, in the linear agency theory model (LEN model), precision does not affect effort allocation under a given performance-based compensation system (Sloof and van Praag 2008, 2010). Including an expectancy factor in performance measurement models, as suggested by Sloof and van Praag (2010) or implementing a multiplicative noise term instead of an additive noise term (Christensen and Feltham 2008) could account for behavioral effects of precision. Second, the findings support the assertion of Schnedler (2006) that information regarding performance measure characteristics alone is not sufficient to predict an agent's effort allocation, as task difficulty can influence the effects of performance measure characteristics in a multi-task environment. Hence, analytical multi-task models in performance measurement could analyze effects of performance measure characteristics while allowing tasks to have different marginal costs of effort. This implies for the design of optimal management control, information, and compensation systems that, besides performance measure characteristics, task characteristics have to be taken into account as well.

This study is subject to limitations that restrict the extent to which the findings can be generalized. As in all experiments, the choice of parameters influences results. With respect to sensitivity it is unlikely that different parameters would significantly change the findings and alter the conclusions of this study. With regard to precision, parameters were chosen for variations in performance measure noise to be *sufficiently* high for the low precision level to exceed a subjective 'tactile threshold' for each participant. However, noise was *not exceedingly* high so that the respective performance measure at the end of each round did not become completely random with respect to a subject's effort.

Furthermore, subjects move through the two tasks sequentially, and the task order is not changed throughout the experiment. This choice was made to reflect the notion of Sloof and van Praag (2010) that when presenting tasks on a single screen subjects may simplify the situation, and aggregate the performance measure precision of both tasks which could lead to undifferentiated behavior with respect to precision. Nevertheless confounding order effects could appear, when participants associate the easy and difficult task with treatment and control conditions. To avoid possible effects of a loss in

concentration due to the close proximity in time in the sequential environment, a mandatory pause of six seconds was set up between the two tasks based on experience from a pre-test. However a loss in concentration cannot be ruled out completely.

Although both tasks employed in this study are simple production tasks, they might differ in task attractiveness. Bailey and Fessler (2011) find a significant three-way interaction between task complexity, task attractiveness, and compensation system on performance. However, they find no interaction effect between task complexity and task attractiveness and no significant main effect of task attractiveness to performance. Thus, task attractiveness only moderates the interaction effect of task complexity and compensation scheme and does not influence task performance directly. Task complexity as a construct subsumes task difficulty (Bonner and Sprinkle 2002). By analogy to the results of Bailey and Fessler (2011), as the hypotheses in this study are stated for a given compensation system, differences in task attractiveness between the easy and the difficult task are not expected to change the results of this study. However, future studies should control for other task dimensions in the postexperimental questionnaire.

Future studies could address these limitations and allow participants to alter the order of the tasks during each round. According to the categorization of Bonner et al. (2000), the two experimental tasks employed in this study are production tasks. Thus, different task types with varying difficulties could be employed sequentially and/or simultaneously to test the generalizability of the results with respect to task type. Further studies could also analyze whether the results apply for judgment and choice tasks as well as for problem-solving tasks.

References

- Bai, G., F. Coronado, and R. Krishnan. 2010. The Role of Performance Measure Noise in Mediating the Relation between Task Complexity and Outsourcing. *Journal of Management Accounting Research* 22 (1): 75-102.
- Bailey, C. D., and N. J. Fessler. 2011. The Moderating Effects of Task Complexity and Task Attractiveness on the Impact of Monetary Incentives in Repeated Tasks. *Journal of Management Accounting Research* 23 (1): 189-210.
- Bailey, C. D., L. D. Brown, and A. F. Cocco. 1998. The effects of monetary incentives on worker learning and performance in an assembly task. *Journal of Management Accounting Research* 10 (1): 119-131.
- Baiman, S. 1982. Agency Research in managerial accounting: a survey. *Journal of Accounting Literature* 1 (1): 154-213.
- Baiman, S. 1990. Agency research in managerial accounting: A second look. *Accounting, Organizations and Society* 15 (4): 341-371.
- Baker, G. P. 1992. Incentive Contracts and Performance Measurement. *Journal of Political Economy* 100 (3): 598-614.
- Baker, G. P. 2002. Distortion and Risk in Optimal Incentive Contracts. *Journal of Human Resources* 37 (4): 728-751.
- Banker, R. D., and S. M. Datar. 1989. Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation. *Journal of Accounting Research* 27 (1): 21-39.
- Bonner, S. E., R. Hastie, G. B. Sprinkle, and S. M. Young. 2000. A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting. *Journal of Management Accounting Research* 12 (1): 19-64.
- Bonner, S. E., and G. B. Sprinkle. 2002. The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27 (4): 303-345.
- Brüggen, A. 2005. *Incentives in multi-task settings*. Dissertation. Maastricht University.
- Brüggen, A. 2011. Ability, Career Concerns, and Financial Incentives in a Multi-Task Setting. *Journal of Management Accounting Research* 23 (1): 211-229.
- Brüggen, A., and F. Moers. 2007. The Role of Financial Incentives and Social Incentives in Multi-Task Settings. *Journal of Management Accounting Research* 19 (1): 25-50.
- Buckless, F. A., and S. P. Ravenscroft. 1990. Contrast Coding: A Refinement of ANOVA in Behavioral Analysis. *The Accounting Review* 65 (4): 933-945.
- Campbell, D. J. 1988. Task Complexity: A Review and Analysis. *Academy of Management Review* 13 (1): 40-52.
- Chow, C. W. 1983. Competitive Manuscript Award: The Effects of Job Standard Tightness and Compensation Scheme on Performance: An Exploration of Linkages. *The Accounting Review* 58 (4): 667-685.
- Chow, C. W., J. C. Cooper, and W. S. Waller. 1988. Participative Budgeting: Effects of a Truth-Inducing Pay Scheme and Information Asymmetry on Slack and Performance. *The Accounting Review* 63 (1): 111-122.

- Christensen, P. O., and G. A. Feltham. 2008. *Performance evaluation. Economics of accounting*. 2. Edition. New York, NY: Springer.
- Datar, S., S. C. Kulp, and R. A. Lambert. 2001. Balancing Performance Measures. *Journal of Accounting Research* 39 (1): 75-92.
- Demski, J. S. 2008. *Managerial uses of accounting information*. 2. Edition. New York, NY: Springer.
- Demski, J. S., J. C. Fellingham, and H. H. Lin. 2009. Tension Relevance. *Journal of Management Accounting Research* 21 (1): 241-248.
- Dillard, J. F., and J. G. Fisher. 1990. Compensation Schemes, Skill Level, and Task Performance: An Experimental Examination. *Decision Sciences* 21 (1): 121-137.
- Feltham, G. A., and M. G. H. Wu. 2000. Public reports, information acquisition by investors, and management incentives. *Review of Accounting Studies* 5 (2): 155-190.
- Feltham, G. A., and J. Xie. 1994. Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations. *The Accounting Review* 69 (3): 429-453.
- Fischbacher, U. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2): 171-178.
- Fisher, J. G., J. R. Frederickson, and S. A. Peffer. 2002. The effect of information asymmetry on negotiated budgets: an empirical investigation. *Accounting, Organizations and Society* 27 (1-2): 27-43.
- Gill, D., and V. Prowse. 2012. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review* 102 (1): 469-503.
- Hannan, R. L., G. P. McPhee, A. H. Newman, and I. D. Tafkov. 2013. The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task Environment. *The Accounting Review* 88 (2): 553-575.
- Hecht, G., I. D. Tafkov, and K. L. Towry. 2012. Performance Spillover in a Multitask Environment. *Contemporary Accounting Research* 29 (2): 563-589.
- Hemmer, T. 1996. On the Design and Choice of "Modern" Management Accounting Measures. *Journal of Management Accounting Research* 8 (1): 87-116.
- Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics* 10 (1): 74-91.
- Holmstrom, B., and P. Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica* 55 (2): 303-328.
- Holmstrom, B., and P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics & Organization* 7 (Special Issue): 24-52.
- IBM Corporation. 2012. IBM SPSS Advanced Statistics 21. Copyright IBM Corporation 1989.
- Ittner, C. D., D. F. Larcker, and M. V. Rajan. 1997. The choice of performance measures in annual bonus contracts. *The Accounting Review* 72 (2): 231-255.
- Jensen, M. C., and K. J. Murphy. 1990. Performance Pay and Top-Management Incentives. *Journal of Political Economy* 98 (2): 225-64.

- Kachelmeier, S. J. 1996. Discussion of 'Tax Advice and Reporting Under Uncertainty: Theory and Experimental Evidence'. *Contemporary Accounting Research* 13 (1): 81-89.
- Kachelmeier, S. J., B. E. Reichert, and M. G. Williamson. 2008. Measuring and Motivating Quantity, Creativity, or Both. *Journal of Accounting Research* 46 (2): 341-373.
- Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2): 263-291.
- Kahneman, D., and A. Tversky. 1996. On the reality of cognitive illusions. *Psychological Review* 103 (3): 582-591.
- Krishnan, R., J. L. Luft, and M. D. Shields. 2005. Effects of Accounting-Method Choices on Subjective Performance-Measure Weighting Decisions: Experimental Evidence on Precision and Error Covariance. *The Accounting Review* 70 (4): 1163-1192.
- Lambert, R. A. 2007. Agency Theory and Management Accounting. In: C. S. Chapman, A. G. Hopwood, and M. D. Shields (Ed.): *Handbook of Management Accounting Research I*. Oxford, U.K.: Elsevier: 247-268.
- Libby, R., R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27 (8): 775-810.
- March, J., and H. A. Simon. 1958. *Organizations*. Oxford, U.K.: Wiley.
- Merchant, K. A., and W. A. van der Stede. 2012. *Management control systems. Performance measurement, evaluation and incentives*. 3. Edition. Harlow, U.K.: Pearson, Financial Times/ Prentice Hall.
- Moers, F. 2000. *The Role of Performance Measure Characteristics in the Design of Incentive Systems: An Empirical Analysis*. MARC Working Paper MARC-WP/3/2000-14.
- Moser, D. V. 1998. Using an experimental economics approach in behavioral accounting research. *Behavioral Research in Accounting* 10 (Supplement): 94-110.
- Prendergast, C. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37 (1): 7-63.
- Prendergast, C. 2002. The tenuous trade-off between risk and incentives. *The Journal of Political Economy* 110 (5): 1071-1102.
- Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2 (6): 110-114.
- Schnedler, W. 2006. *Task Difficulty, Performance Measure Characteristics, and the Trade-Off between Insurance and Well-Allocated Effort*. Discussion Paper Series 425. University of Heidelberg: Department of Economics.
- Shields, M. D., and W. S. Waller. 1988. A behavioral study of accounting variables in performance—incentive contracts. *Accounting, Organizations and Society* 13 (6): 581-594.
- Sillamaa, M. A. 1999. How work effort responds to wage taxation: An experimental test of a zero top marginal tax rate. *Journal of Public Economics* 73 (1): 125-134.

- Sloof, R., and C. M. van Praag. 2008. Performance measurement, expectancy and agency theory: An experimental study. *Journal of Economic Behavior & Organization* 67 (3): 794-809.
- Sloof, R. and C. M. van Praag. 2010. The effect of noise in a performance measure on work motivation: A real effort laboratory experiment. *Labour Economics* 17 (5): 751-765.
- Smith, R. E., J. C. Ascough, and R. F. Ettinger. 1976. The relationship between expectancy of success and task difficulty as a function of absolute and relative success criteria. *Social Behavior and Personality: an international journal* 4 (2): 277-282.
- Sprinkle, G. B., and M. G. Williamson. 2007. Experimental Research in Managerial Accounting. In: C. S. Chapman, A. G. Hopwood, and M. D. Shields (Ed.): *Handbook of Management Accounting Research 1*. Oxford, U.K.: Elsevier: 415-444.
- Stevens, D. E. 2002. The Effects of Reputation and Ethics on Budgetary Slack. *Journal of Management Accounting Research* 14 (1): 153-171.
- Van Eerde, W., and H. Thierry. 1996. Vroom's expectancy models and work-related criteria: A meta-analysis. *Journal of Applied Psychology* 81 (5): 575-586.
- Vroom, V. H. 1964. *Work and motivation*. New York, NY: Wiley.
- Waller, W. S., and C. W. Chow. 1985. The self-selection and effort effects of standard-based employment contracts: A framework and some empirical evidence. *The Accounting Review* 55 (3): 458-476.
- Williams, E. J. 1949. Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Chemistry* 2 (2): 149-168.
- Wood, Robert E. 1986. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes* 37 (1): 60-82.

Budget Negotiation Structure, Fairness, and Agreement – An Experimental Analysis of Performance, Perception of Performance, and Reputation

Christian Meier / Robert Obermaier / Christine Selbitschka[✉]

Abstract

Budgets linked to performance evaluation and compensation influence subordinates' performance and are regularly determined by employing negotiations between superiors and subordinates. In a budget negotiation, initial budget proposal and final authority to determine the budget influence subordinates' fairness concerns, with fairness concerns being important antecedents of subordinates' performance. Therefore, the first issue examined in this study is the effect of initial budget proposal and final authority on subordinates' performance. The second issue analyzed is if superiors' subjective performance evaluations differ from subordinates' performance. It is examined as a third issue, whether negotiation agreement leads to a trade-off between accuracy and leniency when superiors assess subordinates' reputation. To address these three issues, a real-effort laboratory experiment is conducted. Results show that subordinates' performance is high, except for subordinates making the initial proposal and superiors having final authority. In this situation subordinates frame the negotiation outcome as an unfair loss and their performance decreases. Also, superiors' perception of subordinates' performance differs significantly from subordinates' performance. This can be attributed to superiors evaluating subordinates' performance based on their own preferences. Finally, for superiors' assessment of subordinates' reputation, a trade-off between accuracy and leniency is identified. Reputation is only positively related to subordinates' performance, when negotiation agreement is low. This study contributes to an increased understanding of how social preferences affect incentive problems and assessment of reputation. Also, this study is the first to examine perceived performance and reputation from a superior's perspective.

Keywords: Participative Budgeting, Fairness, Negotiation, Final Authority, Initial Budget Proposal, Agreement, Performance, Subjective Performance Evaluation, Reputation.

✉ Christian Meier and Robert Obermaier: University of Passau, Chair of Accounting and Control, Innstraße 27, 94032 Passau, Germany, E-Mail: controlling@uni-passau.de; Phone: 0851-509-3271. Christine Selbitschka: Hindenburgstraße 33, 94227 Zwiesel, E-Mail: christine.selbitschka@gmx.de.

1 Introduction

The use of budgets for performance evaluation and compensation comprises a major aspect of most organizations' management control systems (Hopwood 1976). Budgets linked to performance evaluation and compensation become decision-influencing tools for motivating subordinates' performance (Merchant and van der Stede 2012, Sprinkle and Williamson 2007). Moreover, budgets are often determined participatively by employing negotiations between a superior and a subordinate (Anthony and Govindarajan 2007, Umapathy 1987), with fairness concerns in budget negotiations being important antecedents of subordinates' satisfaction with the budgeting-setting process and subordinates' performance (Fisher et al. 2002a, Libby 2001, Lindquist 1995).

Two important aspects of the budget negotiation structure that determine subordinates' fairness concerns are: (1) which party makes the initial budget proposal and (2) which party has final authority to determine the budget in case of a negotiation impasse.

The initial budget proposal constitutes a reference point relative to a final budget in a budget negotiation. Generally, reference points are important because other outcomes are compared to them and are evaluated in terms of this comparison, with a gain in a negotiation being perceived as fair, and a loss being perceived as unfair (Kahneman 1992). Prior research on two-party negotiations shows that negotiators tend to be highly affected by anchors (Kristensen and Gärling 1997, Whyte and Sebenius 1997, Ritov 1996, Thompson 1995, Kahneman 1992, Northcraft and Neale 1987) with negotiation outcomes being more strongly influenced by initial offers than by subsequent concessionary behavior (Neale and Bazerman 1991, Yukl 1974, Liebert et al. 1968).

Final authority appears to determine the subordinate's framing of the budget-setting process (Rankin et al. 2008). Subordinates tend to frame a budget negotiation where a superior has final authority to determine the budget as a strategic interaction, where each party has egocentric fairness preferences and acts in his or her self-interest (Rankin et al. 2008, Thompson and Loewenstein 1992). Whereas a budget negotiation where a subordinate has final authority to determine the budget and hence to distribute the profit between himself and the superior, may rather be framed as a "resource allocation" or an "ethical dilemma" with reciprocal fairness concerns (Rankin et al. 2008, Falk and Fischbacher 2006, Camerer 2003).

As both initial budget proposal and final authority influence subordinates' fairness concerns in a budget negotiation, with fairness concerns being an important antecedent of subordinates' performance, the first issue examined in this study is, whether initial budget proposal and final authority influence subordinates' performance. Research in managerial accounting analyzing how structural features of the budget-setting process affect incentive problems is fundamental for the design of management control systems (Lambert 2007).

Also, organizations regularly employ subjective performance evaluations of superiors to assess subordinates' performance (Breuer et al. 2013). Subjective performance evaluation has been focus of several studies (Breuer et al. 2013, Maas et al. 2012, Bol 2011, Chang et al. 2008, Moers 2005, Prendergast 1999, Prendergast and Topel 1993). A central result of these studies is that subjective performance evaluations tend to be biased and can lead to an inefficient allocation of employees to tasks or jobs. Furthermore, managers tend to respond to their own incentives and preferences when subjectively evaluating performance (Bol 2011). Overall, results suggest that managers may fail to obtain or fully consider all information available for performance evaluation (Maas et al. 2012). Moreover, Bol (2011) shows that centrality bias negatively affects performance improvement, whereas leniency bias is positively associated with future performance, as it can improve perceived fairness and, in turn, employee motivation.

Consequently, it needs to be examined in a participative budgeting environment whether superiors' subjective performance evaluations differ from subordinates' performance. In prior participative budgeting research either the experimentalist acts as the superior (Fisher et al. 2003, Stevens 2002, Webb 2002, Evans et al. 2001, Chow et al. 1995, 1994, 1991, 1988, Young et al. 1993, Waller 1988, Young 1985), or subjective performance evaluation is not being analyzed (Rankin et al. 2008, 2003, Fisher et al. 2006, 2002a, 2002b, 2000). Research identifying possible bias in subjective performance evaluation based on budgets is important for designers of management control systems, as it can mitigate incentive problems, and can determine how much discretion to allow managers in budget-based compensation plans (Bol 2011). Thus, this study addresses potential subjective performance evaluation bias as a second issue.

As another important aspect of participative budgeting, prior research has examined reputation concerns from the perspective of subordinates. Subordinates have fewer incentives to misrepresent private information when their reputation is based on output

and budget accuracy (Webb 2002). Also, reputation and ethical concerns significantly influence the level of budgetary slack they create (Stevens 2002). However, superiors' assessment of subordinates' reputation has not been examined so far. Theoretically, subordinates' reputation, a criterion of managerial effectiveness, should be positively associated with their performance (Tsui 1984).

However, Breuer et al. (2013) find a trade-off between accuracy and leniency in subjective performance evaluations. Evaluations are more lenient the closer the social ties between supervisor and evaluated employee. In budget negotiations, agreement indicates cooperative behavior between a superior and a subordinate (Pruitt and Carnevale 1993). Also, superiors and subordinates who expect future negotiations act more cooperatively (Fisher et al. 2006). Thus, negotiation agreement should increase social ties in repeated interactions. Therefore, as a third issue, it is examined whether negotiation agreement also leads to a trade-off between accuracy and leniency when superiors assess subordinates' reputation. It is important to examine social motives in managerial accounting, as social ties may lead to nepotism, so that superiors follow their personal social preferences and bias evaluation outcomes or fail to identify training needs of employees when they are judged too leniently (Breuer et al. 2013). Ultimately, analyzing social motives may help explain why certain procedures are observed in practice and may suggest changes in the design of managerial accounting procedures (Sprinkle and Williamson 2007).

To address these three issues, a real-effort laboratory experiment is conducted. Initial budget proposal and final authority are manipulated between-subjects and period is manipulated within-subjects. Furthermore, negotiation agreement and subordinates' performance are measured between-subjects. The following research questions are stated: (1) Do initial budget proposal and final authority influence a subordinate's performance? (2) Does a superior's perception of a subordinate's performance differ from a subordinate's performance? And (3) do budget negotiation agreement and a subordinate's performance influence a subordinate's reputation?

Results show a significant interaction effect of initial budget proposal and final authority on subordinates' performance. Subordinates' performance is high, except for the situation where subordinates make the initial budget proposal and superiors have final authority. In this situation subordinates frame the negotiation outcome as an unfair loss and their performance decreases. Furthermore, results indicate that subordinates' per-

formance is high despite a high budget level, when the budget negotiation is perceived as being fair. The results are consistent with subordinates having egocentric fairness concerns given superior final authority and reciprocal fairness concerns given subordinate final authority. Also, results indicate that superiors' perception of subordinates' performance differs significantly from subordinates' performance, which can be attributed to superiors evaluating subordinates' performance based on their own preferences. Finally, for assessment of subordinates' reputation, a trade-off between accuracy and leniency is identified. Reputation is only positively related to performance, when negotiation agreement between a superior and a subordinate is low. Given high agreement, reputation is high irrespective of subordinates' performance.

The contribution of this study is fourfold. First, this study contributes to an increased understanding of how social preferences affect incentive problems and assessment of reputation. Budget negotiation structure affects fairness concerns, an important antecedent of subordinates' performance. Budget negotiation agreement affects social ties, with social ties being an impediment to accurate reputation assessment. From a theoretical perspective, these results are of particular interest for the development of analytical models of participative budgeting, which should incorporate social preferences for precise predictions of subordinates' and superiors' behavior. Second, besides analyzing subordinates' performance as an objective economic measure of performance in a participative budgeting environment, this study is the first to examine perceived performance and reputation from a superior's perspective as social-psychological measures. These measures are important because parties often do not have the quantitative data necessary to make precise judgments of a negotiation situation, that is, their understanding of the negotiation is based on their perceptions (Thompson 1990). Thus, social-psychological measures can complement objective economic analyses in the budget-setting process, as they allow analyzing if and whether perceptions differ from objective outcomes. Third, in examining the effects of budget negotiation structure on subordinates' performance, this study contributes to the literature identifying how structural features of information, accounting, and compensation systems affect incentive problems, which is fundamental for the design of management control systems. Fourth, this study extends reference cognitions theory (Folger 1986) to participative budget negotiations. Reference cognitions theory has already been successfully applied to a non-participative budgeting setting (Libby 2001).

The remainder of this paper is structured as follows: The next section develops the hypotheses. The two subsequent sections describe the experimental design and results. The final section provides a discussion of the results including their implications and limitations.

2 Theory and Hypotheses

2.1 Setting and Incentives

The budgetary setting in this study is based on Fisher et al. (2006, 2002a, 2002b, 2000). A superior and a subordinate constitute a dyad (pair), which interacts repeatedly for a finite number of periods.

In each period a dyad determines a production budget through a budget negotiation. In a given period, a budget negotiation consists of a finite number of negotiation rounds, and a negotiation round consists of an initial offer and a counteroffer. At the beginning of each negotiation round, depending on which party is entitled to make the initial budget proposal, either the superior or the subordinate makes the initial offer, which is followed by a counteroffer of the other party. If an offer equals a counteroffer in a given negotiation round, this means that a dyad agrees on a budget, the budget negotiation in the given period ends, and the budget is set jointly. Otherwise the budget negotiation continues with the next round. If a dyad fails to agree during the finite number of negotiation rounds in the given period, there is a negotiation impasse and, depending on which party has final authority, either the superior or the subordinate determines the budget unilaterally. In each period, once the budget is set, a subordinate exerts effort on a task, the respective payoffs are determined, and both parties receive information on a subordinate productions' performance and obtain their respective remuneration. With respect to the information environment, in the first period a superior only knows the performance capabilities of subordinates on average and does not know the performance capability of his particular subordinate. However, each dyad knows both compensation schemes from the beginning. The compensation schemes are as follows (Fisher et al. 2006, 2002a, 2002b, 2000):

$$P_{subordinate} = \begin{cases} = F & \text{if } X \leq B \\ = F + A(X - B) & \text{if } X > B \end{cases} \quad (1)$$

$$P_{superior} = \begin{cases} = R(DX - F) & \text{if } X \leq B \\ = R\{DX - [F + A(X - B)]\} & \text{if } X > B \end{cases} \quad (2)$$

where: P = payoff for a period; X = performance; B = budget; F = subordinate fixed wage; A = subordinate piece rate per unit of production over budget; D = profit, exclusive of agent compensation, per unit produced; and R = superior's percentage of firm profit.

The subordinate's compensation scheme in equation (1) is a typical quota scheme based on the production budget (Bonner et al. 2000). A subordinate receives a fixed wage irrespective of performance until the production budget is reached. If a subordinate's performance exceeds the budget, he receives a piece rate for each additional unit of output. Thus, a subordinate's payoff is positively related to performance and negatively related to budget level (Fisher et al. 2000). The compensation scheme gives him an economic incentive to produce up to his performance capabilities for budgets below his performance capabilities, but gives him little if no economic incentive to exert effort when the budget is set above his performance capabilities (Fisher et al 2002a). A subordinate maximizes his payoff, when the budget is minimized, i.e. is equal to zero.

The superior's compensation scheme in equation (2) is a residual contract based on firm profit (Fisher et al. 2006, 2000). His payoff is positively related to both, budget level and subordinate performance. It is maximized in a given period, when the budget, a subordinate's performance and a subordinate's performance capabilities are equal. However, as a subordinate has only an economic incentive to produce up to his performance capabilities for budgets below his performance capabilities, from a superior's perspective, the budget should not be determined above a subordinate's performance capabilities. Thus, the superior faces a trade-off between his payoff and a subordinate's motivation. A high budget level maximizes his payoff, but it needs to stay below a subordinate's performance capabilities (Fisher et al. 2002a, 2000).

Overall, the compensation schemes in equations (1) and (2) lead to an interpersonal conflict. In a budget negotiation, a subordinate wants to negotiate a low budget and a superior wants to negotiate a high budget. Consequently, a subordinate (superior) generally will make a low (high) initial budget proposal. Accordingly, Fisher et al. (2000) provide empirical evidence that subordinates (superiors) take low (high) initial negotiation positions and then make concessions with respect to the final budget. However, both parties have an incentive to keep the budget level below a subordinate's performance capabilities.

2.2 Effects of Initial Budget Proposal and Final Authority on Subordinates' Performance

Traditional assumptions in economic theory are that individuals are rational expected utility maximizers with well-defined preferences that conform to the axioms of expected utility theory (e.g. Baiman 1990, 1982). They are motivated by self-interest only (Eisenhardt 1989) and their utility function consists of only their own wealth and their leisure.¹ They prefer increases in wealth and increases in leisure, and are assumed to exert no effort on a task, unless it contributes to their own wealth (Bonner and Sprinkle 2002). Under the assumption that an economic agent maximizes wealth, this standard reasoning suggests that, if the budget is below his performance capabilities, the subordinate exerts effort until his performance capabilities are reached and exerts no effort if the budget is above his performance capabilities, as effort is assumed to be costly (Fisher et al. 2006, 2002a, 2000). Moreover, a wealth-maximizing superior will set the budget close to a subordinate's performance capabilities and a wealth-maximizing subordinate will minimize the budget level.

However, Fisher et al. (2000) assume that subordinates have fairness concerns and prefer a budget that is above zero but below their performance capabilities (Fisher et al. 2002a, Koford and Penno 1992, Miller 1992, Waller 1988). In contrast to standard economic reasoning and in accordance with fairness preferences, their empirical results show that when subordinates' have final authority to determine the budget in a budget negotiation, the budget level is well above the minimum but below subordinates' average performance (Fisher et al. 2000). In line with this, Fehr and Schmidt (2003) show that individuals prefer outcomes they perceive to be fair, even if these outcomes are costly in monetary terms. Additionally, prior research in experimental economics has shown that many individuals have social preferences, that is, their utility function consists not only of their own wealth, but also of the wealth of others (Fehr and Gintis 2007, Camerer 2003). Furthermore, economic fairness theories state that in addition to their preference for wealth, some people have a preference for fairness (Falk and Fischbacher 2006, Dufwenberg and Kirchsteiger 2004, Charness and Rabin 2002, Bolton and Ockenfels 2000, Fehr and Schmidt 1999, Levine 1998, Rabin 1993).

Prior research in experimental economics indicates that concerns for fairness and reciprocity motivate a substantial number of people (Falk et al. 2008). In experimental

¹ A recent analytical study presenting a theory of participative budgeting also assumes that managers are motivated by self-interest only (Heinle et al. 2014).

economics, Gächter and Falk (2002) show for repeated interactions in a gift exchange game that reciprocity and repeated game incentives have a considerable efficiency-enhancing effect on performance. Fehr and Schmidt (2004) conduct a two-task principal-agent experiment in which only one task is contractible. The principal can offer either a piece-rate contract or a voluntary bonus to the agent. Voluntary bonus contracts strongly outperform piece-rate contracts. Many principals reward high effort on both tasks with substantial voluntary bonuses. Agents anticipate this and provide high effort on both tasks. Again, this behavior contradicts standard economic reasoning but is consistent with fairness concerns. With respect to the budgeting process, economic theories of fairness and the related empirical evidence imply that a budget perceived as fair is attended by high motivation and performance, whereas a budget perceived as unfair decreases motivation and performance.

Referent cognitions theory (Folger 1986) states that decision-making procedures also play an important role in determining individuals' fairness perceptions (besides outcomes). According to referent cognitions theory, in a situation involving outcomes allocated by a decision maker, resentment is maximized when individuals believe they would have obtained better outcomes if the decision maker had used other processes that should have been implemented. Thus, when individuals receive an unfair outcome, their judgments become inherently referential, i.e. they make comparisons between the outcome they receive and a reference point. A reference point constitutes an outcome they feel they should have received given their inputs relative to the inputs of others (Libby 2001, Adams 1965). If the reference point indicates an inequity between the outcome an individual receives and the outcome they feel they should have received relative to others, feelings of resentment result (Libby 2001).

Thus, referent cognitions theory suggests that the fairness of the process used to set budgets, i.e. procedural fairness, is at least as important as the budget level (Lindquist 1995, Lind et al. 1990). A fair budget-setting process increases commitment to the budget, which in turn increases effort and performance (Fisher et al. 2002a). The results of Fisher et al. (2002a) support the notion of procedural fairness and indicate that agreement in a budget negotiation indicates a fair process and increases subordinate performance, whereas performance is lower for a negotiation impasse. Likewise, Fisher et al. (2002a) find that even if superiors set the budgets below subordinates' performance capabilities, subordinates view superiors imposing a budget following a failed negotiation as an unfair process, resulting in dissatisfaction with both the budget-setting

process and the resulting budget. These subordinates then are less committed to achieving their budgets and their performance is low.

In line with referent cognitions theory, the initial budget proposal can be interpreted as a reference point relative to the final budget.² A reference point separates the domain into regions of desirable outcomes, i.e. gains and undesirable ones, i.e. losses. When an alternative is compared to the reference point, the comparison is coded in terms of the advantages and disadvantages of that alternative (Kahneman 1992). Therefore, in a budget negotiation, from the perspective of a subordinate, a final budget which is higher than the initial budget proposal will be perceived as a loss, and a final budget which is lower than the initial budget proposal will be perceived as a gain. In addition to their effect on the valuation of outcomes, reference points also affect negotiations by influencing judgments of what is fair or unfair (Kahneman 1992). In negotiations perceptions of fairness tend to be egocentric and negotiators recall more information about the situation that supports their own position compared to that of the other party (Thompson and Loewenstein 1992). Therefore, a gain in a negotiation will be perceived as fair, whereas a loss will be perceived as unfair (Kahneman 1992), i.e. from the perspective of a subordinate a final budget which is lower than the initial budget proposal will be perceived as fair, and a final budget which is higher than the initial budget proposal will be perceived as unfair.

In addition, final authority is an important determinant of the budget negotiation structure (Fisher et al. 2000), which appears to determine subordinates' framing of the budget-setting process (Rankin et al. 2008). According to referent cognitions theory, an individual who is responsible for choosing his own level of outcomes cannot find fault with somebody else due to procedural reasons. On the other hand, reasons for challenging a procedure are more present when somebody else was responsible for decisions that led to the outcomes obtained, and referent cognitions theory predicts resentment of unfair treatment when unjustified procedures result in unsatisfactory outcomes (Cropanzano and Folger 1989).

² Prior research has shown that negotiation outcomes are more strongly influenced by initial offers than by subsequent concessionary behavior (Neale and Bazerman 1991, Yukl 1974, Liebert et al. 1968). Also, Rubin and Brown (1975) note that early moves in negotiation are critical to constructing the parameters of the negotiation. Specifically, research on dyad negotiations suggests that negotiators tend to be highly affected by anchors in negotiation (Kristensen and Gärling 1997, Whyte and Sebenius 1997, Ritov 1996, Thompson 1995, Kahneman 1992, Northcraft and Neale 1987).

When a superior has final authority, in case of a negotiation impasse, he determines the final budget. Thus, according to referent cognitions theory, he is responsible for the budget and subordinates feel resentment of unfair treatment when an unfair budget negotiation leads to an unfair budget. In line with this, subordinates tend to frame a budget negotiation where a superior has final authority as a strategic interaction where each party acts in his or her self-interest (Rankin et al. 2008). In such a situation, negotiators' judgments of fair outcomes are biased in an egocentric direction³ (Thompson and Loevenstein 1992) and the initial budget proposal constitutes a reference point relative to the final budget.

Also, according to referent cognitions theory, a subordinate who is responsible for choosing his own budget level cannot fault a superior on procedural reasons. Thus, subordinates may frame the situation with a subordinate having final authority rather as a "resource allocation"⁴ or an "ethical dilemma" (Rankin et al. 2008) than as a negotiation. If a subordinate has final authority, he is responsible to allocate the payoff between him and his superior depending on his individual fairness preferences and he will determine a budget which he believes to be fair. Moreover, conforming to referent cognitions theory (Folger 1986), when a fair budget level is determined, information about the process used to set the budget is unimportant to an individual's motivation to achieve the budget level, i.e. when a subordinate has final authority the initial budget proposal no longer constitutes a reference point relative to the final budget.

The above discussion suggests that if a superior makes the initial budget proposal, in line with his incentives, he will set a high initial proposal. However, during the budget negotiation process, the superior will make concessions and the final budget will be lower than the initial budget proposal. Furthermore, if the superior has final authority a subordinate's reference-dependent fairness preferences will be egocentric and during the budget negotiation he will perceive concessions from the superior as a fair gain.

³ For example, negotiators who are unable to reach mutually acceptable agreements blame their counterparty for the failure, whereas they usually attribute success to themselves (Brandstätter et al. 1983).

⁴ Also, from an economic perspective, the situation where subordinates have final authority can be compared to a dictator game, where one player, the "dictator," divides an amount of money between himself and another person, the "responder" (Camerer 2003). In this game, a rational utility maximizing dictator, with self-interest only, will allocate the whole amount to himself. However, in several experimental studies dictators allocate to responders a significant amount of money, which can be attributed to reciprocity and outcome fairness concerns (Falk and Fischbacher 2006). These findings support reciprocal fairness concerns and are in line with participative budgeting studies, where subordinates with final authority consistently determine a budget significantly deviating from their payoff-maximizing optimum and distribute payoffs to superiors (Chow et al. 1988, Waller 1988, Young 1985).

Thus, he will perceive the negotiation as a fair process and exert effort up to his performance capabilities, irrespective of the budget level.

Now, if a subordinate makes the initial budget proposal, in line with his incentives, he will set a low initial proposal. However, during the budget negotiation process, the subordinate will make concessions and the final budget will be higher than the initial budget proposal. Furthermore, if a superior has final authority, a subordinate's reference-dependent fairness preferences will be egocentric and during the budget negotiation he will perceive his concessions to the superior as an unfair loss. This unfair budget negotiation process demotivates him and even if the budget is set below his performance capabilities he exerts less effort than his performance capabilities. In line with this prediction, Libby (2001) shows for a nonparticipative budgeting setting that an unfair budget level determined under a fair budgeting process leads to higher performance than an unfair budget level determined under an unfair budgeting process.

Referent cognitions theory predicts that if a subordinate has final authority to determine the budget he is responsible for the budget and cannot blame a superior for an unfair process. Hence, he will frame the situation rather as a resource allocation than as a negotiation and the budget negotiation process becomes unimportant to him. Thus, irrespective of initial budget proposal, he will determine a budget below his performance capabilities, which he believes is fair, and exerts effort up to his performance capabilities. Accordingly, Libby (2001) finds that individuals which are assigned to a fair budget level perform equally well regardless of the fairness or the unfairness of the budgeting process employed. Overall, a subordinate's performance is lowest when a subordinate makes the initial budget proposal and a superior has final authority, and approximately equal otherwise. The following hypothesis can be stated:

H1: *If a superior makes the initial budget proposal, a subordinate's performance will be high, irrespective of final authority. If a subordinate makes the initial budget proposal, his performance will be high, when he has final authority and low, when a superior has final authority.*

2.3 Superiors' Perception of Subordinates' Performance

Negotiators typically do not perform a fully rational analysis of a negotiation situation either because the relevant information is not available, or because they ignore or distort information (Neale and Bazerman 1991). Therefore, negotiators' perceptions may differ substantially from objective economic analyses (Thompson 1990, Thompson and

Hastie 1990). Luft and Libby (1997) show in the domain of transfer pricing that because of self-serving biases, negotiating managers have different expectations regarding what constitutes a 'fair' transfer price. Also in transfer price negotiations, Chang et al. (2008) show that a loss frame, compared to a gain frame, exacerbates managers' self-serving biases.

Furthermore, supervisors tend to distort subjective performance ratings by not sufficiently differentiating good from bad performance in their ratings. Two forms of compression are centrality bias and leniency bias (Prendergast 1999). Centrality bias refers to a practice where superiors offer all employees ratings that differ little from a norm. Leniency bias implies that superiors simply overstate the performance of poor performers. Both biases are well documented in prior research (Bailey et al. 2011, Bol 2011, Chang et al. 2008, Moers 2005, Lipe and Salterio 2000, Jawahar and Williams 1997, Thompson 1990). For example, Moers (2005) shows that managers exhibit a centrality bias, such that differences in employees' performance are understated. Related experimental work suggests that in allocating discretionary bonus pools managers tend to allocate the bonus pool evenly or to rely too much on available accounting information (Bailey et al. 2011). Overall, these results show that managers regularly fail to obtain or fully consider all available information.

Bol (2011) further examines the determinants and performance effects of centrality bias and leniency bias. Her results indicate that managers tend to respond to their own incentives and preferences when subjectively evaluating performance. That is, they have incentives to keep the time and effort invested in the performance evaluation process at a minimum, circumvent conflicts, avoid harm to personal relationships, and restrict criticism (Bol 2011).

In the present study a dyad determines a production budget through a budget negotiation process in each period. Once the budget has been set the subordinate exerts effort on a task, the respective payoffs are determined and both parties receive information on subordinate production performance and their respective profit. This implies that (1) a superior's performance judgments might be affected by self-serving biases due to gains and losses in the budget negotiation process and that (2) a superior might also respond to his own compensation scheme (see equation 2) and evaluates performance in accordance with his own profit. However, from an ex-ante perspective the direction of possible

performance evaluation bias can't be predicted in this rich setting. Therefore, the following hypothesis is stated non-directionally:

H2: *A superior's perception of a subordinate's performance differs from the respective subordinate's performance.*

2.4 Effects of Negotiation Agreement and Subordinates' Performance on Subordinates' Reputation

Many economic relations are long-term relationships and reputation plays an important role in repeated interactions (Gächter and Falk 2002). Reputation is a 'characteristic or attribute ascribed to one person by another' (Wilson 1985, p. 27) and is based on an individual's performance and actions over a period of time (Dejong et al. 1985, Kreps and Wilson 1982). Gächter and Falk (2002) find for a gift-exchange game that reputation contributes to increased effort levels in a multi-period setting relative to a single-period setting. Accordingly, Tsui (1984) finds that the most reputationally effective managers tend to be more successful in their careers than the least reputationally effective managers.

In this study a superior and a subordinate interact repeatedly for a finite number of periods. In theory, a subordinate's performance should be positively associated with a subordinate's reputation (Tsui 1984). Thus, in a budget negotiation, being assessed by a superior, a high performing subordinate should receive a high reputation, whereas a low performing subordinate should receive a low reputation. However, previous empirical studies have shown that strategic behavior such as reciprocity is likely to arise when participants interact over multiple periods (Dopuch et al. 2001), which is attributable to expectation formation and reputation building (Schatzberg and Stevens 2008). Also, Maas et al. (2012) show that social preferences affect the incorporation of information into subjective performance evaluations. These findings indicate that reputation assessments can be affected by managers' social preferences.

The assumption that a supervisor has both, an intrinsic preference for accurately reporting a subordinate's performance and cares for the welfare of a subordinate, leads to a basic trade-off between accuracy and leniency in subjective performance evaluations (Prendergast 2002, Prendergast and Topel 1993). Evaluations are the more lenient, the stronger the supervisor's social preferences are towards the evaluated subordinate. Breuer et al. (2013) find evidence for this trade-off and show that a subjective performance evaluation is biased when there is a close social proximity between a supervisor

and subordinates. In their study, employees in large groups receive worse subjective performance evaluations than employees in smaller groups, i.e. closer social ties lead to a leniency bias.

Austin (1980) examines individuals involved in a continuous relationship (college roommates) and strangers that distribute rewards for a task under conditions of either high or low performance. He predicts and finds that roommates overlook differences in performance and choose an equality distribution rule, whereas strangers choose equality when performance is low and merit when performance is high. The results suggest that a closer social proximity leads to a leniency bias with respect to distribution of rewards.

In budget negotiations, social proximity, i.e. the strength of the personal relationship between supervisor and subordinate, is assumed to increase with negotiation agreement. Negotiation agreement indicates cooperative behavior by both the superior and subordinate (Pruitt and Carnevale 1993). Also, people experience positive feelings when they reach an agreement (Izard 1977, Deci 1975) and repeated agreements lead actors to attribute their positive feelings to the relation (Lawler 1992). Furthermore, once established, cooperative relationships tend to persist (Pruitt and Carnevale 1993) and superiors and subordinates who expect future negotiations act more cooperatively (Fisher et al. 2006).

This suggests that in budget negotiations with repeated interactions an increasing rate of negotiation agreement is expected to be positively related to social proximity which leads to a trade-off between accuracy and leniency with respect to a superior's assessment of a subordinate's reputation. The following hypothesis can be stated:

H3: *Given high agreement, a subordinate's performance does not influence a subordinate's reputation; given low agreement, a subordinate's performance has a positive effect on a subordinate's reputation.*

3 Method and Design

3.1 Experimental Design

To analyze the effects of initial budget proposal and final authority on subordinates' performance and perceived subordinates' performance from a superior's perspective, this study employs a 2×2 (between-subjects) ×3 (within-subjects) full factorial mixed experimental design. Final authority and initial budget proposal are *manipulated* between-subjects. The number of periods is *manipulated* within-subjects. To examine the

effects of subordinates' performance and negotiation agreement on subordinates' reputation, this study utilizes a 2×2 (between-subjects) ×3 (within-subjects) full factorial mixed experimental design. Subordinates' performance and negotiation agreement are *measured* between-subjects. The number of periods is *manipulated* within-subjects. Both experimental designs are based on the same subject pool.

3.2 Independent Variables

Final Authority is a manipulated independent between-subjects variable. It defines which party determines the budget level in case of a negotiation impasse. Either the superior has final authority (*Superior Authority*), or the subordinate has final authority (*Subordinate Authority*) (Fisher et al. 2000).

Initial Budget Proposal is a manipulated independent between-subjects variable. It determines which party makes the initial budget proposal in the budget negotiation. Either the superior makes the initial budget proposal (*Superior Proposal*), or the subordinate makes the initial budget proposal (*Subordinate Proposal*) (Fisher et al. 2000).

Agreement is a measured independent between-subjects variable. In each period *Agreement* equals 0 for a dyad when a budget negotiation ends in an impasse and 1 when it ends in an agreement. With three periods in the multi-period setting, *Agreement* is defined as the fraction of negotiation agreements relative to the maximum amount of possible negotiation agreements, and can thus take values of 0, 1/3, 2/3 or 1. To classify high agreement versus low agreement, a dummy variable is specified for agreement that equals 0 for *Low Agreement* and 1 for *High Agreement*, based on a median split of the measured variable *Agreement* (median = 2/3).

Performer is a measured independent between-subjects variable. *Performer* is defined as the average number of correct decodings performed by a subordinate during the three periods of the multi-period setting. To classify high versus low performers, a dummy variable is specified for performance that equals 0 for *Low Performer* and 1 for *High Performer*, based on a median split of the measured performance of a subordinate averaged on three periods (median = 57.17).

Period is a manipulated independent within-subjects variable. There are three periods in the multi-period setting.

3.3 Dependent Variables

Performance as a dependent variable is the number of correct decodings performed by a subordinate in the experimental production task in one period (Fisher et al. 2000).

Perceived Performance as a dependent variable is determined by asking a superior from a dyad in the postexperimental questionnaire to rate the performance of his respective subordinate. Superiors are asked to respond to the statement: 'I am satisfied with the performance of my subordinate' on a seven-point Likert scale anchored by '1 = strongly disagree' and '7 = strongly agree'.

Reputation as a dependent variable is determined by asking a superior of a dyad in the postexperimental questionnaire to rate the reputation of his respective subordinate. Superiors are asked to respond to the statement 'How high would you describe the reputation of your subordinate?' on a seven-point Likert scale anchored by '1 = absolutely no reputation' and '7 = very high reputation'.

3.4 Participants

In total, 120 students from a public university in Germany participated in the study. The mean age of participants was 22.53 years, with the youngest participants being 19 years old and the oldest participant being 32 years old. There were 90 female participants and 30 male participants. Superiors and subordinates were paired anonymously, creating a total of 60 dyads. As in other studies, which examine outcomes in participative budget negotiations (Fisher et al. 2006, 2002a, 2002b, 2000), the participants were undergraduates and graduate students. Students self-registered as participants in response to an invitation mail from the university's laboratory. Participants took part in a computer-based laboratory experiment programmed and conducted with the software z-Tree (Fischbacher 2007). The laboratory consists of an experimenter cubicle and 20 cubicles, each with a computer. On average one session took 75 minutes.

3.5 Experimental Procedures

Experimental procedures have been adopted largely by Fisher et al. (2006, 2000) and consist of the following steps:

1. Participants gather in front of the experimental laboratory and draw a random number. This number corresponds to a respective cubicle in the laboratory and ensures a random assignment. When all participants are present, they enter the laboratory. Each

subject sits down at the assigned cubicle equipped with a laptop as well as a folder with the required decryption keys for the decode production task. Furthermore, printed copies of the instructions, a sheet of paper, and a pen are available. Cubicles are separated by partitions in order to disable communication between subjects. This controls for interpersonal factors and enables an anonymous budget negotiation.

2. After all subjects have entered the cubicles, the instructions are shown on the computer screen. Instructions tell participants not to communicate during the experiment, that anonymity is assured and that, if any questions come up, the experimenter can be addressed. In order to perceive themselves as subordinates or superiors, the experiment is framed in a firm context for participants. Participants receive information that they will work at their current workplace for one period. After the instructions the real-effort production task is explained. The production task is to decode letters into two-digit numbers. The number of correct decodings determines a subordinate's performance. This decode task is similar to studies by Fisher et al. (2006, 2002a, 2002b, 2000), Stevens (2002), and Chow et al. (1988). Subsequently, each participant performs a 3-minute trial of the decode task, to ensure that the production task is understood (Fisher et al. 2000). Finally, each participant is told whether he acts as a subordinate or as a superior during the experiment.

3. Subordinates perform three 3-minute practice rounds to assess their performance capabilities. In each round each correct decoding is rewarded with 0.08 L€ (Laboreuro). All monetary amounts are denoted in this experimental currency. One L€ equals 45 Eurocent. Subordinates receive information about the number of correct, incorrect and total decodings after each practice round. Superiors also perform three 3-minute practice rounds to familiarize themselves with the task. However, to ensure that precise knowledge of their performance capabilities does not influence the budget negotiation process, superiors are neither paid nor informed about the number of correct, incorrect, and total decodings after each practice round (Fisher et al. 2000).

4. The two compensation schemes (Equation 1 and 2) are explained and illustrated using numerical examples. The parameters are fixed with $F = 1.00$ L€, $A = 0.05$ L€, $D = 0.08$ L€ and $R = 100\%$. In addition, it is explained that a superior is endowed with 1.00 L € in each round to cover the fixed wage of the subordinate.

5. The budget negotiation process is described. Participants are informed which party makes the initial budget proposal and which subject has final authority. Moreover,

it is explained that in a given period the budget negotiation consists of 4 rounds and each round consists of one offer and one counteroffer. If a dyad agrees on a budget during the negotiation, i.e. in a given round an offer matches a counteroffer, the budget is determined jointly by agreement. If there is a negotiation impasse at the end of the budget negotiation process, either the superior or the subordinate determines the budget, depending on who has final authority.

6. To ensure that participants understand the instructions, forced manipulation checks of a general understanding of the compensation schemes and the budget negotiation process are performed. If a question is answered incorrectly, a participant is told the correct answer.

7. Superiors and subordinates receive information on the average performance of subordinates. Information on average performance is taken from a pretest. On average 50.0 letters were decoded correctly in the pretest.

8. Subordinates have to assess how many correct decodings they can perform in three minutes. This information is used to measure subordinates' performance capability. Subordinates are reassured that the superior receives no information about their self-assessment. Nevertheless, subordinates are again asked, if they have given their self-assessment truthfully. If they answer with no, they may adjust their self-assessment. After the self-assessment, the budget negotiation begins. The initial budget proposal and a counteroffer are entered. An offer and a counteroffer constitute one round, with four rounds in total in one period. If an offer and a counteroffer match, this indicates agreement, the budget is set accordingly, and the negotiation ends. If a negotiation dyad fails to reach agreement within the four rounds of negotiation, the party who has final authority sets the budget. After the budget has been determined, the subordinate performs the production task. Finally both parties are informed about the amount of correct, incorrect and total decodings as well as their respective payoffs.

9. Participants are told that their workplace in the firm has changed and thus they interact with a new subject for three periods. This constitutes the multi-period setting with repeated interactions. Participants are informed that their role as superior or subordinate as well as final authority and initial budget proposal do not change and they remain the same during the three periods. Before the first budget negotiation of the multi-period setting begins, participants are asked manipulation checks about a general under-

standing of the new situation. Then, the new dyads interact for three periods, i.e. step 8 (above) is repeated three times (Fisher et al. 2006).

10. Subjects have to answer a postexperimental questionnaire. Finally, participants leave the experimental laboratory one by one and receive their respective payoff.

4 Results

4.1 Manipulation Checks

After the last practice round and after the single-period setting, participants are asked several questions regarding a general understanding of the experimental task and payoff. No participant can proceed unless all questions are answered correctly. If an answer is wrong, a help box appears on the screen explaining the correct answer. All participants succeed to proceed. Also, general task understanding is analyzed with the postexperimental statements: ‘I knew, what I had to do’ and ‘I did not understand something’. The answer categories to both questions are on a seven-point Likert scale anchored by ‘1 = strongly disagree’ and ‘7 = strongly agree’. The median value with respect to the first question is 7 and the median value with respect to the second question is 2. This indicates that participants had a high understanding of the experimental task on average.

4.2 Descriptive Statistics

To analyze the effects of *Final Authority* and *Initial Budget Proposal* on *Performance* and *Perceived Performance*, experimental cells are classified into four categories based upon (1) whether a superior or a subordinate has final authority in the event of a negotiation impasse and (2) whether a superior or a subordinate makes the initial budget proposal in the budget negotiation. Descriptive statistics for *Initial Proposal*⁵, *Budget*⁶, $\Delta(\text{Initial} - \text{Budget})$ ⁷, *Performance*⁸, *Perceived Performance* and *Superior Profit*⁹ are reported in Table 1, Panel A.

⁵ *Initial Proposal* as a dependent variable is the initial budget proposal in a given period. Panel A in Table 1 shows the average value of Initial Proposal across the three periods in the multi-period setting.

⁶ *Budget* as a dependent variable is the final budget in a given period. Panel A in Table 1 shows the average value of *Budget* across the three periods in the multi-period setting.

⁷ $\Delta(\text{Initial} - \text{Budget})$ is the difference between *Initial Proposal* and *Budget*. A positive value for $\Delta(\text{Initial} - \text{Budget})$ indicates that the final budget is lower than the initial budget proposal; a negative value indicates that the final budget is higher than the initial budget proposal. Panel A in Table 1 shows the average value of $\Delta(\text{Initial} - \text{Budget})$ across the three periods in the multi-period setting.

⁸ Panel A and Panel B in Table 1 show the average value of *Performance* across the three periods in the multi-period setting.

Table 1: Descriptive Statistics

Panel A: Descriptive Statistics for Initial Proposal^a, Budget^b, $\Delta(\text{Initial} - \text{Budget})^c$, Performance^d, Perceived Performance^e, and Superior Profit^f; Mean; [Standard Deviation]; (Number of Observations)				
Initial Budget Proposal^h	Dependent Variable	Final Authority^g		Means
		Superior Authority	Subordinate Authority	
Superior Proposal	Initial Proposal	58.44 [13.59] (n = 45)	52.73 [15.16] (n = 45)	55.59 [14.60] (n = 90)
	Budget	51.82 [12.21] (n = 45)	32.44 [23.56] (n = 45)	42.13 [21.05] (n = 90)
	$\Delta(\text{Initial} - \text{Budget})$	6.62 [7.78] (n = 45)	20.29 [21.55] (n = 45)	13.46 [17.51] (n = 90)
	Performance	56.80 [12.59] (n = 45)	56.82 [10.53] (n = 45)	56.81 [11.54] (n = 90)
	Perceived Performance	6.07 [1.75] (n = 15)	3.80 [2.21] (n = 15)	4.93 [2.27] (n = 30)
	Superior Profit	4.18 [1.04] (n = 45)	3.28 [1.22] (n = 45)	3.73 [1.21] (n = 90)
Subordinate Proposal	Initial Proposal	35.42 [13.98] (n = 45)	34.16 [16.58] (n = 45)	34.79 [15.26] (n = 90)
	Budget	54.98 [12.07] (n = 45)	38.51 [17.65] (n = 45)	46.74 [17.16] (n = 90)
	$\Delta(\text{Initial} - \text{Budget})$	-19.56 [12.86] (n = 45)	-4.36 [8.97] (n = 45)	-11.96 [13.41] (n = 90)
	Performance	47.58 [21.05] (n = 45)	61.02 [7.42] (n = 45)	54.30 [17.09] (n = 90)
	Perceived Performance	4.80 [2.11] (n = 15)	4.60 [1.80] (n = 15)	4.70 [1.93] (n = 30)
	Superior Profit	3.57 [1.54] (n = 45)	3.76 [.79] (n = 45)	3.66 [1.22] (n = 90)
Means	Initial Proposal	46.93 [17.94] (n = 90)	43.44 [18.35] (n = 90)	45.19 [18.18] (n = 180)
	Budget	53.40 [12.18] (n = 90)	35.48 [20.92] (n = 90)	44.44 [19.29] (n = 180)
	$\Delta(\text{Initial} - \text{Budget})$	-6.47 [16.88] (n = 90)	7.97 [20.56] (n = 90)	.75 [20.11] (n = 180)
	Performance	52.19 [17.86] (n = 90)	58.92 [9.30] (n = 90)	55.56 [14.60] (n = 180)
	Perceived Performance	5.43 [2.01] (n = 30)	4.20 [2.02] (n = 30)	4.82 [2.10] (n = 60)
	Superior Profit	3.87 [1.34] (n = 90)	3.52 [1.05] (n = 90)	3.69 [1.22] (n = 180)

Panel B: Descriptive Statistics for Performance^d, and Reputationⁱ; Mean; [Standard Deviation]; (Number of Observations)				
Agreement^k	Dependent Variable	Performer^j		Means
		Low Performer	High Performer	
Low Agreement	Performance	43.52 [21.78] (n = 42)	64.21 [7.15] (n = 39)	53.48 [19.38] (n = 81)
	Reputation	3.29 [1.49] (n = 14)	4.85 [1.41] (n = 13)	4.04 [1.63] (n = 27)
High Agreement	Performance	51.10 [7.69] (n = 48)	63.04 [4.80] (n = 51)	57.25 [8.72] (n = 99)
	Reputation	5.00 [1.10] (n = 16)	5.00 [1.62] (n = 17)	5.00 [1.37] (n = 33)
Means	Performance	47.57 [16.26] (n = 90)	63.54 [5.93] (n = 90)	55.56 [14.60] (n = 180)
	Reputation	4.20 [1.54] (n = 30)	4.93 [1.51] (n = 30)	4.93 [1.51] (n = 60)

^a *Initial Proposal* as a dependent variable is the initial budget proposal in a given period. Panel A shows the average value of *Initial Proposal* across the three periods in the multi-period setting.

^b *Budget* as a dependent variable is the final budget in a given period. Panel A shows the average value of *Budget* across the three periods in the multi-period setting.

⁹ *Superior Profit* as a dependent variable is the profit of a superior in a given period measured in Laboreuro. Panel A in Table 1 shows the average value of *Superior Profit* across the three periods in the multi-period setting.

Table 1: Descriptive Statistics cont.

<p>^b <i>Budget</i> as a dependent variable is the final budget in a given period. Panel A shows the average value of <i>Budget</i> across the three periods in the multi-period setting.</p> <p>^c $\Delta(\text{Initial} - \text{Budget})$ as a dependent variable is the difference between the initial budget proposal and the final budget in a given period. A positive value indicates that the final budget is lower than the initial budget proposal; a negative value indicates that the final budget is higher than the initial budget proposal. Panel A shows the average value of $\Delta(\text{Initial} - \text{Budget})$ across the three periods in the multi-period setting.</p> <p>^d <i>Performance</i> as a dependent variable is the number of correct decodings performed by a subordinate in the experimental production task in one period. Panel A and Panel B show the average value of <i>Performance</i> across the three periods in the multi-period setting.</p> <p>^e <i>Perceived Performance</i> as a dependent variable is determined by asking a superior from a dyad in the postexperimental questionnaire to rate the performance of his subordinate. Superiors are asked to respond to the statement: 'I am satisfied with the performance of my subordinate' on a seven-point Likert scale anchored by 'strongly disagree' and 'strongly agree'.</p> <p>^f <i>Superior Profit</i> as a dependent variable is the profit of a superior in a given period measured in Laboreuro. Panel A shows the average value of <i>Superior Profit</i> across the three periods in the multi-period setting.</p> <p>^g <i>Final Authority</i> is a manipulated independent between-subjects variable. It defines which party determines the budget level in case of a negotiation impasse. This variable has two factor levels: (1) superior has final authority (<i>Superior Authority</i>) and (2) subordinate has final authority (<i>Subordinate Authority</i>).</p> <p>^h <i>Initial Budget Proposal</i> is a manipulated independent between-subjects variable. It determines which party makes the initial budget proposal in the budget negotiation. This variable has two factor levels: (1) superior makes the initial budget proposal (<i>Superior Proposal</i>) and (2) subordinate makes the initial budget proposal (<i>Subordinate Proposal</i>).</p> <p>ⁱ <i>Reputation</i> as a dependent variable is determined by asking the superior from a dyad in the postexperimental questionnaire to rate the reputation of his subordinate. Superiors are asked to respond to the statement 'How high would you describe the reputation of your subordinate?' on a seven-point Likert scale anchored by 'absolutely no reputation' and 'very high reputation'.</p> <p>^j <i>Performer</i> is a measured independent between-subjects variable. Performer is defined as the average number of correct decodes performed by a subordinate during the three periods of the multi-period setting. To classify high versus low performance, a dummy variable is specified for performance that equals 0 for <i>Low Performer</i> and 1 for <i>High Performer</i>, based on a median split of performance averaged across the three periods in the multi-period setting (median = 57.17).</p> <p>^k <i>Agreement</i> is a measured independent between-subjects variable. In each period Agreement equals 0 when the budget negotiation ends in an impasse and 1 when the budget negotiation ends in agreement. With three periods in the multi-period setting, Agreement is defined as the fraction of negotiation agreements relative to the maximum amount of possible negotiation agreements, and can take values of 0, 1/3, 2/3 or 1. To classify high agreement versus low agreement, a dummy variable is specified for agreement that equals 0 for <i>Low Agreement</i> and 1 for <i>High Agreement</i>, based on a median split of the measured variable Agreement (median = 2/3).</p>
--

Also, to examine the effects of *Agreement* and *Performer* on *Reputation* and *Performance*, experimental cells are classified into four categories based upon (1) whether a dyad has low agreement or high agreement in the budget negotiations, and (2) whether the respective subordinate is a high or low performer. Descriptive statistics for *Performance* (averaged on *Period*) and *Reputation* are reported in Table 1, Panel B.

4.3 Hypothesis Tests

The SPSS generalized linear mixed model procedure¹⁰ (GENLINMIXED) is used to analyze the effects of the independent between-subjects variables *Final Authority* and

¹⁰ The generalized linear mixed model extends the general linear model in order that the dependent variable is linearly related to the independent factors and covariates via a specified link function. The generalized linear model covers widely used statistical models, such as linear regression and accordingly analysis of variance (ANOVA) for normally distributed responses, plus many other statistical models through its very general model formulation (IBM Corporation 2012).

Initial Budget Proposal, the independent within-subjects variable *Period* on *Performance*, the effects of the independent between-subjects variables *Final Authority* and *Initial Budget Proposal* on *Perceived Performance*. Specifically, four unique between-subjects cell means can be created for each *Period* – *Final Authority* with *Superior Authority* and *Subordinate Authority*, *Initial Budget Proposal* with *Superior Proposal* and *Subordinate Proposal*.

Because generalized linear mixed models are overall tests of effects of independent variables, additional follow up tests are performed to better understand the effects of *Final Authority* and *Initial Budget Proposal*. The descriptive statistics are depicted in Table 1, Panel A. The generalized linear mixed models as overall tests of the effects are shown in Table 2, Panel A and the detailed mean by mean comparisons are reported in Table 3, Panel A1 and Panel A2.

Effects of Final Authority and Initial Budget Proposal on Subordinates' Performance

A generalized linear model analysis using *Performance* as dependent variable is performed (Panel A in Table 2). An identity link-function, a robust compound symmetry covariance matrix estimator for repeated measures, and Satterthwaite (1946) degrees of freedom are specified. *Final Authority*, *Initial Budget Proposal* and *Period* are specified as independent variables. *Agree*¹¹ and *Capability*¹² are specified as covariates to control for fairness effects of negotiation agreement and subordinates' performance capabilities (Fisher et al. 2002a).

There is a significant main effect of *Final Authority* on *Performance* ($F = 5.79$; $p = .028$), no significant main effect of *Initial Budget Proposal* on *Performance* ($F = 1.83$; $p = .196$), and a marginally significant interaction effect between *Final Authority* and *Initial Budget Proposal* ($F = 3.32$; $p = .095$). There is no significant main effect of *Period* on *Performance* ($F = .15$; $p = .866$), no significant interaction effect between *Final Authority* and *Period* ($F = 1.27$; $p = .333$), no significant interaction effect between *Initial Budget Proposal* and *Period* ($F = .22$; $p = .804$), and no significant interaction effect between *Final Authority*, *Initial Budget Proposal*, and *Period* ($F = 1.37$; $p = .307$).

¹¹ *Agree* is a measured independent between-subjects variable. *Agree* is a dummy variable, in each period *Agree* equals 0 when the budget negotiation ends in impasse and 1 when the budget negotiation ends in agreement.

¹² *Capability* is a measured independent between-subjects variable. *Capability* measures subordinate performance capability in each period and is based on a subjects self-assessment of how many correct decodings he can perform in three minutes.

Table 2: Linear (Mixed) Models (GENLINMIXED)^a with Performance, Perceived Performance, and Reputation as Dependent Variables

Panel A – Performance and Perceived Performance							
Dependent Variable	Performance ^b			Perceived Performance ^c			
Source of Variation	df	F-Statistic	p-Value ^e	df	F-Statistic	p-Value ^e	
<i>Factors</i>							
Final Authority (F)	1	5.79	.028**	1	6.24	.015**	
Initial Budget Proposal (I)	1	1.83	.196	1	.13	.638	
F x I Interaction	1	3.32	.095*	1	3.11	.041**	
Period (P)	2	.15	.866	-	-	-	
F x P Interaction	2	1.27	.333	-	-	-	
I x P Interaction	2	.22	.804	-	-	-	
F x I x P Interaction	2	1.37	.307	-	-	-	
<i>Covariates^f</i>							
Agree	1	1.62	.215	-	-	-	
Capability	1	13.5	.006***	-	-	-	
			<i>N = 180</i>				<i>N = 60</i>
Panel B – Performance and Reputation							
Dependent Variable	Performance ^b			Reputation ^d			
Source of Variation	df	F-Statistic	p-Value ^e	df	F-Statistic	p-Value ^e	
<i>Factors</i>							
Agreement (A)	1	1.91	.185	1	6.94	.011**	
Performer (PF)	1	49.48	<.001***	1	4.84	.032**	
A x PF Interaction	1	3.56	.077*	1	4.84	.032**	
Period (P)	2	.97	.430	-	-	-	
A x P Interaction	2	.74	.516	-	-	-	
PF x P Interaction	2	.52	.620	-	-	-	
A x PF x P Interaction	2	.93	.446	-	-	-	
			<i>N = 180</i>				<i>N = 60</i>

^a The generalized linear model extends the general linear model in order that the dependent variable is linearly related to the independent factors and covariates via a specified link function. The generalized linear model covers widely used statistical models, such as linear regression or analysis of variance, plus many other statistical models through its very general model formulation (IBM Corporation 2012). For definitions of independent and dependent variables, see Table 1.

^b *Performance* as a dependent variable is assumed to have a normal distribution. The link function is specified as identity, so that $f(x) = x$. The dependent variable is not transformed (IBM Corporation 2012).

^c *Perceived Performance* as a dependent variable is assumed to have a normal distribution. The link function is specified as identity, so that $f(x) = x$. The dependent variable is not transformed (IBM Corporation 2012).

^d *Reputation* as a dependent variable is assumed to have a normal distribution. The link function is specified as identity, so that $f(x) = x$. The dependent variable is not transformed (IBM Corporation 2012).

^e ***, **, * Significance levels of 1 %, 5 % and 10 %.

^f *Agree* is a dummy variable, in each period *Agree* equals 0 when the budget negotiation ends in impasse and 1 when it ends in agreement. *Capability* measures a subordinate's performance capability in each period and is based on a subject's self-assessment of how many correct decodings he can perform in three minutes.

Table 3: Mean by Mean Comparisons

Panel A: Mean by Mean Comparisons for Final Authority and Initial Budget Proposal						
Panel A1: Mean by Mean Comparisons for Performance^a						
Contrast	Comparison^f	Treat- ment 1	Treat- ment 2	Std.Error	t-Statistic	p-Value^g
(1) Superior Proposal	Superior Authority vs. Subordinate Authority	56.79	56.80	2.48	-.00	.998
(2) Subordinate Proposal	Superior Authority vs. Subordinate Authority	49.71	58.39	3.77	-3.33	.014** ^h
(3) Superior Authority	Superior Proposal vs. Subordinate Proposal	56.79	49.71	4.11	1.73	.057* ^h
(4) Subordinate Authority	Superior Proposal vs. Subordinate Proposal	56.80	58.39	1.65	-.97	.335
Panel A2: Mean by Mean Comparisons for Perceived Performance^b						
Contrast	Comparison^f	Treat- ment 1	Treat- ment 2	Std.Error	t-Statistic	p-Value^g
(1) Superior Proposal	Superior Authority vs. Subordinate Authority	6.07	3.80	.70	3.22	.002***
(2) Subordinate Proposal	Superior Authority vs. Subordinate Authority	4.80	4.60	.69	.29	.774
(3) Superior Authority	Superior Proposal vs. Subordinate Proposal	6.07	4.80	.68	1.85	.069*
(4) Subordinate Authority	Superior Proposal vs. Subordinate Proposal	3.80	4.60	.71	-1.12	.266
Panel A3: Mean by Mean Comparisons for Superior Profit^c						
Contrast	Comparison^f	Treat- ment 1	Treat- ment 2	z-Statistic	p-Value^g	
(1) Superior Proposal	Superior Authority vs. Subordinate Authority	4.18	3.28	-3.75	<.001***	
(2) Subordinate Proposal	Superior Authority vs. Subordinate Authority	3.57	3.76	-1.07	.285	
(3) Superior Authority	Superior Proposal vs. Subordinate Proposal	4.18	3.57	-2.18	.029**	
(4) Subordinate Authority	Superior Proposal vs. Subordinate Proposal	3.28	3.76	-1.84	.065*	
Panel B: Mean by Mean Comparisons for Agreement and Performer						
Panel B1: Mean by Mean Comparisons for Performance^d						
Contrast	Comparison^f	Treat- ment 1	Treat- ment 2	Std.Error	t-Statistic	p-Value^g
(1) Low Agreement	Low Performer vs. High Performer	43.52	64.21	4.22	-4.90	<.001***
(2) High Agreement	Low Performer vs. High Performer	51.10	63.04	1.93	-6.20	<.001***
(3) Low Performer	Low Agreement vs. High Agreement	43.52	51.10	4.10	-1.85	.091*
(4) High Performer	Low Agreement vs. High Agreement	64.21	63.04	2.17	.54	.592
Panel B2: Mean by Mean Comparisons for Reputation^e						
Contrast	Comparison^f	Treat- ment 1	Treat- ment 2	Std.Error	t-Statistic	p-Value^g
(1) Low Agreement	Low Performer vs. High Performer	3.29	4.85	.54	-2.91	.003*** ^h
(2) High Agreement	Low Performer vs. High Performer	5.00	5.00	.46	.00	1.00
(3) Low Performer	Low Agreement vs. High Agreement	3.29	5.00	.47	-3.68	.001***
(4) High Performer	Low Agreement vs. High Agreement	4.85	5.00	.53	-0.29	.774

Table 3: Mean by Mean Comparisons cont.

- ^a Mean by mean comparisons for *Performance* are performed using simple contrasts based on a generalized linear model for *Performance* as a dependent variable with *Agree* and *Capability* as covariates as specified in Table 2. Tabulated means for *Performance* are estimated marginal means. The covariate for *performance capability* is fixed at the mean of 53.7.
- ^b Mean by mean comparisons for *Perceived Performance* are performed using simple contrasts based on a generalized linear model for *Perceived Performance* as a dependent variable as specified in Table 2. Tabulated means for *Perceived Performance* are estimated marginal means.
- ^c Mean by mean comparisons for *Superior Profit* are performed using non-parametric Mann-Whitney U tests.
- ^d Mean by mean comparisons for *Performance* are performed using simple contrasts based on a generalized linear model for *Performance* as a dependent variable as specified in Table 2. Tabulated means for *Performance* are estimated marginal means.
- ^e Mean by mean comparisons for *Reputation* are performed using simple contrasts based on a generalized linear model for *Reputation* as a dependent variable as specified in Table 2. Tabulated means for *Reputation* are estimated marginal means.
- ^f Treatment 1 refers to the left hand side of the comparison. Treatment 2 refers to the right hand side
- ^g ***, **, * represents two-tailed significance at the 1 %, 5 % or 10 % level; ***, **, *h represents one-tailed significance at the 1 %, 5 % or 10 % level.

To further analyze the interaction of *Final Authority* and *Initial Budget Proposal* on *Performance* (Panel A in Figure 1) mean by mean comparisons (Panel A1 in Table 3) are performed with contrast analysis (Buckless and Ravenscroft 1990). The covariate for performance capability (*Capability*) is fixed at the mean of 53.7. Given *Superior Proposal* a mean by mean comparison shows no significant difference between *Superior Authority* (mean *Performance* = 56.79) and *Subordinate Authority* (56.80) with respect to *Performance* ($t = -.00$; $p = .998$). Given *Subordinate Proposal* a mean by mean comparison shows a significant difference between *Superior Authority* (49.71) and *Subordinate Authority* (58.39) with respect to *Performance* ($t = -3.33$; $p = .014$; one-tailed). Given *Superior Authority* a mean by mean comparison shows a marginal significant difference between *Superior Proposal* (56.79) and *Subordinate Proposal* (49.71) with respect to *Performance* ($t = 1.73$; $p = .057$; one-tailed). Given *Subordinate Authority* a mean by mean comparison shows no significant difference between *Superior Proposal* (56.80) and *Subordinate Proposal* (58.39) with respect to *Performance* ($t = -.97$; $p = .335$).

If superiors make the initial budget proposal, there is no significant effect of final authority on subordinates' performance. If subordinates make the initial budget proposal, their performance is significantly higher when subordinates have final authority, compared to superiors having final authority. Generally, subordinates' performance is high, except for the situation where subordinates make the initial proposal and superiors have final authority. Thus, the results support H1.

Figure 1: Effects of Final Authority and Initial Budget Proposal on Performance, Perceived Performance, and Superior Profit^a

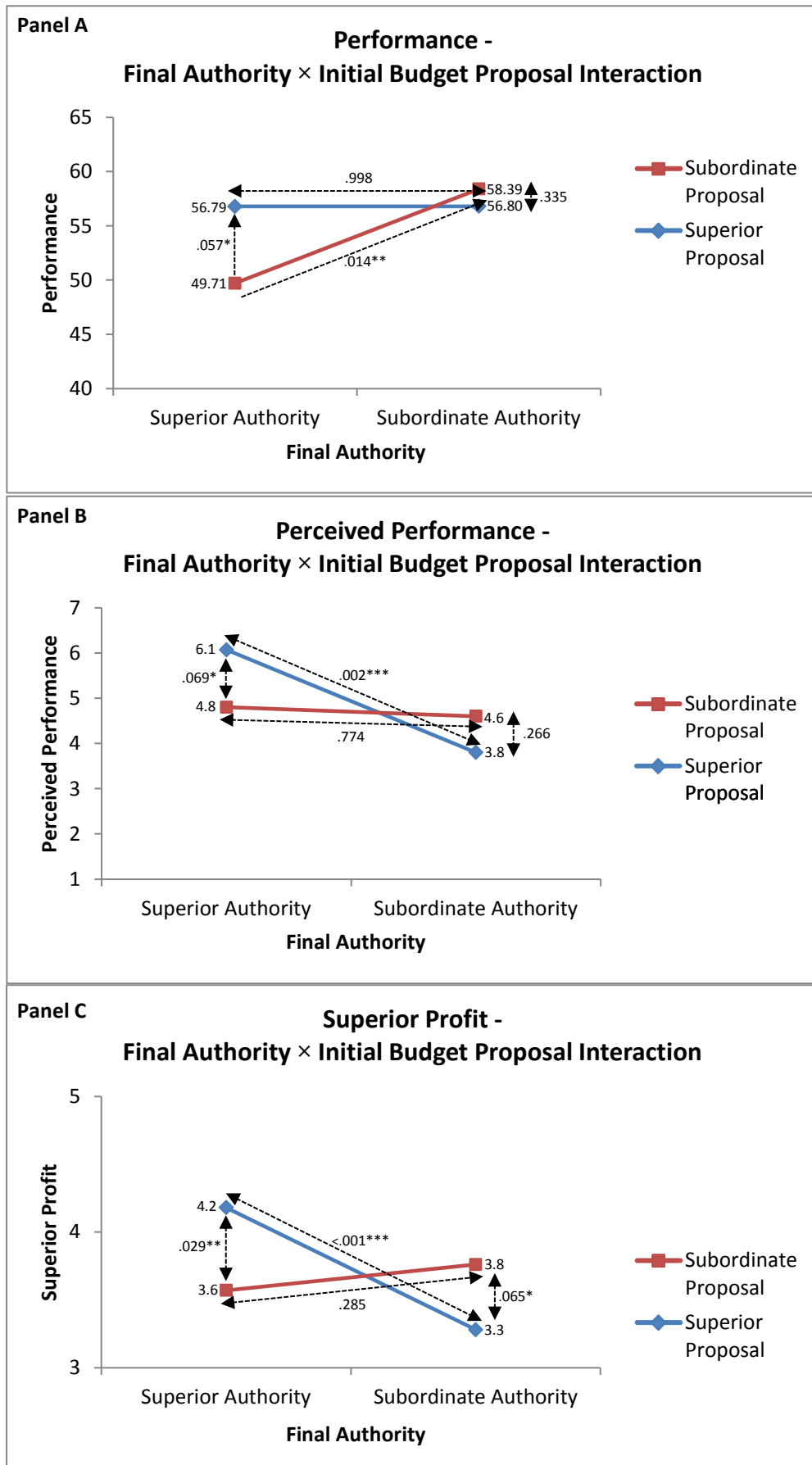


Figure 1: Effects of Final Authority and Initial Budget Proposal on Performance, Perceived Performance, and Superior Profit^a cont.

^a Variable definitions in Table 1. Detailed analyses of the *Final Authority* × *Initial Budget Proposal* interaction for *Performance* (Panel A), *Perceived Performance* (Panel B) and *Superior Profit* (Panel C). The horizontal axes represent *Superior Authority* and *Subordinate Authority* in case of a negotiation impasse. The full lines represent *Initial Budget Proposal* (*Superior Proposal*, *Subordinate Proposal*). The vertical axes display *Performance* (Panel A; the covariate for performance capability is fixed at the mean of 53.7), *Perceived Performance* (Panel B) and *Superior Profit* (Panel C). One-sided dotted arrows and associated figures present the direction and the p-value of a one-sided mean by mean comparison. Two-sided dotted arrows and associated figures present the p-value of a two-sided mean by mean comparison. ***, **, * indicate significance levels of 1 %, 5 % or 10 %. Mean by mean comparisons are based on analyses in Table 3.

Given *Superior Authority* it is further analyzed whether the difference between initial budget proposal and final budget $\Delta(\text{Initial} - \text{Budget})$, is negative on average for *Subordinate Proposal*, indicating an unfair loss for subordinates. Furthermore, it is positive on average for *Superior Proposal*, indicating a fair gain for subordinates. Given *Superior Authority* a Wilcoxon signed-rank test indicates that $\Delta(\text{Initial} - \text{Budget})$ is significantly negative for *Subordinate Proposal* (mean = -19.56; $z = -5.76$; $p < .001$) and significantly positive for *Superior Proposal* (mean = 6.62; $z = 4.93$; $p < .001$). These results are consistent with concessionary behavior during budget negotiations (Fisher et al. 2000). From a subordinate's perspective, the negative sign for *Subordinate Authority* supports the notion of being perceived as an unfair loss, whereas the positive sign for *Superior Authority* supports the notion of being perceived as a fair gain.

Given *Superior Authority* descriptive statistics (Panel A in Table 1) indicate that the budget levels for *Superior Proposal* (mean = 51.82) and *Subordinate Proposal* (54.98) are close to the grand mean of *Performance* (55.56) and can thus be considered as high budget levels. Also, given *Subordinate Authority* it is additionally analyzed whether *Budget* is positive on average for *Subordinate Proposal* and *Superior Proposal*, indicating subordinates' fairness concerns for superiors. Given *Subordinate Authority* a Wilcoxon signed-rank test indicates that *Budget* is significantly positive for both *Subordinate Proposal* (mean = 38.51; $z = 5.85$; $p < .001$) and *Superior Proposal* (mean = 32.44; $z = 5.72$; $p < .001$). These results are consistent with subordinates having reciprocal fairness concerns for superiors (Fisher et al. 2000).

Superiors' Perception of Subordinates' Performance

A generalized linear model analysis using *Perceived Performance* as the dependent variable is performed (Panel A in Table 2). An identity link-function, a robust covari-

ance matrix estimator, and Satterthwaite (1946) degrees of freedom are specified. *Final Authority* and *Initial Budget Proposal* are specified as independent variables.

There is a significant main effect of *Final Authority* on *Perceived Performance* ($F = 6.24$; $p = .015$), no significant main effect of *Initial Budget Proposal* on *Perceived Performance* ($F = .13$; $p = .638$), and a significant interaction effect between *Final Authority* and *Initial Budget Proposal* ($F = 3.11$; $p = .041$).

To further analyze the interaction effect of *Final Authority* and *Initial Budget Proposal* on *Perceived Performance* (Panel B in Figure 1) mean by mean comparisons are performed with contrast analysis (Panel A2 in Table 3). Given *Superior Proposal* a mean by mean comparison shows a significant difference between *Superior Authority* (mean *Perceived Performance* = 6.07) and *Subordinate Authority* (3.80) with respect to *Perceived Performance* ($t = 3.22$; $p = .002$). Given *Subordinate Proposal* a mean by mean comparison shows no significant difference between *Superior Authority* (4.80) and *Subordinate Authority* (4.60) with respect to *Perceived Performance* ($t = .29$; $p = .774$).

In line with the results for *Performance*, these results indicate that superiors' performance perception differs from subordinates' performance. Given *Superior Proposal*, *Performance* does not differ significantly between *Superior Authority* and *Subordinate Authority* (Panel A in Figure 1). However, given *Superior Proposal*, *Perceived Performance* is significantly higher for *Superior Authority* compared to *Subordinate Authority* (Panel B in Figure 1). Therefore, although on average *Performance* does not differ across these conditions, superiors perceive performance to be higher, when they have final authority. Also, given *Subordinate Proposal*, *Performance* is significantly lower for *Superior Authority* compared to *Subordinate Authority* (Panel A in Figure 1). However, given *Subordinate Proposal*, *Perceived Performance* does not differ significantly between *Superior Authority* and *Subordinate Authority* (Panel B in Figure 1). Again, although on average *Performance* differs across these conditions, superiors rate performance equally, when subordinates have final authority. Overall, these results support H2.

Superiors receive information on subordinates' performance as well as their own payoff. Bol (2011) shows that managers tend to respond to their incentives and preferences when subjectively evaluating performance. In line with this, Pearson product-moment correlation coefficient indicates a strong relationship between *Perceived Per-*

formance and *Superior Profit* ($r = .607$; $p < .001$). Therefore an exploratory analysis based on non-parametric Mann-Whitney U tests is conducted to show how *Perceived Performance* is related to *Superior Profit* (Panel A3 in Table 3). Given *Superior Proposal*, *Perceived Performance* is significantly higher for *Superior Authority* (mean *Superior Profit* in Laboreuro = 4.18) compared to *Subordinate Authority* (3.28) (Mann-Whitney U; $z = -3.75$; $p < .001$). Given *Subordinate Proposal*, *Perceived Performance* does not differ significantly between *Superior Authority* (3.57) and *Subordinate Authority* (3.76) (Mann-Whitney U; $z = -1.07$; $p = .285$). This pattern parallels the results for *Perceived Performance* as described above (see Panel B and C in Figure 1) and can be interpreted as evidence that superiors respond to their own incentives when evaluating subordinate performance.

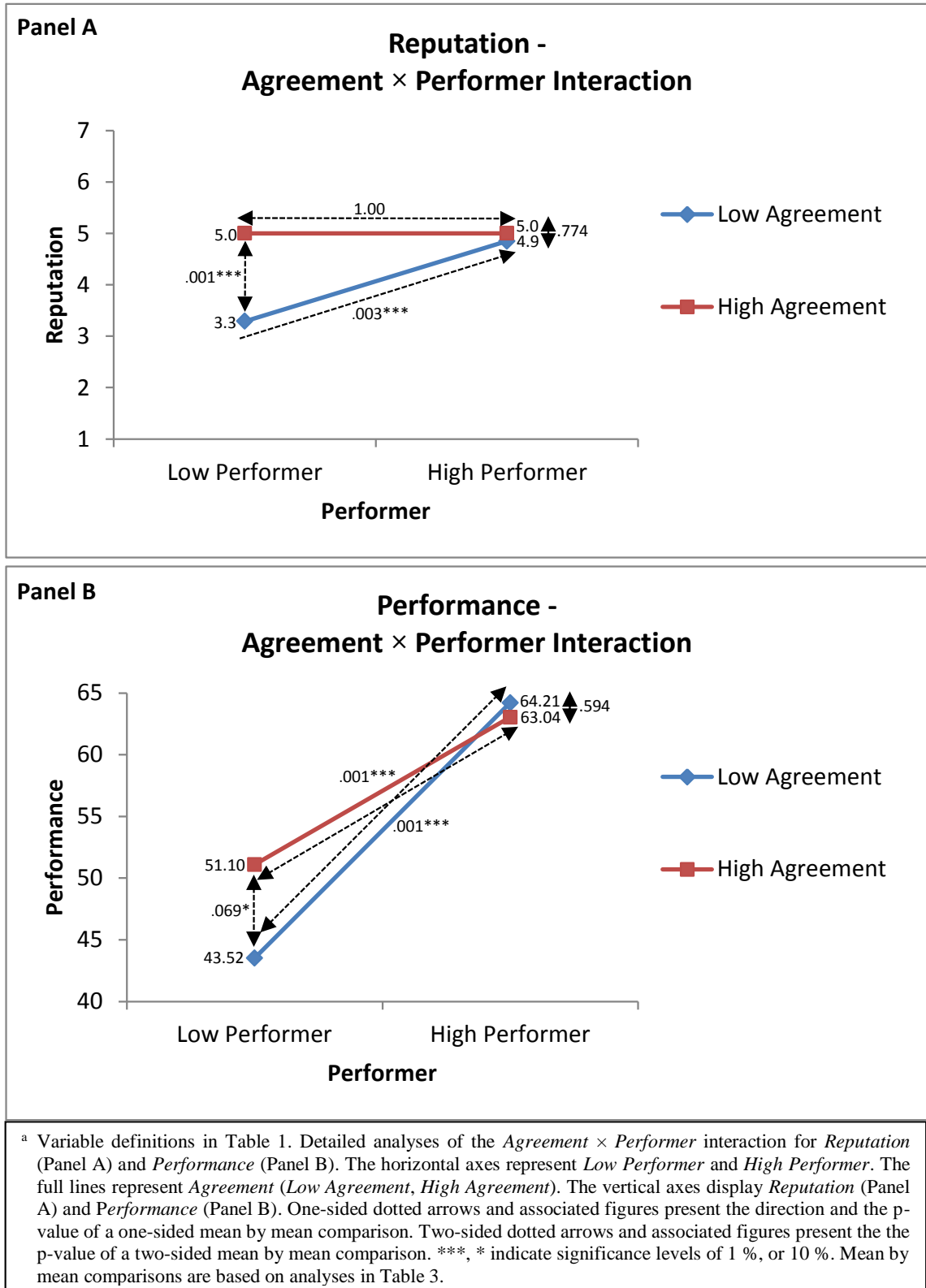
Effects of Negotiation Agreement and Subordinates' Performance on Subordinates' Reputation

The SPSS generalized linear mixed model procedure (GENLINMIXED) is used to analyze the effects of the independent between-subjects variables *Agreement* and *Performer* and the independent within-subjects variable *Period* on *Performance* and the effects of the independent between-subjects variables *Agreement* and *Performer* on *Reputation*. Specifically, four unique between-subjects cell means can be created for each *Period* – *Agreement* with *Low Agreement* and *High Agreement*, and *Performer* with *Low Performer* and *High Performer*. Additional follow up tests are performed to better understand the effects of *Agreement* and *Performer*. The generalized linear mixed models as overall tests of the effects are shown in Table 2, Panel B and the detailed mean by mean comparisons are reported in Table 3, Panel B1 and Panel B2.

Performance

Before examining the effects of *Agreement* and *Performer* on *Reputation*, it is analyzed whether *Low Performers* differ significantly from *High Performers*. Therefore, a generalized linear model analysis using *Performance* as the dependent variable is performed (Panel B in Table 2). An identity link-function, a robust compound symmetry covariance matrix estimator for repeated measures, and Satterthwaite (1946) degrees of freedom are specified. *Agreement*, *Performer* and *Period* are specified as independent variables.

Figure 2: Effects of Agreement and Performer on Reputation and Performance^a



There is no significant main effect of *Agreement* on *Performance* ($F = 1.91$; $p = .185$), a significant main effect of *Performer* on *Performance* ($F = 49.48$; $p < .001$), and a marginally significant interaction effect between *Agreement* and *Performer* ($F = 3.56$;

$p = .077$). There is no significant main effect of *Period* on *Performance* ($F = .97$; $p = .430$), no significant interaction effect between *Agreement* and *Period* ($F = .74$; $p = .516$), no significant interaction effect between *Performer* and *Period* ($F = .52$; $p = .620$) and no significant interaction effect between *Agreement*, *Performer*, and *Period* ($F = .93$; $p = .446$).

To further analyze the interaction effect of *Agreement* and *Performer* on *Performance* (Panel B in Figure 2) mean by mean comparisons (Panel B1 in Table 3) are performed with contrast analysis. Given *Low Agreement* a mean by mean comparison shows a significant difference between *Low Performer* (mean *Performance* = 43.52) and *High Performer* (64.21) with respect to *Performance* ($t = -4.90$; $p < .001$). Given *High Agreement* a mean by mean comparison shows a significant difference between *Low Performer* (51.10) and *High Performer* (63.04) with respect to *Performance* ($t = -6.20$; $p < .001$). Results show that low performers have a significantly lower performance than high performers.

Also, it is analyzed whether low performers and high performers react differently to negotiation agreement (Panel B1 in Table 3). Given *Low Performer* a mean by mean comparison shows a marginally significant difference between *Low Agreement* (mean *Performance* = 43.52) and *High Agreement* (51.10) with respect to *Performance* ($t = -1.85$; $p = .091$). Given *High Performer* a mean by mean comparison shows no significant difference between *Low Agreement* (64.21) and *High Agreement* (63.04) with respect to *Performance* ($t = .54$; $p = .592$). This indicates that only low performers perform significantly lower, when budget negotiation ends with an impasse. High performers perform high, irrespective of budget negotiation outcome.

Reputation

A generalized linear model analysis using *Reputation* as the dependent variable is performed (Panel B in Table 2). An identity link-function, a robust covariance matrix estimator, and Satterthwaite (1946) degrees of freedom are specified. *Agreement* and *Performer* are specified as independent variables.

There is a significant main effect of *Agreement* on *Reputation* ($F = 6.94$; $p = .011$), a significant main effect of *Performer* on *Reputation* ($F = 4.84$; $p = .032$), and a significant interaction effect between *Agreement* and *Performer* ($F = 4.84$; $p = .032$).

To further analyze the interaction effect between *Agreement* and *Performer* on *Reputation* (Panel B in Figure 2) mean by mean comparisons (Panel B2 in Table 3) are performed with contrast analysis. Given *Low Agreement* a mean by mean comparison shows a significant difference between *Low Performer* (mean *Reputation* = 3.29) and *High Performer* (4.85) with respect to *Reputation* ($t = -2.91$; $p = .003$; one-tailed). Given *High Agreement* a mean by mean comparison shows no difference between *Low Performer* (mean = 5.00) and *High Performer* (5.00) with respect to *Reputation* ($t = .00$; $p = 1.00$). These results indicate that subordinates' reputation does not differ between low performers and high performers, when the rate of agreement between subordinates and superiors is high. High performers only receive a higher reputation than low performers, when the rate of agreement between superiors and subordinates is low. Therefore, a trade-off between accuracy and leniency with respect to subordinates' reputation is found and H3 is supported.

5 Discussion and Conclusion

This study addresses three issues associated with participative budgeting: effects of budget negotiation structure on subordinates' performance, superiors' perception of subordinates' performance, as well as effects of negotiation agreement, and subordinates' performance on superiors' assessment of subordinates' reputation.

Supporting referent cognitions theory, results show a significant interaction between initial budget proposal and final authority with respect to subordinates' performance. Subordinates' performance is high, except for the situation where subordinates make the initial budget proposal and superiors have final authority.

If superiors have final authority, in line with their incentives, final budget levels are high. Given superior final authority, if a superior makes the initial budget proposal, the mean difference between the initial budget proposal and the final budget is positive from a subordinate's perspective, indicating a fair budget negotiation process, and accordingly subordinates' performance is high. In contrast, given superior final authority, if a subordinate makes the initial budget proposal the mean difference between the initial budget proposal and the final budget is negative from a subordinate's perspective, revealing an unfair budget negotiation process, and consequently subordinates' performance is low. These results indicate egocentric fairness preferences and support referent cognitions theory, as a(n) fair (unfair) budget negotiation process leads to high (low) performance. Moreover, these findings contradict Fisher et al. (2000), who assume that

subordinates are likely to view a high initial budget proposal from superiors as unreasonable, and the resulting budget negotiation process and budgets as unfair. In contrast to Fisher et al. (2000), the results of this study suggest that subordinates assess the fairness of the budget negotiation in relative terms and view the initial budget proposal as a reference point relative to the final budget. These results support the view of subordinates having reference-dependent fairness preferences. Despite superiors' high initial budget proposals, the difference between initial budget proposals and final budgets is significantly positive from subordinates' perspectives, indicating a fair gain. In line with this, results show that subordinate performance is high.

The situation is different, when subordinates have final authority. First of all, results show that, if subordinates have final authority, budget levels are significantly higher than subordinates' profit-maximizing minimum budget, but well below subordinates' performance. Thus, in line with reciprocal fairness concerns, subordinates share a substantial amount of their payoffs with superiors. Furthermore, when a subordinate has final authority and a subordinate makes the initial proposal, the mean difference between the initial budget proposal and the final budget is negative from a subordinate's perspective, indicating an unfair loss and an unfair budget negotiation process. However, despite the negotiation seeming unfair, in this situation subordinates' performance is high. This supports referent cognitions theory in the sense that a subordinate who has final authority, and thus is responsible for the final budget in case of a negotiation impasse, does not blame a superior for an unfair process. Hence, the information about the process used to set the budget is unimportant, the initial budget proposal no longer constitutes a reference point, and a subordinate's motivation and performance is high. Also, when a subordinate has final authority and a superior makes the initial proposal, the mean difference between the initial budget proposal and the final budget is positive from a subordinate's perspective, indicating a fair budget negotiation process and subordinates' performance is high.

The results of this study extend referent cognitions theory to budget negotiations and are in line with results of the nonparticipative budgeting setting from Libby (2001). Libby (2001) examines the effects of the fairness of both, budget level and budgeting process, on subordinates' performance. In contrast to this study, the fairness of the budget level and the fairness of the budgeting process are manipulated at high (fair) and low (unfair) levels. Libby (2001) predicts and finds an interaction effect similar to the interaction effect in this study. When both, the budget level and the budgeting process,

are unfair, subjects' performance is lower than in all other conditions. When the budget level is fair, the fairness of procedures used to set the budget target has no apparent effect on subjects' performance. Moreover, the results of this study extend the findings of Fisher et al. (2002a) and indicate that budget negotiation structure is an important determinant of subordinates' fairness perceptions and influences subordinates' performance. Fisher et al. (2002a) show that negotiation agreement has a positive impact on subordinate performance as it indicates a fair budget negotiation process. Taken together, this implies that both, structural determinants of the budget negotiation and negotiation agreement, influence fairness concerns. Overall, findings provide support for fairness models and challenge the traditional assumption in economic theory that all individuals are expected utility maximizers motivated only by self-interest.

The second issue analyzed is the existence of potential bias of perceived subordinates' performance from a superior's perspective. An important issue, as identifying possible bias in performance evaluation can mitigate incentive problems (Bol 2011).

Results indicate that superiors' perception of subordinates' performance differs from subordinates' performance. When superiors make the initial budget proposal, superiors' perceived performance is significantly higher given superior final authority compared to subordinate final authority, whereas subordinates' performance does not differ. However, if subordinates make the initial budget proposal, superiors' perceived performance does not differ significantly between superior final authority and subordinate final authority although subordinates' performance is significantly lower given superior final authority compared to subordinate final authority.

Further analyses show that the pattern of superiors' subjective performance evaluations parallels the pattern of their profit. In line with Bol (2011) these results indicate that superiors may respond to their own incentives, when evaluating performance. Bol (2011) asserts that the personal costs associated with rating employee performance decreases the accuracy of performance ratings, indicating that subjective performance ratings are explained partly by managers' incentives and preferences. However in her field study and in contrast to this study, managers' subjective performance evaluations are rather based on employee performance, despite managers having the discretion to bias ratings. Thus, the different results might stem from additional factors influencing the likelihood of managers biasing their subjective performance evaluations in line with their incentives, which could be a fruitful avenue for further research.

The final issue examined in this study is the effect of subordinates' performance and negotiation agreement on subordinates' reputation. This issue is examined, as subordinate reputation plays an important role in participative budgeting (Stevens 2002, Webb 2002).

First of all, results show significant performance differences between individuals. Furthermore, negotiation agreement increases performance of low performers. However, negotiation agreement has no effect on performance of high performers. These results extend the findings of Fisher et al. (2002a), who also find a positive effect of negotiation agreement on performance but do not differentiate between low and high performers.

With respect to subordinates' reputation, results show that, given high agreement between superior and subordinate, subordinates' reputation, as being assessed by superiors, is high irrespective of subordinates' performance. However, given low agreement, reputation is positively associated with performance. These findings indicate that negotiation agreement increases social ties, extend the results of Breuer et al. (2013) and Bol (2011), and indicate a trade-off between accuracy and leniency with respect to subordinates' reputation. Breuer et al. (2013) provide empirical evidence that subjective performance evaluations tend to be lenient, when there is a close social proximity between a supervisor and subordinates. Bol (2011) shows that the strength of the employee-manager relationship increases centrality bias and leniency bias.

These results have several implications. First, designers of management control systems should take into account that budget negotiation structure can significantly affect subordinates' performance. If a subordinate makes the initial proposal and a superior has final authority a subordinate is likely to perceive the budget negotiation as being unfair. Therefore, when considering a top-down budgeting approach, upper management should make the initial budget proposal in the budget-setting process, as employees perceive concessions from upper management as fair. Second, this study implies that without further specifying on which information a superior's subjective performance evaluation should be based upon, the possibility is given that a superior chooses different reference points than those being intended, which increases the risk of adversely evaluating performance. Third, a performance-accuracy trade-off with respect to negotiation agreement can be identified. Low performers perform significantly higher, when negotiation agreement is high. However, when negotiation agreement is high,

their reputation rating is too lenient, leading to the above mentioned negative consequences. Nevertheless, leniency bias is also positively associated with future performance (Bol 2011). Disentangling these effects could be an avenue for future research.

This experimental study is subject to limitations that restrict the extent to which the findings can be generalized. First, the dependent variables perceived performance and reputation are based on items from the postexperimental questionnaire measured on a seven-point Likert scale. They must be interpreted with care, as the validity of Likert scale measurement can be compromised due to social desirability. However, the experiment was anonymous and offering anonymity on self-administered questionnaires should further reduce social pressure, and thus may likewise reduce social desirability bias.

The compensation scheme of superiors is a special case, since generally superiors in superior-subordinate relationships are not residual claimants, as most companies are multi-layered (Prendergast 1999). Nevertheless, this compensation scheme is adopted from Fisher et al. (2006, 2002a, 2002b, 2000) as it is well-established and comparable to prior empirical studies.

Furthermore, this study does not formally derive predictions or expectations of equilibrium behavior from an analytical model (Maas et al. 2012). Analytical accounting research has only just begun to develop participative budgeting models (e.g. Heinle et al. 2014) and relies on the standard wealth-maximizing assumption. Also, in the economics literature formal theories including fairness preferences have not been applied to budgetary settings. Therefore, in developing the hypotheses, this study relies on behavioral referent cognitions theory, regular empirical findings, and the intuition provided by recent economic fairness models (Falk and Fischbacher 2006). However, the results of this study suggest that future analytical models of participative budgeting should integrate social preferences into the utility function of subordinates.

References

- Adams, J. S. 1965. Inequity in social exchange. In L. Berkowitz (Ed.): *Advances in experimental social psychology* 2. New York, NY: Academic Press: 267-299.
- Anthony, R. N., and V. Govindarajan. 2007. *Management control systems*. Chicago, IL: Irwin McGraw-Hill.
- Austin, W. 1980. Friendship and fairness effects of type of relationship and task performance on choice of distribution rules. *Personality and Social Psychology Bulletin* 6 (3): 402-408.
- Bailey, W. J., G. Hecht, and K. L. Towry. 2011. Dividing the Pie: The Influence of Managerial Discretion Extent on Bonus Pool Allocation. *Contemporary Accounting Research* 28 (5): 1562-1584.
- Baiman, S. 1982. Agency research in managerial accounting: A survey. *Journal of Accounting Literature* 1 (1): 154-213.
- Baiman, S. 1990. Agency research in managerial accounting: A second look. *Accounting, Organizations and Society* 15 (4): 341-371.
- Bol, J. C. 2011. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review* 86 (5): 1549-1575.
- Bolton, G. E., and A. Ockenfels. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review* 90 (1): 166-193.
- Bonner, S. E., H. Reid, G. B. Sprinkle, and S. M. Young. 2000. A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting. *Journal of Management Accounting Research* 12 (1): 19-64.
- Bonner, S. E., and G. B. Sprinkle. 2002. The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27 (4): 303-345.
- Brandstätter, H., G. Kette, and J. Sageder. 1983. Expectations, attribution, and behavior in bargaining with liked and disliked partners. In: R. Tietz (Ed.): *Aspiration levels in bargaining and economic decision making*. Berlin: Springer: 136-152.
- Breuer, K., P. Nieken, and D. Sliwka. 2013. Social ties and subjective performance evaluations: an empirical investigation. *Review of Managerial Science* 7 (2): 141-157.
- Buckless, F. A., and S. P. Ravenscroft. 1990. Contrast Coding: A Refinement of ANOVA in Behavioral Analysis. *The Accounting Review* 65 (4): 933-945.
- Camerer, C. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: University Press.
- Chang, L., M. Cheng, and K. T. Trotman. 2008. The effect of framing and negotiation partner's objective on judgments about negotiated transfer prices. *Accounting, Organizations and Society* 33 (7-8): 704-717.
- Charness, G., and M. Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117 (3): 817-869.

- Chow, C. W., J. C. Cooper, and W. S. Waller. 1988. Participative Budgeting: Effects of a Truth-Inducing Pay Scheme and Information Asymmetry on Slack and Performance. *The Accounting Review* 63 (1): 111-122.
- Chow, C. W., J. C. Cooper, and K. Haddad. 1991. The effects of pay schemes and ratchets on budgetary slack and performance: A multiperiod experiment. *Accounting, Organizations and Society* 16 (1): 47-60.
- Chow, C. W., M. K. Hirst, and M. D. Shields. 1994. Motivating Truthful Subordinate Reporting: An Experimental Investigation in a Two Subordinate Context. *Contemporary Accounting Research* 10 (2): 699-720.
- Chow, C. W., M. K. Hirst, and M. D. Shields. 1995. The effects of pay schemes and probabilistic management audits on subordinate misrepresentation of private information: An experimental investigation in a resource allocation context. *Behavioral Research in Accounting* 7 (1): 1-16.
- Cropanzano, R., and R. Folger. 1989. Referent cognitions and task decision autonomy: Beyond equity theory. *Journal of Applied Psychology* 74 (2): 293-299.
- Deci, E. L. 1975. *Intrinsic Motivation*. New York, NY: Plenum Press.
- Dejong, D. V., R. Forsythe, and R. J. Lundholm. 1985. Ripoffs, lemons, and reputation formation in agency relationships: A laboratory market study. *The Journal of Finance* 40 (3): 809-820.
- Dopuch, N., R. R. King, and R. Schwartz. 2001. An experimental investigation of retention and rotation requirements. *Journal of Accounting Research* 39 (1): 93-117.
- Dufwenberg, M., and G. Kirchsteiger. 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47 (2): 268-298.
- Eisenhardt, K. M. 1989. Agency theory: an assessment and review. *Academy of Management Review* 14 (1): 57-74.
- Evans III, J. H., R. L. Hannan, R. Krishnan, and D. V. Moser. 2001. Honesty in managerial reporting. *The Accounting Review* 76 (4): 537-559.
- Falk, A., and U. Fischbacher. 2006. A theory of reciprocity. *Games and Economic Behavior* 54 (2): 293-315.
- Falk, A., E. Fehr, and U. Fischbacher. 2008. Testing theories of fairness—Intentions matter. *Games and Economic Behavior* 62 (1): 287-303.
- Fehr, E., and H. Gintis. 2007. Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology* 33 (1): 43-64.
- Fehr, E., and K. M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114 (3): 817-868.
- Fehr, E., and K. M. Schmidt. 2003. Theories of fairness and reciprocity—Evidence and economic applications. In: M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Ed.): *Advances in Economic Theory: Eighth World Congress of the Economic Society*. Cambridge, U.K.: Cambridge University Press.
- Fehr, E., and K. M. Schmidt. 2004. Fairness and incentives in a multi-task principal-agent model. *Scandinavian Journal of Economics* 106 (3): 453-474.
- Fischbacher, U. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2): 171-178.

- Fisher, J. G., J. R. Frederickson, and S. A. Pfeffer. 2000. Budgeting: An Experimental Investigation of the Effects of Negotiation. *The Accounting Review* 75 (1): 93-114.
- Fisher, J. G., J. R. Frederickson, and S. A. Pfeffer. 2002a. The effect of information asymmetry on negotiated budgets: an empirical investigation. *Accounting, Organizations and Society* 27 (1-2): 27-43.
- Fisher, J. G., L. A. Maines, S. A. Pfeffer, and G. B. Sprinkle. 2002b. Using Budgets for Performance Evaluation: Effects of Resource Allocation and Horizontal Information Asymmetry on Budget Proposals, Budget Slack, and Performance. *The Accounting Review* 77 (4): 847-865.
- Fisher, J. G., S. A. Pfeffer, and G. B. Sprinkle. 2003. Budget-Based Contracts, Budget Levels, and Group Performance. *Journal of Management Accounting Research* 15 (1): 51-74.
- Fisher, J. G., J. R. Frederickson, and S. A. Pfeffer. 2006. Budget negotiations in multi-period settings. *Accounting, Organizations and Society* 31 (6): 511-528.
- Folger, R. 1986. Rethinking Equity Theory. In: H. W. Bierhoff, R. L. Cohen, and J. Greenberg (Ed.): *Justice in Social Relations*. New York, NY: Plenum Press: 145-162.
- Gächter, S., and A. Falk. 2002. Reputation and reciprocity: Consequences for the labour relation. *The Scandinavian Journal of Economics* 104 (1): 1-26.
- Heinle, M. S., N. Ross, and R. E. Saouma. 2014. A Theory of Participative Budgeting. *The Accounting Review* 89 (3): 1025-1050.
- Hopwood, A. G. 1976. *Accounting and Human Behavior*. New York, NY: Prentice Hall.
- IBM Corporation. 2012. *IBM SPSS Advanced Statistics 21*. Copyright IBM Corporation 1989.
- Izard, C. E. 1977. *Human emotions*. New York, NY: Plenum Press.
- Jawahar, J., and C. R. Williams. 1997. Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology* 50 (4): 905-925.
- Kahneman, D. 1992. Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes* 51 (2): 296-312.
- Koford, K., and M. Penno. 1992. Accounting, Principal-Agent Theory and Self-Interested Behavior. In: N. E. Bowie, and E. R. Freeman (Ed.): *Ethics and Agency Theory – An Introduction*. Oxford, NY: Oxford University Press: 127-142.
- Kreps, D. M., and R. Wilson. 1982. Reputation and imperfect information. *Journal of Economic Theory* 27 (2): 253-279.
- Kristensen, H., and T. Gärling. 1997. The effects of anchor points and reference points on negotiation process and outcome. *Organizational Behavior and Human Decision Processes* 71 (1): 85-94.
- Lambert, R. A. 2007. Agency Theory and Management Accounting. In: C. S. Chapman, A. G. Hopwood, and M. D. Shields (Ed.): *Handbook of Management Accounting Research I*. Oxford, U.K.: Elsevier: 247-268.
- Lawler, E. J. 1992. Affective attachments to nested groups: A choice-process theory. *American Sociological Review* 57 (3): 327-339.

- Levine, D. K. 1998. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics* 1 (3): 593-622.
- Libby, T. 2001. Referent cognitions and budgetary fairness: A research note. *Journal of Management Accounting Research* 13 (1), S. 91-105.
- Liebert, R. M., W. P. Smith, J. H. Hill, and M. Keiffer. 1968. The effects of information and magnitude of initial offer on interpersonal negotiation. *Journal of Experimental Social Psychology* 4 (4): 431-441.
- Lind, E. A., R. Kanfer, and P. C. Earley. 1990. Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments. *Journal of Personality and Social Psychology* 59 (5): 952-959.
- Lindquist, T. M. 1995. Fairness as an antecedent to participative budgeting: Examining the effects of distributive justice, procedural justice and referent cognitions on satisfaction and performance. *Journal of Management Accounting Research* 7 (1): 122-147.
- Lipe, M. G., and S. E. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 75 (3): 283-298.
- Luft, J. L., and R. Libby. 1997. Profit comparisons, market prices and managers' judgments about negotiated transfer prices. *The Accounting Review* 72 (2): 217-229.
- Maas, V. S., M. van Rinsum, and K. L. Towry. 2012. In Search of Informed Discretion: An Experimental Investigation of Fairness and Trust Reciprocity. *The Accounting Review* 87 (2): 617-644.
- Merchant, K. A., W. A. van der Stede. 2012. *Management control systems. Performance measurement, evaluation and incentives*. 3. Edition. Harlow, U.K.: Pearson, Financial Times/ Prentice Hall.
- Miller, G. 1992. Ethics and the new game theory. In: N. E. Bowie, and R. E. Freeman (Ed.): *The Ruffin Series in Business Ethics*. Oxford, NY: Oxford University Press: 117-126.
- Moers, F. 2005. Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society* 30 (1): 67-80.
- Neale, M. A., and M. H. Bazerman. 1991. *Cognition and rationality in negotiation*. New York, NY: Free Press.
- Northcraft, G. B., and M. A. Neale. 1987. Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes* 39 (1): 84-97.
- Prendergast, C. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37 (1): 7-63.
- Prendergast, C. 2002. Uncertainty and incentives. *Journal of Labor Economics* 20 (2): 115- 137.
- Prendergast, C., and R. Topel. 1993. Discretion and bias in performance evaluation. *European Economic Review* 37 (2): 355-365.
- Pruitt, D. G., and P. J. Carnevale. 1993. *Negotiation in social conflict*. Pacific Grove, CA: Thomson Brooks/ Cole Publishing Co.

- Rabin, M. 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83 (5): 1281-1302.
- Rankin, F. W., S. T. Schwartz, and R. A. Young. 2003. Management Control Using Nonbinding Budgetary Announcements. *Journal of Management Accounting Research* 15 (1): 75-93.
- Rankin, F. W., S. T. Schwartz, and R. A. Young. 2008. The effect of honesty and superior authority on budget proposals. *The Accounting Review* 83 (4): 1083-1099.
- Ritov, I. 1996. Anchoring in simulated competitive market negotiation. *Organizational Behavior and Human Decision Processes* 67 (1): 16-25.
- Rubin, J. Z., and B. R. Brown. 1975. *The social psychology of bargaining and negotiation*. New York, NY: Academic Press.
- Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. *Biometrics bulletin* 2 (6): 110-114.
- Schatzberg, J. W., and D. E. Stevens. 2008. Public and Private Forms of Opportunism within the Organization: A Joint Examination of Budget and Effort Behavior. *Journal of Management Accounting Research* 20 (1): 59-81.
- Sprinkle, G. B., and M. G. Williamson. 2007. Experimental Research in Managerial Accounting. In: C. S. Chapman, A. G. Hopwood, and M. D. Shields (Ed.): *Handbook of Management Accounting Research 1*. Oxford, U.K.: Elsevier: 415-444.
- Stevens, D. E. 2002. The Effects of Reputation and Ethics on Budgetary Slack. *Journal of Management Accounting Research* 14 (1): 153-171.
- Thompson, L. 1990. Negotiation behavior and outcomes: Empirical evidence and theoretical issues. *Psychological Bulletin* 108 (3): 515-532.
- Thompson, L. 1995. The impact of minimum goals and aspirations on judgments of success in negotiations. *Group Decision and Negotiation* 4 (6): 513-524.
- Thompson, L., and R. Hastie. 1990. Social perception in negotiation. *Organizational Behavior and Human Decision Processes* 47 (1): 98-123.
- Thompson, L., and G. Loewenstein. 1992. Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes* 51 (2): 176-197.
- Tsui, A. S. 1984. A role set analysis of managerial reputation. *Organizational Behavior and Human Performance* 34 (1): 64-96.
- Umapathy, S. 1987. *Current budgeting practices in US industry: The state of the art*. New York, NY: Quorum Books.
- Waller, W. S. 1988. Slack in participative budgeting: The joint effect of a truth-inducing pay scheme and risk preferences. *Accounting, Organizations and Society* 13 (1): 87-98.
- Webb, R. A. 2002. The impact of reputation and variance investigations on the creation of budget slack. *Accounting, Organizations and Society* 27 (4-5): 361-378.
- Whyte, G., and J. K. Sebenius. 1997. The effect of multiple anchors on anchoring in individual and group judgment. *Organizational Behavior and Human Decision Processes* 69 (1): 74-85.
- Wilson, R. 1985. Reputations in games and markets. In: A. Roth (Ed.): *Game-theoretic models of bargaining*. Cambridge, U.K.: Cambridge University Press: 27-62.

- Young, S. M. 1985. Participative Budgeting: The Effects of Risk Aversion and Asymmetric Information on Budgetary Slack. *Journal of Accounting Research* 23 (2): 829-842.
- Young, S. M., J. Fisher, and T. M. Lindquist. 1993. The effects of intergroup competition and intragroup cooperation on slack and output in a manufacturing setting. *The Accounting Review* 68 (3): 466-481.
- Yukl, G. 1974. Effects of the opponent's initial offer, concession magnitude and concession frequency on bargaining behavior. *Journal of Personality and Social Psychology* 30 (3): 323-335.
- Zhang, Y. 2008. The effects of perceived fairness and communication on honesty and collusion in a multi-agent setting. *The Accounting Review* 83 (4): 1125-1146.