

Institut für Visualisierung und Interaktive Systeme

Universität Stuttgart  
Universitätsstraße 38  
D-70569 Stuttgart

Bachelorarbeit Nr. 128

# **Untersuchung von Visualisierungsalternativen zur Vermittlung von unsicheren Daten an Laien**

Thorsten Ohler

<b>Studiengang:</b>	Informatik
<b>Prüfer/in:</b>	Jun.-Prof. Dr. Niels Henze
<b>Betreuer/in:</b>	Dipl.-Inf. Miriam Greis
<b>Beginn am:</b>	1. Mai 2014
<b>Beendet am:</b>	31. Oktober 2014
<b>CR-Nummer:</b>	H.1.2, J.4



## Kurzfassung

Wir leben in einer Welt in der wir überall und tagtäglich mit Risiken und Unsicherheiten umgehen müssen. Jeder muss im Laufe der Zeit immer wieder auf Grund von Statistiken wichtige Entscheidungen treffen, die mit Risiken und Unsicherheiten verbunden sind. Aber noch ist unklar, in wie weit die Angabe von Informationen zur Unsicherheit die Menschen bei ihren Entscheidungen unterstützt und welche Visualisierungen und Darstellungen sich am besten eignen, um unsichere Daten an Laien zu vermitteln. Obwohl bereits einige Forschung in Sachen Kommunikation von unsicheren Daten betrieben wurde, blieb diese meist sehr theoretisch und wurde häufig nur durch Umfragen untermauert, denn es fehlten empirische Daten aus echten Entscheidungen. In dieser Arbeit wurde daher ein Spiel entwickelt, bei dem die Teilnehmer auf Basis von unsicheren Daten Entscheidungen fällen mussten. Zuerst wurde hierfür mit Hilfe einer Online-Umfrage aus zwölf vorgeschlagenen Darstellungen vier unterschiedliche Abbildungen für die Darstellung eines Wetterberichtes im Spiel ausgewählt. Eine Darstellung enthielt nur den textuellen Erwartungswert ohne Angaben zur Unsicherheit. Die drei anderen Darstellungen waren allesamt graphisch und enthielten unterschiedlich differenzierte Informationen zur Unsicherheit. Nachdem das Spiel mit den vier gewählten Darstellungen entwickelt und implementiert war, wurde es sowohl öffentlich und unmoderiert, als auch unter kontrollierten Bedingungen gespielt. Die Analyse der Daten und Entscheidungen aus den gespielten Partien ergab, dass für die meisten Menschen die Angabe von Informationen zur Unsicherheit ein Mehrwert für ihre Entscheidungen bedeutet. Außerdem zeigte sich, dass die Qualität der Entscheidungen der Laien mit zunehmenden Angaben zur Unsicherheit stieg. Die kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen stellte sich in diesem Kontext als beste Visualisierungsalternative zur Vermittlung von unsicheren Daten an Laien heraus. Das Verständnis über die Fähigkeit zur Kommunikation von unsicheren Daten der verschiedenen Visualisierungsalternativen und Darstellungsformen stiftet für viele Entscheidungen einen großen Nutzen. Es kann dazu beitragen, dass unsichere Daten seltener deterministisch dargestellt werden und stattdessen die probabilistische Kommunikation von unsicheren Daten weitere Verbreitung findet. Denn es wurde gezeigt, dass die probabilistische Darstellung von unsicheren Daten einen großen Mehrwert für die Entscheidungen der Menschen bedeuten kann.

## **Abstract**

We live in a world where we have to deal with risks and uncertainties everywhere and every day. Over time, everyone has to make important decisions on the basis of statistics that involve risks and uncertainties. But it is still unclear to what extent the provision uncertainty information supports people in their decisions and which visualizations and representations are best suited to communicate uncertain data to laymen. Although some research in communication of uncertain data has already been conducted, it was mostly very theoretical and was often supported only by surveys, because it lacked empirical data from real decisions. In this paper, a game was developed, in which the participants had to make decisions based on uncertain data. First, for this purpose four out of twelve suggested representations were selected with the help of an online survey for the representation of a weather report in the game. One representation contained only the textual expected value without uncertainty information. The other three representations were all graphically and contained different differentiated uncertainty information. After the game was developed and implemented with the four selected representations, it was played publicly and unmoderated, as well as under controlled conditions. The analysis of the data and decisions from the games played showed that for most people the provision of uncertainty information means more value for their decisions. Furthermore, it was found that the more detailed the uncertainty informations are, the better decisions are made by the layman. The continuous representation of a histogram in the form of a function graph turned out in this context as the best visualization alternative for communicating uncertain data to laymen. The understanding of the ability to communicate uncertain data using different visualization alternatives and representations can help to ensure that uncertain data are shown less often deterministic and instead probabilistic communication of uncertain data is more widely used. The probabilistic representation of uncertain data can mean a great added value for people's decisions.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>11</b>
<b>2. Problemstellung</b>	<b>13</b>
2.1. Thematische Eingrenzung . . . . .	13
2.2. Wettervorhersagen . . . . .	14
<b>3. Verwandte Arbeiten</b>	<b>15</b>
3.1. Bedarf der Kommunikation von Unsicherheiten . . . . .	15
3.2. Schwierigkeiten bei der Kommunikation von Unsicherheiten . . . . .	16
3.2.1. Rechenschwächen . . . . .	16
3.2.2. Qualitative vs. quantitative Angaben . . . . .	17
3.2.3. Menge an Angaben . . . . .	17
3.2.4. Gewinn- vs. Verlust-Formulierung . . . . .	17
3.2.5. Absolute vs. relative Veränderung des Risikos . . . . .	18
3.2.6. Natürliche Frequenzen vs. Prozentangaben . . . . .	18
3.2.7. Zeitspanne . . . . .	19
3.3. Interpretation der Öffentlichkeit von unsicheren Prognosen . . . . .	19
3.4. Darstellungen für die Kommunikation von Unsicherheiten . . . . .	21
3.5. Zusammenfassung und Diskussion . . . . .	23
<b>4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten</b>	<b>25</b>
4.1. Art der darzustellenden Daten . . . . .	25
4.2. Identifikation und Klassifikation von Darstellungen . . . . .	25
4.3. Methodik der Online-Umfrage . . . . .	30
4.3.1. Befragung . . . . .	30
4.3.2. Vorbereitung und Veröffentlichung . . . . .	31
4.3.3. Teilnehmer . . . . .	32
4.4. Ergebnisse . . . . .	32
4.4.1. Entscheidungshilfe . . . . .	33
4.4.2. Vertrautheit . . . . .	35
4.4.3. Einfaches Verständnis . . . . .	36
4.4.4. Optisch ansprechend . . . . .	38
4.4.5. Korrelation zwischen den vier Fragestellungen . . . . .	39
4.5. Kommentare der Teilnehmer . . . . .	39
4.6. Diskussion . . . . .	40
4.7. Auswahl für das Experiment . . . . .	41

<b>5. Anwendung Farm-Smart</b>	<b>43</b>
5.1. Spielablauf und Konzept . . . . .	43
5.2. Implementierung . . . . .	48
<b>6. Öffentliches Experiment</b>	<b>51</b>
6.1. Auswertung der Daten . . . . .	51
6.2. Veröffentlichung . . . . .	56
6.3. Teilnehmer . . . . .	56
6.4. Ergebnisse . . . . .	57
6.4.1. Optimale Entscheidungen . . . . .	57
6.4.2. Suboptimale Entscheidungen . . . . .	58
6.4.3. Dauer der Einsicht des Wetterberichts . . . . .	58
6.4.4. Häufigkeit der Einsicht des Wetterberichts . . . . .	60
6.4.5. Fragestellungen aus der Online-Umfrage . . . . .	61
6.5. Kommentare der Teilnehmer . . . . .	63
6.6. Diskussion . . . . .	63
<b>7. Kontrolliertes Experiment</b>	<b>67</b>
7.1. Auswertung der Daten . . . . .	67
7.2. Veröffentlichung . . . . .	67
7.3. Teilnehmer . . . . .	67
7.4. Ergebnisse . . . . .	68
7.4.1. Optimale Entscheidungen . . . . .	68
7.4.2. Suboptimale Entscheidungen . . . . .	69
7.4.3. Dauer der Einsicht des Wetterberichts . . . . .	70
7.4.4. Häufigkeit der Einsicht des Wetterberichts . . . . .	71
7.4.5. Fragestellungen aus der Online-Umfrage . . . . .	72
7.4.6. Befragung nach der besten Darstellung . . . . .	73
7.5. Kommentare der Teilnehmer . . . . .	73
7.6. Diskussion . . . . .	74
<b>8. Zusammenfassung und Fazit</b>	<b>77</b>
<b>A. Anhang</b>	<b>79</b>
A.1. Mockup der Anwendung Farm-Smart . . . . .	79
<b>Literaturverzeichnis</b>	<b>87</b>

# Abbildungsverzeichnis

---

4.1.	Darstellungen 1-6 - Darstellung 1: textueller Erwartungswert, Darstellung 2: textueller Erwartungswert mit Standardabweichung, Darstellung 3: textuelle Quantilangaben (0.25-, 0.50-, 0.75-Quantile), Darstellung 4: Liniendiagramm des Erwartungswerts mit Spannweite als Flächendiagramm, Darstellung 5: Box-Whisker-Plot, Darstellung 6: Balkendiagramm des Erwartungswert mit Fehlerbalken. . . . .	28
4.2.	Darstellungen 7-12 - Darstellung 7: Histogramm in Form eines Balkendiagramms, Darstellung 8: kumulatives Histogramm in Form eines gestapeltem Balkendiagramms, Darstellung 9: kumulatives Histogramm in Form eines gestapeltem Flächendiagramms, Darstellung 10: transparentes Wahrscheinlichkeitsdichtediagramm ohne Variation in der Höhe, Darstellung 11: kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen, Darstellung 12: kontinuierliche Darstellung eines kumulativen Histogramms in Form eines Funktionsgraphen. . . . .	29
4.3.	Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der 12 den Teilnehmern vorgestellten Darstellungen. Bewertungen kleiner als 2.4 wurden in rot, zwischen 2.4 und 3.0 in orange, zwischen 3.0 und 3.6 in hellgrün und größer als 3.6 in dunkelgrün dargestellt. . . . .	32
4.4.	Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer half eine Entscheidung zu treffen. . . . .	33
4.5.	Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer vertraut war. . . . .	35
4.6.	Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer einfach zu verstehen war. . . . .	36
4.7.	Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer optisch ansprechend war. . . . .	38
4.8.	Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der 12 Darstellungen, die den Teilnehmern vorgestellt wurden. Bewertungen kleiner als 2.4 wurden in rot, zwischen 2.4 und 3.0 in orange, zwischen 3.0 und 3.6 in hellgrün und größer als 3.6 in dunkelgrün dargestellt. . . . .	41
5.1.	Der Begrüßungstext im Spiel. Eine kleine Einweisung wird gegeben und die Aufgabe für den Spieler kurz beschrieben. . . . .	44
5.2.	Die Auswahl der Pflanzen im Menü des Spiels. Unter der jeweiligen Abbildung der Pflanze sind die Angaben zu den Kosten, Wert, Anbauzeit, benötigte Mindesttemperatur, Mindestniederschlagsmenge und maximale Windstärke abgebildet. . . . .	45
5.3.	Die geänderte Darstellung 11 des Wetterberichts im Spiel mit 3-Tages-Vorschau für Temperatur, Niederschlag und Windstärke, den der Spieler täglich einsehen konnte. .	46

5.4.	Die vier der fünf Zustände, die eine Ackerfläche annehmen konnte: „ungepflügt“, „gepflügt“, „gesät“ und „gestorben“ (von links nach rechts). Der fünfte Zustand „reif“ wurde mit der Grafik der jeweilig reifen Pflanze angezeigt und ist in Abbildung 5.2 zu sehen. . . . .	48
5.5.	Das isometrisches Tile-basierte 2D Spielfeld mit den 36 Ackerflächen auf denen jeweils eine Pflanze angebaut werden kann. Damit eine Pflanze angebaut werden konnte, musste der Spieler zuerst das Feld für 10 Euro pflügen. Dazu musste er im unteren Menü das Feld „Pflügen“ auswählen und danach das zu pflügende Ackerfeld klicken. Daraufhin konnte er die Pflanze seiner Wahl anbauen, indem er wieder im Menü zuerst die Pflanze auswählte und danach auf ein bereits gepflügte Ackerfeld klickte.	49
6.1.	Die untersuchten Darstellungen 1-4 - Darstellung 1: textueller Erwartungswert, Darstellung 2: Liniendiagramm des Erwartungswerts mit Spannweite als Flächendiagramm, Darstellung 3: Histogramm in Form eines Balkendiagramms, Darstellung 4: kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen. . .	52
6.2.	Optimale Entscheidungen abhängig von der Darstellung im öffentlichen Experiment.	57
6.3.	Sinnvolle Entscheidungen abhängig von der Darstellung und dem Risikoverhalten im öffentlichen Experiment. . . . .	58
6.4.	Durchschnittliche Dauer der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im öffentlichen Experiment. Fehlerbalken zeigen den Standardfehler. . .	59
6.5.	Anteil der Runden in denen der Wetterbericht mindestens einmal angezeigt wurde in Abhängigkeit der Darstellung und Runde im öffentlichen Experiment. . . . .	60
6.6.	Durchschnittliche Häufigkeit der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im öffentlichen Experiment. Fehlerbalken zeigen den Standardfehler.	60
6.7.	Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der vier Darstellungen, die die Teilnehmer abgegeben hatten, nachdem sie ein Spiel vollständig mit der jeweiligen Darstellung gespielt hatten. Die Teilnehmer wurden gefragt, ob die Darstellung ihnen geholfen hatte eine Entscheidung zu treffen, ob die Darstellung ihnen bereits vertraut war, ob die Darstellung für sie einfach zu verstehen war und ob die Darstellung optisch ansprechend war. Fehlerbalken zeigen den Standardfehler. . . . .	61
6.8.	Gegenüberstellung der durchschnittlichen Bewertungen aus der vorausgegangenen Online-Umfrage, dem öffentlichen Experiment und dem kontrollierten Experiment bezüglich der vier Fragestellungen: Entscheidungshilfe, Vertrautheit, Einfaches Verständnis, Optisch ansprechend. Fehlerbalken zeigen den Standardfehler. . . . .	62
7.1.	Optimale Entscheidungen abhängig von der Darstellung im kontrollierten Experiment.	68
7.2.	Sinnvolle Entscheidungen abhängig von der Darstellung und dem Risikoverhalten im kontrollierten Experiment. . . . .	69
7.3.	Durchschnittliche Dauer der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im kontrollierten Experiment. Fehlerbalken zeigen den Standardfehler. .	70
7.4.	Anteil der Runden in denen der Wetterbericht mindestens einmal angezeigt wurde in Abhängigkeit der Darstellung und Runde im kontrollierten Experiment. . . . .	71
7.5.	Durchschnittliche Häufigkeit der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im kontrollierten Experiment. Fehlerbalken zeigen den Standardfehler.	71



7.6.	Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der vier Darstellungen, die die Teilnehmer abgegeben hatten, nachdem sie ein Spiel vollständig mit der jeweiligen Darstellung gespielt hatten. Die Teilnehmer wurden gefragt, ob die Darstellung ihnen geholfen hatte eine Entscheidung zu treffen, ob die Darstellung ihnen bereits vertraut war, ob die Darstellung für sie einfach zu verstehen war und ob die Darstellung optisch ansprechend war. Fehlerbalken zeigen den Standardfehler. . . . .	72
7.7.	Die Verteilung der zwölf Teilnehmerstimmen im kontrollierten Experiment bezüglich der Frage, welche Darstellung ihnen am besten gefallen hat. . . . .	73
A.1.	Logo und Fortschrittsbalken (Preloading der Bilder) . . . . .	79
A.2.	Willkommensbildschirm, Aufgabenbeschreibung und kleine Einweisung . . . . .	80
A.3.	Dem Spieler wird der erste Wetterbericht für den morgigen Tag angezeigt . . . . .	80
A.4.	Ansicht des ersten Wetterberichts für den übermorgigen Tag. Es werden verschiedene Visualisierungen der Wahrscheinlichkeitsverteilungen angezeigt. . . . .	81
A.5.	Der Wetterbericht ist vom Spieler unten rechts immer noch einzusehen. Der Spieler kann nun darauf basierend entscheiden, welche Pflanzen er anbauen möchte. . . . .	81
A.6.	Bevor der Spieler etwas anpflanzen kann, muss er zuerst ein Ackerfeld pflügen. . . . .	82
A.7.	Die Pflanzen haben verschiedene Anbaukosten, Werte beim Ernten, Anbauzeiten und Anforderungen an Regen, Sonne und Wind. Der Spieler entscheidet sich aufgrund des Wetterberichts Kürbisse anzubauen. . . . .	82
A.8.	Vier Tage sind vergangen. Die Kürbisse sind reif, da sie eine Anbauzeit von 4 Tagen haben. Der Spieler kann sie nun ernten. . . . .	83
A.9.	Es vergehen weitere Tage. Der Spieler bekommt täglich neue Wetterberichte und baut darauf basierend neue Pflanzen an und erntet reife Pflanzen. . . . .	83
A.10.	Tritt das falsche Wetter ein, können die Pflanzen kaputt gehen und er macht Verlust. Die kaputten Pflanzen muss der Spieler entgeltlich entfernen. . . . .	84
A.11.	Die Spieldauer ist um. Das Spiel endet und der endgültige Geldstand wird angezeigt und gespeichert. . . . .	84
A.12.	Der Highscore der bisher gespielten Partien werden dem Spieler gezeigt. . . . .	85

## Tabellenverzeichnis

---

4.1.	Klassifikation der 12 Darstellungen nach dem Grad der Darstellung der Unsicherheit. . . . .	27
4.2.	Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer half eine Entscheidung zu treffen, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von $p < 0,0008$ führte. Statistisch signifikante Werte sind mit einem (*), zwei (**) oder (***) Sternen gekennzeichnet. * $p < 0.0008$ , ** $p < 0.00016$ , *** $p < 0,000016$ . . . . .	34

4.3.	Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer vertraut war, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von $p < 0,0008$ führte. Statistisch signifikante Werte sind mit einem (*), zwei (**) oder (***) Sternen gekennzeichnet. * $p < 0.0008$ , ** $p < 0.00016$ , *** $p < 0,000016$ . . . . .	34
4.4.	Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer einfach zu verstehen war, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von $p < 0,0008$ führte. Statistisch signifikante Werte sind mit einem (*), zwei (**) oder (***) Sternen gekennzeichnet. * $p < 0.0008$ , ** $p < 0.00016$ , *** $p < 0,000016$ . . . . .	37
4.5.	Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung optisch ansprechend ist, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von $p < 0,0008$ führte. Statistisch signifikante Werte sind mit einem (*), zwei (**) oder (***) Sternen gekennzeichnet. * $p < 0.0008$ , ** $p < 0.00016$ , *** $p < 0,000016$ . . . . .	37
4.6.	Pearson-Korrelation zwischen den vier verschiedenen Fragestellungen. Alle Werte sind statistisch signifikant, $p < 0,0005$ . . . . .	39
6.1.	Gewinne pro Anbautag für jede Pflanze. Wenn das Wetter es zuließ verschiedene Pflanzen anzubauen, war es sinnvoll, sich nach dieser Tabelle zu richten, um den höchsten Gewinn zu erzielen. . . . .	53
6.2.	Die Berechnungen der Grenzwerte für Niederschlag, Temperatur und Windstärke für jede Darstellung und jedes Risikoverhalten. . . . .	55

# 1. Einleitung

“In this world, nothing is certain except death and taxes” stellte schon Benjamin Franklin 1789 fest. Diese Feststellung hat nach wie vor Gültigkeit. Sie verdeutlicht uns, dass nichts in dieser Welt sicher ist, auch wenn es oft so scheinen mag. Wir leben in einer Welt in der wir überall und tagtäglich mit Risiken und Unsicherheiten umgehen müssen. Viele der modernen technischen Errungenschaften in unserer heutigen Welt schaffen die besten Voraussetzungen, um Risiken zu minimieren, indem sie berechnet und numerisch dargestellt werden. Jedoch haben Studien [GM10] gezeigt, dass es auch in unserer modernen, technisch geprägten Gesellschaft in großen Teilen der Bevölkerung noch immer ein großes Defizit beim Verständnis von statistischen Unsicherheiten und numerischen Problemen gibt. Dieses Verständnis wird jedoch benötigt, um aus den berechneten Daten Entscheidungen abzuleiten, die dabei helfen Risiken tatsächlich zu minimieren.

Der Umgang mit Unsicherheiten gehört zu unserem täglichen Leben. Deshalb muss jeder im Laufe der Zeit immer wieder auf Grund von Statistiken wichtige Entscheidungen treffen, die mit Risiken und Unsicherheiten verbunden sind. So erläutern beispielsweise Ärzte den Patienten bei Arztbesuchen die gesundheitlichen Risiken von Medikamenten oder operativen Eingriffen anhand von Wahrscheinlichkeiten. Doch nicht nur bei außergewöhnlichen und wichtigen Entscheidungen, sondern auch bei alltäglichen Dingen, wie der Tagesplanung, die oftmals von der Wettervorhersage abhängt, treffen wir auf diese Unsicherheiten. Abhängig von der Wettervorhersage entscheiden wir, welche Kleidung wir anziehen und wann wir das Haus verlassen.

Für den Umgang mit Unsicherheiten ist es bei weitem nicht ausreichend numerische Werte, statistische Angaben, Wahrscheinlichkeiten und Verteilungen nur zu bestimmen. Ein ebenso wichtiger Schritt ist es, diese den Entscheidungsträgern vermitteln zu können. Risikovorwarnung ist nicht nur abhängig von der Zuverlässigkeit der Risikoabschätzung, sondern auch von der Effektivität mit der die Informationen an die Entscheidungsträger vermittelt werden. Die Daten müssen in klare und effektive Visualisierungen übersetzt werden, so dass sie unter Umständen auch Laien verstehen und ihrer Entscheidung zugrunde legen können. Studien haben gezeigt, dass das Design und der Aufbau von unsicheren Daten einen großen Einfluss darauf haben, wie die Menschen die Informationen verstehen und verwenden (vgl. Kapitel 3). Allerdings gibt es bisher kaum empirische Daten aus Studien mit einem breiteren Spektrum der Bevölkerung. Insbesondere fehlen Daten für den Umgang von Laien mit unsicheren Daten bei Wettervorhersagen.

Im Rahmen dieser Bachelorarbeit soll daher untersucht werden, wie Wahrscheinlichkeitsverteilungen verständlich kommuniziert werden können, sodass Laien intuitiv mit Unsicherheiten umgehen können. Ziel ist es, mit Hilfe einer experimentellen Untersuchung herauszufinden, ob bestimmte Darstellungen und Visualisierungen besser geeignet sind als andere, um die Schwierigkeiten beim Vermitteln von Unsicherheiten an Laien zu überwinden. Dazu wurde zuerst eine Auswahl an möglichen Visualisierungen für Unsicherheiten identifiziert und klassifiziert. Aus dieser Auswahl wurden dann, mit Hilfe

## 1. Einleitung

---

einer vorausgehenden Umfrage, vier Darstellungen ausgewählt. Daraufhin wurde ein Spiel entwickelt, welches Entscheidungen auf Grund von dargestellten Wahrscheinlichkeiten erfordert. Die im Spiel auftretenden Wahrscheinlichkeitsverteilungen wurden dem Spieler dann mit Hilfe der vier, in der Umfrage identifizierten, Darstellungen veranschaulicht. Abhängig von diesen Informationen musste der Spieler seine Entscheidungen im Spiel treffen. Diese Entscheidungen wurden ausgewertet, um Rückschlüsse auf die Eignung der Visualisierungen ziehen zu können. Die Daten verschiedener Spiele und Teilnehmer wurden zum einen öffentlich und zum anderen unter kontrollierten Bedingungen gesammelt.

Die Arbeit ist wie folgt aufgebaut. Zunächst findet in Kapitel 2 eine thematische Eingrenzung statt und die Problemstellung wird näher erläutert. In Kapitel 3 wird auf die bereits vorhandenen verwandten Arbeiten eingegangen. Dabei werden Bedarf, Schwierigkeiten und Darstellungen der Kommunikation von Unsicherheiten, sowie die Interpretation der Öffentlichkeit von unsicheren Prognosen, genauer betrachtet. Daraufhin werden in Kapitel 4 mögliche Darstellungen für die Kommunikation von Unsicherheiten identifiziert, klassifiziert und anschließend mit Hilfe einer Online-Umfrage evaluiert. Vier Darstellungen werden für das eigentliche Experiment ausgewählt. Kapitel 5 umfasst das Konzept und die Anwendung für die experimentelle Untersuchung. Hierbei werden der Spielablauf, das Konzept und die Implementierung des Spiels beschrieben. Anschließend werden in Kapitel 6 die Auswertung und die Ergebnisse der Daten des öffentlichen Experiments behandelt. In Kapitel 7 werden die Ergebnisse des kontrollierten Experiments dargestellt. Zuletzt enthält Kapitel 8 eine kurze Zusammenfassung der Arbeit sowie ein Fazit.

## 2. Problemstellung

Die Untersuchung von Visualisierungsalternativen kann nicht kontextfrei geschehen. Daher findet im folgenden eine thematische Eingrenzung statt.

### 2.1. Thematische Eingrenzung

Abhängig vom Kontext und der Art des Risikos sind unterschiedliche Darstellungen und Visualisierungen vorteilhaft, um bestimmte Risiken zu kommunizieren [Cou06]. Daher ist es sinnvoll den behandelten Risikobereich und -kontext klar abzugrenzen und nur ein Risikogebiet zu behandeln. Da es weiterhin auch für die experimentelle Studie dieser Arbeit zweckmäßig war ein Thema zu finden, wird hier die Kommunikation meteorologischer Risiken am Beispiel von Wettervorhersagen thematisiert.

Bei der Kommunikation von meteorologischen Risiken mit Hilfe von Wettervorhersagen liegen im Vergleich zu anderen Prognosen und Risiken verschiedene interessante Ansatzpunkte für die Forschung vor. Da Wettervorhersagen schon sehr lange erstellt werden, haben die Menschen ausführliche und differenzierte persönliche Erfahrungen mit Wettervorhersagen und dem anschließend eintretenden Wetter gemacht. Wettervorhersagen sind den Menschen bereits sehr vertraut und sie wenden die Informationen aus den Wettervorhersagen häufig an.

Ein weiterer Grund für die Auswahl von Wettervorhersagen als Untersuchungsgegenstand ist, dass es Wettervorhersagen mit Angaben zur Wahrscheinlichkeit des eintretenden Wetters noch nicht überall und nicht überall gleich lange gibt [GG05]. So sind Menschen, denen beispielsweise die Angabe einer Regenwahrscheinlichkeit nicht vertraut ist, unsicher bei der Interpretation der dargestellten Wahrscheinlichkeit und deuten die Information oft falsch [GG05].

Außerdem verwenden viele Menschen Wettervorhersagen jeden Tag und beziehen sie in ihre täglichen Entscheidungen ein. Die Menschen werden über verschiedene Informationsplattformen wie beispielsweise Printmedien, das Fernsehen sowie das Internet mit unterschiedlichen Wettervorhersagen konfrontiert. Je nach Medium und Anbieter bekommen die Menschen verschiedene Formate und Darstellungen präsentiert und müssen deshalb die Informationen immer wieder neu interpretieren. Dies könnte für die experimentelle Studie von Vorteil sein, da den Probanden bereits bekannt ist, dass jede Wettervorhersage Unsicherheiten beinhaltet [REM08]. Risiken, Unsicherheiten und Wahrscheinlichkeiten sind für viele Menschen ein abstraktes Themengebiet. Mit Hilfe der Wettervorhersage soll das Gebiet greifbarer und verständlicher werden, sowie einen Alltagsbezug erhalten.

Wie bereits erwähnt, sind die Menschen an unterschiedliche Darstellungen von Wettervorhersagen gewöhnt und sind in der Lage sich an neue Darstellung anzupassen. Werden Darstellung falsch

## 2. Problemstellung

---

interpretiert oder schlecht verstanden, liefert dies einen Hinweis darauf, dass die Darstellung an sich ungeeignet ist und nicht, dass diese dem Probanden unbekannt ist.

In dieser Arbeit liegt das Hauptaugenmerk auf Laien, welche einen bedeutenden Anteil der Zielgruppe von Wettervorhersagen ausmachen. In diesem Zusammenhang waren die folgenden Fragen von besonderem Interesse: Wie vertraut sind Menschen mit verschiedenen Typen von Darstellungen für Wettervorhersagen mit Unsicherheiten? Wie einfach fällt es den Menschen die Darstellungen zu verstehen und zu interpretieren? Wie schnell können die Menschen die Darstellungen interpretieren? Welche Darstellungen bevorzugen Menschen um Prognosen mit Unsicherheiten zu erhalten? Inwiefern bevorzugen Menschen deterministische Prognosen<sup>1</sup> den unsicheren Prognosen? Und leiten Menschen Unsicherheiten in deterministische Prognosen ab?

### 2.2. Wettervorhersagen

Da die Atmosphäre ein dynamisches System ist, das nur eine begrenzte Vorhersagbarkeit aufweist, wissen Meteorologen schon seit den frühen Tagen der modernen Wettervorhersage, dass die Prognosen immer ein gewisses Maß an Unsicherheit enthalten werden.

Um das dynamische System der Atmosphäre vollständig zu erfassen, muss einerseits deren initialer Zustand bekannt sein, andererseits ihre zukünftige Entwicklung vorausgesagt werden. Ist nur einer dieser beiden Faktoren fehlerhaft, kann keine sichere Wettervorhersage erstellt werden. Die zwei hauptverantwortlichen Gründe für die Unsicherheiten bei Wettervorhersagen sind zum einen Unsicherheiten beim Erfassen und der Analyse des initialen Zustand der Atmosphäre und zum anderen Annäherungen im Model, die benutzt werden, um die Zukunft vorzuschätzen. Fehler oder Unsicherheiten in der Analyse des initialen Zustands der Atmosphäre können beispielsweise durch zu wenig Beobachtungs- und Messpunkte entstehen oder durch Limitierungen des Datenintegrationssystems verursacht werden. Fehler in den Modellen, die benutzt werden, um die Zukunft vorzuschätzen entstehen vor allem durch die Notwendigkeit die hochkomplexen tatsächlichen Zustände durch Annäherungen bei der Formulierung des Models zu vereinfachen. Diese kleinen Abweichungen können sich aufgrund der Nichtlinearität der atmosphärischen Entwicklung im Laufe der Vorhersage verstärken und zu großen Fehlern führen [Myl02].

Durch präzisere Modelle, bessere Systeme und leistungsstarke Hardware werden die Wettervorhersagen jedoch immer präziser. Eine präzise Wettervorhersage wird erst dadurch wertvoll, dass Verbraucher sie anwenden und eine Entscheidungen aus ihr ableiten können. Ein essentieller Bestandteil bei der Erstellung von Vorhersagen ist daher nicht nur deren Entwicklung, sondern auch die Kommunikation.

<sup>1</sup>deterministische Prognosen sagen einen einzelnen zukünftigen Zustand voraus ohne Informationen über die Unsicherheiten zu geben.

## 3. Verwandte Arbeiten

Arbeiten, die einen ähnlichen Forschungsgegenstand zum Inhalt haben und deren Ergebnisse in dieser Arbeit verwendet wurden, werden hier vorgestellt und diskutiert. Dieses Kapitel beschäftigt sich insbesondere mit verwandten Arbeiten zu den Themen Bedarf und Schwierigkeiten bei der Kommunikation von Unsicherheiten, Interpretation von unsicheren Daten und der Evaluation von möglichen Darstellungen für unsichere Daten.

### 3.1. Bedarf der Kommunikation von Unsicherheiten

Bei vielen Prognosen, die einer breiten Masse zur Verfügung gestellt werden, wie beispielsweise Wettervorhersagen oder Fahrpläne von öffentlichen Verkehrsmitteln, werden vorhandene Unsicherheiten nicht kommuniziert. Stattdessen werden die Prognosen der Öffentlichkeit deterministisch dargestellt. Dies liegt häufig daran, dass nicht eindeutig geklärt ist, ob diese Informationen die Entscheidungsfindung beeinflussen und inwiefern die Entscheidungsträger einen Mehrwert aus dargestellten Unsicherheiten erhalten. Um diese Fragestellung zu untersuchen, führten Mark S. Roulston et. al [MSR06] ein Experiment mit Studenten durch, bei dem die Studenten ein Entscheidungsspiel spielen sollten. Die Spieler erhielten Wettervorhersagen, anhand welcher sie entscheiden mussten, ob sie die Straßen streuen sollten oder nicht. Ein Teil der im Spiel verwendeten Prognosen enthielt Angaben zu Unsicherheiten, der andere Teil beinhaltete keine Angaben zu Unsicherheiten. Sie fanden heraus, dass Laien bessere Entscheidungen treffen, das heißt ihren erwarteten Gewinn bei einer gleichzeitigen Verringerung des Risikos erhöhen konnten, wenn ihnen Prognosen mit Angaben über Unsicherheiten zur Verfügung gestellt werden.

Susan L. Joslyn et. al [JL12] kamen bei sehr ähnlichen Experimenten zu den gleichen Ergebnissen. Die Angabe von Unsicherheiten erhöhte die Entscheidungsqualität als auch das Vertrauen in die Vorhersage. Teilnehmer, denen Informationen über die Unsicherheiten zur Verfügung gestellt wurden, trafen bessere Sicherheitsvorkehrungen, wenn sie nötig waren und weniger wenn sie unnötig waren, als die Teilnehmer ohne Informationen über die Unsicherheiten. Die besten Entscheidungen wurden getroffen, wenn die Teilnehmer mit einer Entscheidungshilfe und Unsicherheitsabschätzungen versorgt wurden.

Wardekker et. al [JAW08] befragten in einem Workshop 118 Entscheidungsträger und Wissenschaftler zu Ihrer persönlichen Meinung zum Thema Unsicherheit. Sie sollten grundsätzlich bewerten ob Unsicherheiten für Ihre Arbeit von Nöten sind oder ob man versuchen sollte Unsicherheiten komplett zu vermeiden. Die Teilnehmer spalteten sich in zwei Gruppen von etwa gleicher Größe. Die eine Gruppe fand, dass Unsicherheiten nicht erwünscht, aber leider unvermeidbar sind. Die zweite Gruppe

### 3. Verwandte Arbeiten

---

befürwortete Unsicherheiten und meinte, dass sie Chancen und Möglichkeiten bietet. Nur ein verschwindend kleiner Teil fand, dass man Unsicherheiten komplett vermeiden sollte. Ein Großteil der Befragten war außerdem der Meinung, dass die Kommunikation von Unsicherheiten Informationen lieferte, die wichtig sind um Entscheidungen zu fällen, wissenschaftliche und soziale Debatten zu führen und einen Mehrwert für ihre eigene Arbeit bringt. Wardekker et. al hielten fest, dass Informationen zu Unsicherheiten für die Öffentlichkeit klar, verständlich, differenziert und präzise sein müssen. Sie sollten aber nicht jede Nuance von Unsicherheit enthalten und sich stattdessen auf die relevantesten Informationen beziehen.

Weiterhin haben Morss et. al [REM08] in den USA ein bundesweite Befragung mit 1465 Teilnehmern zu den bevorzugten Angaben bei Wettervorhersagen durchgeführt. Die Teilnehmer sollten angeben, ob sie eine deterministische, probabilistische Wettervorhersage, beides oder keine von beiden bevorzugen. 22 Prozent der Befragten gaben an lieber eine deterministische Wettervorhersage (z.B. 76 °F) zu bekommen, aber über doppelt so viele, nämlich 54 Prozent der Teilnehmer, favorisierten eine probabilistische (z.B. zwischen 74 °F und 78 °F) Wettervorhersage. In Summe bevorzugten also über 70 Prozent der Teilnehmer Vorhersagen mit Angaben zur Unsicherheit. Außerdem wurden die Teilnehmer zu ihren Präferenzen bei bimodalen Verteilungen (80% Wahrscheinlichkeit von 85 °F und 20% Wahrscheinlichkeit von 70 °F) der Wettervorhersage befragt. Den Teilnehmern wurde erklärt, dass diese ungewöhnliche Verteilung durch eine eventuelle Kaltwetterfront zustande kommt, die einen enormen Temperatursturz hervorrufen würde. In diesem Fall präferierten 90% der Teilnehmer eine nicht deterministische Vorhersage und wollten weitergehende Informationen über die Wahrscheinlichkeiten.

## 3.2. Schwierigkeiten bei der Kommunikation von Unsicherheiten

Folgende Schwierigkeiten müssen bei der Kommunikation von quantitativen Informationen, insbesondere im Zusammenhang mit Wahrscheinlichkeiten, beachtet werden:

### 3.2.1. Rechenschwächen

Rechenschwäche, also fehlendes arithmetisches Verständnis, ist bei vielen Erwachsenen weit verbreitet. So haben viele Erwachsene Probleme bei der Lösung einfachster mathematischer Probleme. Ungefähr 50% der Amerikaner können das Trinkgeld nicht genau berechnen [KI93]. Fast ein Viertel aller Erwachsenen mit College-Abschluss wissen nicht was ein höheres Risiko ist: 1%, 5% oder 10% [LI01]. Dieses Defizit stellt höhere Anforderungen an die Kommunikation von Unsicherheiten, denn akkurate Angaben zu Risiken, ausschließlich mit Hilfe von Wahrscheinlichkeitsangaben zu machen, ist in so einem Fall nicht ausreichend. In diesen Fällen ist es beispielsweise nicht genug in einem Wetterbericht einfach die Regenwahrscheinlichkeit von 30% anzugeben, denn diese Information kann von diesen Menschen nicht verstanden und somit nicht in ihren Entscheidungsprozess eingebunden werden. Visualisierungen, die die Risiken beispielsweise über Farben, Verhältnissen von Flächen oder Höhen von Balken darstellen, können helfen diese Probleme zu überwinden.



### 3.2.2. Qualitative vs. quantitative Angaben

Risiken und Vorteile können qualitativ oder quantitativ kommuniziert werden. Während quantitative Angaben eine eindeutige Einschätzung des Risikos, ohne Möglichkeiten zur Interpretation, vermitteln, bieten qualitative Beschreibungen einen Interpretationsspielraum und machen die objektive und exakte Angabe eines Risikos schwierig.

Während bei der quantitative Angabe „das Regenrisiko beträgt 1%“ oder „in 1 aus 100 Fällen wird es morgen regnen“ das Risiko auf Regen exakt bestimmt wird, stellt eine qualitative Angabe wie beispielsweise „es besteht für morgen ein geringes Regenrisiko“, keine genaue Angabe dar. Die Interpretation einer quantitativen Angabe ist stark vom subjektiven Empfinden des Einzelnen beeinflusst. Während für einen Menschen ein Regenrisiko von 1% ein „geringes Regenrisiko“ darstellt, kann es sein, dass für einen anderen Menschen 5% oder sogar 10% ein „geringes Regenrisiko“ bedeuten [WT86]. Wie man sieht, liegt ein signifikantes Problem darin, dass es einen Mangel an Übereinstimmung gibt, was die quantitativen Begriffe wie „geringes Risiko“ exakt bedeuten.

Weiterhin haben Forschungen [BD03] [SL09] im Bereich von gesundheitlichen Risiken gezeigt, dass Laien eine erhöhte, teilweise übersteigerte Wahrnehmung von Risiken zeigten, wenn nur verbale, qualitative Angaben (sehr häufig, häufig, ungewöhnlich, selten, sehr selten) gemacht wurden. Das Ergebnis war, dass die Probanden eine signifikant höhere Einschätzung ihres gesundheitlichen Risikos hatten und eine signifikant geringere Wahrscheinlichkeit aufwiesen, die Medikamente einzunehmen. Diese Probleme der Interpretation traten nicht nur mit diesem einen verbalen Deskriptor auf. Auch andere Risikodeskriptoren wurden untersucht und wiesen ähnliche Probleme der Missinterpretation auf.

### 3.2.3. Menge an Angaben

Auch die Menge und Qualität von Informationen beeinflusst Entscheidungen. Im Rahmen einer Studie zur Entscheidungsfindung bei der Wahl der Therapiemethode für Brustkrebs kamen Zikmund-Fisher et. al [ZFB08] zu dem Ergebnis, dass wenige, qualitativ hochwertige, vorselektierte Daten, oft zu besseren Entscheidungen führen, als wenn den Probanden zu viele, ungefilterte Informationen auf einmal dargestellt werden.

Normalerweise wird den Brustkrebspatienten jede theoretisch mögliche Behandlungsmethode gezeigt, wobei jedoch im Normalfall nur zwei der möglichen Methoden sinnvoll sind. In ihrer Untersuchung zeigten Zikmund-Fisher et. al [ZFB08] den Patienten nicht mehr alle theoretisch möglichen Behandlungsmethoden, sondern nur noch zwei passende. Die Forscher stellten fest, dass die Patienten die unterschiedlichen Methoden besser und schneller verstanden, wenn sie weniger Optionen und Wahlmöglichkeiten zur Verfügung hatten.

### 3.2.4. Gewinn- vs. Verlust-Formulierung

Die meisten Risiken und Vorteile können auf zwei Arten formuliert werden. Einerseits als positive Gewinn-Formulierung wie beispielsweise: „es besteht eine neunzigprozentige Überlebenschance“, andererseits als negative Verlust-Formulierung wie: „ die Sterblichkeitsrate beträgt zehn Prozent“.

### 3. Verwandte Arbeiten

---

Untersuchungen [DA84] [AA03] [BS95] [RA99] [RA97] [RA93] haben gezeigt, dass Entscheidungsträger von der Auswahl dieser Formulierung beeinflusst werden, obwohl sie objektiv das gleiche ausdrücken.

In einer Studie zu dieser Problemstellung haben beispielsweise McNeil et. al [MB82] Patienten, Doktoranden und Physiker gebeten sich vorzustellen, sie hätten Lungenkrebs. Anschließend mussten sich die Probanden, für eine Behandlungsmethode, entweder einen chirurgischen Eingriff oder eine Strahlungstherapie, entscheiden. Als Entscheidungshilfe wurde ihnen eine Beschreibung jeder Behandlungsmethode mit kumulativen Wahrscheinlichkeiten und Angaben zur Lebenserwartung zur Verfügung gestellt. Die Autoren änderten bei einem Teil der Teilnehmer die Beschreibung des chirurgischen Eingriffs ab und stellten die Überlebensrate als Sterblichkeitsrate dar. Unabhängig von der Teilnehmerklasse wurde der chirurgische Eingriff häufiger gewählt, wenn die Überlebenswahrscheinlichkeit positiv, als neunzig prozentige Überlebensrate dargestellt wurde. Bei der negativen Beschreibung, in Form einer zehn prozentigen Sterblichkeitsrate, entschieden sich mehr Probanden gegen den chirurgischen Eingriff.

#### 3.2.5. Absolute vs. relative Veränderung des Risikos

Häufig ist es nötig, nicht nur ein Risiko, sondern auch dessen Veränderung, wie etwa eine Verringerung des Regenrisikos, mitzuteilen. Hierbei kann die Größe der Veränderung des Risikos entweder als absolute oder relative Angabe formuliert werden. Eine absolute Formulierung könnte beispielsweise lauten: „Das Regenrisiko hat sich von 10% auf 5% verringert.“ Die entsprechende relative Formulierung, „das Regenrisiko hat sich um 50% verringert“ hat dieselbe Bedeutung.

Dass die Wahl der Formulierung Einfluss auf Entscheidungen hat, wurde von Malenka et. al [MD93] und Forrow et. al [FL92] gezeigt. Diese stellten bei ihren Untersuchungen fest, dass die relative Formulierung einer Änderungsrate eine größere Änderung suggerierte als eine absolute Formulierung.

Malenka et. al [MD93] fanden bei einer Befragung von 470 Probanden heraus, dass wenn die Vorteile des Medikaments im relativen Format dargestellt wurden, den Probanden der Vorteil des Medikaments größer erschien und einer medikamentösen Behandlung eher zugestimmt wurde, als wenn der Vorteil in absoluten Zahlen dargestellt wurde. Die Ergebnisse sind über alle Altersgruppen hinweg ähnlich ausgefallen. Der Bildungsstand machte dabei keinen Unterschied.

#### 3.2.6. Natürliche Frequenzen vs. Prozentangaben

Vorstellbar ist ebenfalls, dass es für Entscheidungsprozesse eine Rolle spielt, ob Angaben über unsichere Daten als natürliche Frequenz, wie „an einem von zehn Tagen“ oder in Prozent, als „zehn Prozent der Tage“ dargestellt werden. Um herauszufinden, welche Angabe zu besseren Entscheidungen führt, wurden verschiedene Studien [UH00] [GG03] [SP00] durchgeführt. Dabei stellten die Forscher fest, dass die Wahl der Angabe die Wahrnehmung des Risikos bei Menschen beeinflussen kann, [SP00].

Ulrich Hoffrage et. al [UH00] fanden in mehreren Experimenten, mit jeweils unterschiedlichem Kontext und verschiedenen Teilnehmern heraus, dass sowohl Experten, als auch Laien Statistiken und unsichere Daten besser verstehen, wenn sie in natürlichen Frequenzen ausgedrückt werden.

Signifikant war der Unterschied vor allem bei der Angabe von bedingten Wahrscheinlichkeiten also, wenn statistische Inferenz auf Basis von unsicheren Daten betrieben werden muss, um eine korrekte Schlussfolgerung zu ziehen und damit die richtige Entscheidung zu treffen. Als Erklärung für dieses Phänomen geben Ulrich Hoffrage et. al [UH00] an, dass bei natürlichen Frequenzen einerseits weniger Berechnungen nötig sind, um Schlüsse zu ziehen und andererseits natürliche Frequenzen eher dem entsprechen, wie Menschen im Laufe der Geschichte statistische Informationen erfahren und verwendet haben. Ein weiterer Vorteil dieser Darstellung ist laut Gigerenzer et. al [GG03], dass bei der Darstellung von natürlichen Frequenzen die Referenzklasse, auf die sich eine Angabe bezieht, implizit mit angegeben wird und so weniger Verwirrungen entstehen.

#### 3.2.7. Zeitspanne

Des Weiteren kann die gewählte Zeitspanne die Risikowahrnehmung erheblich beeinflussen. Vielen Menschen fällt es schwer ihre Risikowahrnehmung längeren Zeitspannen anzupassen. Untersucht wurde dies [SDM82] anhand der Bereitschaft von Probanden sich beim Autofahren einen Sicherheitsgurt anzulegen. Die Probanden waren häufiger bereit, einen Sicherheitsgurt zu tragen, wenn sie darüber informiert wurden, dass das Risiko im Laufe des Lebens beim Autofahren ohne Sicherheitsgurt eine schwere Verletzung zu erleiden bei 33% lag. Wenn man ihnen hingegen nur das sehr viel kleinere Verletzungsrisiko einer einzelnen Fahrt nannte, sank die Bereitschaft ihren Sicherheitsgurt anzulegen deutlich.

### 3.3. Interpretation der Öffentlichkeit von unsicheren Prognosen

Bei der Vermittlung von Prognosen an die Öffentlichkeit muss bedacht werden, dass die Menschen über unterschiedlich ausgeprägtes Vorwissen verfügen, vielfältige Erfahrungen gemacht haben und unterschiedliche Bildung genossen haben, wodurch sehr viele verschiedene Interpretationen derselben Prognose entstehen können.

Um unter anderem zu untersuchen, wie die Öffentlichkeit Prognosen, insbesondere jene mit Informationen zu Unsicherheiten und Wahrscheinlichkeiten, auffasst und interpretiert, führten Morss et. al [REM10] eine Umfrage mit 1465 Teilnehmern durch. Ähnlich wie in der vorliegenden Arbeit wurde zudem untersucht, wie die Prognoseinformationen in die Entscheidungsfindung einbezogen werden. Um dies herauszufinden, wurden empirische Daten aus Entscheidungsszenarien benutzt, die in einer landesweiten US-Umfrage gewonnen wurden.

Die Resultate wiesen darauf hin, dass Menschen unterschiedliche persönliche probabilistische Schwellenwerte haben, ab denen sie Schutzmaßnahmen gegen die in der Studie genannten Gefahren einleiten würden und, dass der Kontext und die Darstellung der Informationen die Nutzung der Prognose beeinflussen. Dieses Ergebnis erhielten sie, indem sie nach den Schwellenwerten für die Einleitung von Schutzmaßnahmen in zwei völlig verschiedenen von einander unabhängigen Szenarien fragten. Die dabei ermittelten durchschnittlichen Schwellenwerte der Probanden für die beiden Szenarien unterschieden sich signifikant voneinander.

### 3. Verwandte Arbeiten

---

Die festgestellten Ergebnisse deuteten außerdem darauf hin, dass viele Menschen Unsicherheiten in deterministische Prognosen ableiten, um sie ihrem Entscheidungsprozess zu Grunde zu legen. Vielen Teilnehmern ist es möglich, die verwendeten Darstellungen so weit zu interpretieren, dass sie diese in die Entscheidungsfindung einbeziehen konnten und einen Mehrwert daraus gewannen.

Außerdem analysierten Morss et. al die Daten der Teilnehmer mit unterschiedlichen Interpretationen der Regenwahrscheinlichkeiten und verglichen diese mit den Daten, die die unterschiedlichen Schwellenwerte für Schutzmaßnahmen der Teilnehmer beschrieben. Dabei gelang es Morss et. al Rückschlüsse auf die Beziehung zwischen der Nutzung der Prognoseinformationen und der Interpretation von Unsicherheiten allgemein und Wettervorhersagen des Niederschlages im speziellen zu ziehen. Im Gegensatz zu anderen Studien fand er einen Zusammenhang, der allerdings sehr komplex ist und weiterer Untersuchung Bedarf.

Murphy et al. [GM10] schlussfolgerten schon 1980 bei einem Experiment mit Studenten, dass die Kommunikation von Unsicherheiten nicht nur am Verständnis von Wahrscheinlichkeiten scheiterte, sondern häufig ein Missverständnis der Grund für die falsche Interpretation ist. Oftmals war den Probanden nicht bekannt auf welches Ereignis sich eine bestimmte Wahrscheinlichkeitsangabe eigentlich bezieht. In [GG05] gingen Gigerenzer et. al daher der Frage nach, wie Wettervorhersagen von der Öffentlichkeit überhaupt verstanden werden. Es wurde deutlich, dass die richtige Interpretation eines Wetterberichts häufig daran scheiterte, dass den Probanden nicht klar war, auf welches Ereignis sich die angegebene Regenwahrscheinlichkeit eigentlich bezieht. Diese Information wird in der Praxis im Normalfall nicht angeführt, wodurch die Öffentlichkeit gezwungen ist, selbst zu deuten worauf sich die Wahrscheinlichkeitsangabe bezieht. Diese sind leider oft nicht dieselben, welche die Meteorologen vorsahen. Gigerenzer fand heraus, dass vor allem diejenigen Probanden die Wahrscheinlichkeitsangabe falsch interpretierten, in deren Wohnort Wetterberichte nicht oder noch nicht sehr lange mit Angaben zur Regenwahrscheinlichkeit ergänzt werden.

Die Niederschlagswahrscheinlichkeit gibt nach offizieller Definition des National Weather Service [Tul09] „die Wahrscheinlichkeit eines Auftretens (ausgedrückt in Prozent) einer messbaren Menge flüssigen Niederschlags... zu jeder Zeit während einer spezifizierten Zeitperiode in dem Vorhersagegebiet“ an. Das heißt, eine dreißigprozentige Niederschlagswahrscheinlichkeit gibt an, dass, wenn die Wetterbedingungen wie heute sind, in 3 aus 10 Fällen am nächsten Tag wenigstens ein Minimum an Niederschlag fallen wird. Der Großteil der Probanden interpretierte jedoch, dass es in dreißig Prozent des Gebiets sicher regnet oder, dass es dreißig Prozent der Zeit sicher regnen wird. Um Missverständnisse zu vermeiden, empfiehlt Gigerenzer, zusätzlich zur Wahrscheinlichkeitsangabe die gewünschte Interpretation mitzuteilen.

Morss et al. [REM08] stellten sich die Frage, wie die Öffentlichkeit deterministische Angaben bei Wettervorhersagen interpretiert. Vor allem interessierten sie sich dafür, ob den Menschen bewusst ist, dass die Angabe eines einzelnen festen Wertes bei einer Wettervorhersage eigentlich nicht der Wahrheit entspricht, sondern nur eine vereinfachte Darstellung ist. Sie befragten dazu 1465 Menschen und fanden heraus, dass 95% der Teilnehmer bewusst war, dass sich hinter der deterministischen Prognose ein gewisser möglicher Wertebereich verbirgt. Diese Wertebereiche unterschieden sich allerdings deutlich. Bei einer deterministische Vorhersage von 75 °F reichten sie von einer geringen Abweichung von nur +/-1 °F bis zu Abweichungen von +/- 10 °F. Die meisten Befragten gingen von einer Abweichung von +/- 2 °F aus.

Mit weiteren Befragungen fanden sie heraus, dass den meisten Teilnehmern auch bewusst war, dass eine Prognose, die weiter in die Zukunft reicht auch potentiell weniger vertrauenswürdig ist. Weiterhin war den Teilnehmern auch bewusst, dass Wettervorhersagen unterschiedliche Grade an Vertrauenswürdigkeit besitzen, je nachdem ob Angaben zu Temperatur, Niederschlagswahrscheinlichkeit oder Niederschlagsmenge gemacht werden.

## 3.4. Darstellungen für die Kommunikation von Unsicherheiten

Wie bereits in Kapitel 3.2 dargelegt, müssen bei der Kommunikation von Unsicherheiten viele Besonderheiten beachtet werden. Da schon Unterschiede in der Formulierung der Wahrscheinlichkeiten zu abweichenden Interpretationen und Entscheidungen führen, war es sinnvoll zu untersuchen, wie unterschiedliche graphische Darstellungen die Interpretation von Unsicherheiten beeinflussen und welche Darstellung sich zur Kommunikation von Unsicherheiten am besten eignet.

Harald Ibrenk und M. Granger Morgan versuchten deshalb herauszufinden, welche Visualisierungen sich eignen um Unsicherheiten an Laien zu kommunizieren. In einer Studie [HI87] mit 45 Teilnehmern stellten sie neun alternative Darstellungen zur Kommunikation von quantitativen unsicheren Daten vor und baten die Teilnehmer aus den Darstellungen unterschiedliche Angaben wie beispielsweise einen Durchschnittswert zu ermitteln.

Das Resultat war, dass je nach Fragestellung die einzelnen Darstellungen unterschiedlich gut ihre Aufgabe erfüllten. Bei der Frage nach der Abschätzung des Durchschnittswertes für jede der neun Darstellungen, schnitt die einfachste Darstellung, in Form einer Punktangabe des Erwartungswertes auf einem Zahlenstrahl, mit Abbildung des 95%-Erwartungsbereiches mit Hilfe eines Fehlerbalkens, am besten ab. Weiterhin schnitt bei dieser Frage ein modifizierter Box-Plot sehr gut ab, der anstatt des Medians den Durchschnittswert explizit angab. Die kumulative, kontinuierliche Darstellung, in Form eines Liniendiagramms schnitt dagegen verheerend ab. Die Teilnehmer hatten große Probleme aus einer kumulativen Darstellung einen Durchschnittswert abzuleiten.

Die Forscher stellten fest, dass die Teilnehmer, je nach Darstellung verschiedene Strategien entwickelten, um die beste Abschätzung des Durchschnittswertes zu erhalten. Bei der Darstellung eines Histogramms in Form eines Balkendiagrammes und eines Funktionsgraphen wählten die Teilnehmer den höchsten Punkt der Kurve als Durchschnittswert. Bei denjenigen Darstellungen, bei denen der Durchschnittswert nicht auf Grund des Höhenprofils ermittelt werden konnte, hatten die Teilnehmer eine andere Strategie entwickelt. Wurde bei den Darstellungen mit enger einander liegenden gleichgroßen Punkten bzw. Balken für die Kommunikation der Wahrscheinlichkeit gearbeitet, wählten die Teilnehmer die am dichtesten gezeichneten Stellen als beste Abschätzung des Durchschnittswertes.

Trotz des Umstandes, dass der kumulative Funktionsgraph bei der Frage nach der Abschätzung des Durchschnittswertes sehr schlecht abgeschnitten hat, schnitt ebendiese Darstellung, zusammen mit dem Kreisdiagramm, bei anderen Fragestellungen sehr gut ab. Fragten die Forscher nach der Wahrscheinlichkeit, dass ein Wert  $x$  größer als ein Wert  $a$  ( $x > a$ ) ist, oder dass ein Wert  $x$  zwischen zwei Werten  $a$  und  $b$  ( $b > x > a$ ) liegt, präferierten die Teilnehmer den kumulativen Funktionsgraph und das Kreisdiagramm. Verblüffenderweise schnitt das Histogramm in Form eines Balkendiagramms

### 3. Verwandte Arbeiten

---

im Vergleich dazu schlecht ab, obwohl auch bei dieser Darstellung die Wahrscheinlichkeit für diese Fragestellung praktisch direkt abgelesen werden konnte.

Die Befragung wurde einmal mit und einmal ohne Erklärung der Darstellungen durchgeführt. Die vorherige Erklärung der Darstellungen hatte nur einen sehr schwachen Effekt auf das Ergebnis. Erst nach der Erklärung konnten Teilnehmer sich erschließen, dass die Informationen aus der Darstellung in Form einer einfachen Punktangabe des Erwartungswerts auf einem Zahlenstrahl mit Angabe des 95%-Erwartungsbereich nicht dazu ausreicht, um die Frage nach der Wahrscheinlichkeit für  $b > x > a$  zu beantworten.

Die Darstellung des kumulativen Funktionsgraphen der Dichtefunktion alleine hatte bei Teilnehmern mit wenigen statistischen Vorkenntnissen häufig zur Annahme eines falschen Mittelwertes geführt. Eine solche Fehlinterpretation des Ergebnisses war bei Teilnehmern mit ausgeprägteren Statistikkenntnissen beinahe ausgeschlossen. Im Gesamtergebnis zeigte sich aber, dass auch ein gutes statistisches Basiswissen nicht zu einem signifikant besseren Ergebnis in der Entscheidungsfindung führte. Die Autoren empfehlen, zusätzlich zur normalen Wahrscheinlichkeitsdichtefunktion eine kumulative Wahrscheinlichkeitsdichtefunktion mit der gleichen horizontalen Skalierung darzustellen, um die Entscheidungsfindung bestmöglich zu unterstützen. Weiterhin sollte der Mittelwert klar auf beiden Kurven eingezeichnet sein.

Chris Olston et. al beschäftigten sich ebenfalls mit der Frage, wie Unsicherheiten bestmöglich visualisiert werden [CO02]. Sie unterschieden dabei allerdings zwischen statistischer Unsicherheit und beschränkter Unsicherheit. Dabei liegt der statistischen Unsicherheit, prinzipiell eine unbegrenzte Verteilung von Werten zugrunde und sie verfügt meistens über einen wahrscheinlichsten Höhepunkt. Bei der beschränkten Unsicherheit kann dagegen keine Verteilung angenommen werden. Es ist nur bekannt, dass der Wert irgendwo zwischen der exakt gegebenen unteren und oberen Grenze liegt. Liegt eine beschränkte Unsicherheit vor empfehlen Olston et. al keinen einzelnen Fehlerbalken in der gegebenen Darstellung abzubilden, sondern das graphische Element so zu verändern, dass die Grenzen erweitert werden und stattdessen transparent überlappend dargestellt werden. Diese Technik lässt sich unter anderem auf Balkendiagramme, Streudiagramme, Liniendiagramme, Kreisdiagramme und gestapelte Balkendiagramme anwenden.

Wardekker et. al untersuchten [JAW08] wie Probanden die Begrifflichkeiten für Wahrscheinlichkeiten, die beispielsweise von der IPCC (Intergovernmental Panel on Climate Change) im Rahmen einer festen Skala in den meteorologischen Berichten verwendet werden, interpretieren. Die meisten der verwendeten Begriffe riefen die erwünschte Interpretation hervor, jedoch gab es auch Ausnahmen. Beispielsweise stellte sicher der Begriff „medium likelihood“, der einem Wahrscheinlichkeitsspielraum von 33%-66% angeben soll, in der Nutzung und Übersetzung als schwierig heraus. Als der Begriff in „about as likely as not“ geändert wurde, ergaben sich sehr viel bessere Ergebnisse.

Morss et al. [REM08] untersuchten unter anderem auch, welche Abbildung bei textuellen Darstellungen von unsicheren Daten bezüglich des Niederschlages bevorzugt werden. Dabei stellte sich heraus, dass Prozentangaben wie zum Beispiel 20% oder 50% und nicht-numerische Angaben wie beispielsweise „es besteht eine geringe Chance“, mit großem Abstand am beliebtesten waren. Relative Häufigkeiten wie „in 1 aus 5 Fällen“ oder in „1 aus 2 Fällen“ wurden nur von einer Minderheit gewählt. Noch weiter abgeschlagen war die Angabe der Wahrscheinlichkeit als Chance, wie etwa „eine Chance von eins zu vier“ oder „eine Chance von eins zu eins“.

Während sich die bisher vorgestellten Arbeiten mit der Vermittlung von Unsicherheiten an Laien beschäftigten versuchten Pappenberger et al. [FP13] zu ermitteln, welche Visualisierungen zur Kommunikation von Unsicherheiten unter Experten präferiert werden. 57 Meteorologen bekamen die Aufgabe eine probabilistische 10-Tages-Prognose für Hochwasser für einen einzelnen Ort, mit Hilfe einer beliebigen Darstellung, zu skizzieren. Es ergab sich eine sehr breit gefächerte Menge an vielen verschiedenen Abbildungen. Von den Meteorologen häufig gewählte Darstellungen waren Spaghetti Plots, unterschiedlichste Quantile, Box Plots, vollständige Wahrscheinlichkeitsverteilungen, Min- und Max-Angaben, Mittelwert und Median.

Bei der Darstellung der Unsicherheiten gab es auch unter den Experten keinen klaren Konsens, welche Darstellung die bessere ist, weshalb eine individuelle nutzerdefinierte Einstellung für die Experten als Lösung in Betracht gezogen wurde. Hauptsache sei, dass die gezeigte Darstellung nicht zu Missverständnissen führen kann, da beispielsweise festgestellt wurde, dass einige Teilnehmer die den Spaghetti Plot bevorzugten, diesen falsch interpretierten und ihn deterministisch benutzten. Das heißt, dass sie einen einzigen willkürlich gewählten Pfad am ehesten glaubten und diesen zur Entscheidungsfindung nutzten, obwohl jeder Pfad gleich wahrscheinlich ist.

### 3.5. Zusammenfassung und Diskussion

Die hier vorgestellten Arbeiten zeigen, dass, obwohl es schon eine Reihe von Untersuchungen zur Kommunikation von Wahrscheinlichkeiten gibt, auf diesem Gebiet weitere Forschung benötigt wird. Es wurde bereits eine Reihe der Aspekte der Kommunikation von unsicheren Daten im Kontext der Wettervorhersagen untersucht. Jedoch ist weitere Arbeit nötig, um die Ergebnisse, die seither nur aus Umfragen stammen, mit empirischen Daten aus echten Entscheidungen unter realen Bedingungen zu bestätigen.

Vor allem blieb bisher offen, wie unsichere Daten unter realen Bedingungen bestmöglich vermittelt werden können, damit diese bei Entscheidungsprozessen effektiv eingesetzt werden und wie stark die Entscheidungen von der jeweiligen Darstellung abhängen. Es soll untersucht werden welche Visualisierungsalternative sich am besten zur Kommunikation von unsicheren Daten an Laien eignet. Weiterhin muss bestätigt werden, ob und wie unterschiedlich Entscheidungen ausfallen, wenn sie aufgrund einer deterministischen Darstellung und einer probabilistischen Darstellung getroffen werden. Diese Untersuchungen sollen Meteorologen ein praktisches Verständnis davon vermitteln, was die Nutzer der Prognosen von den Angaben über die Unsicherheit tatsächlich verstehen können und wie sie die Daten für ihre Entscheidungen nutzen. Bisher können die Meteorologen nur annehmen was der Nutzer mögen oder nutzen könnte. Auch andere Experten, die unsichere Daten bestmöglich an Laien vermitteln möchten, könnten die Ergebnisse auf eigene Untersuchungen anwenden und von den Erkenntnissen profitieren.





## **4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten**

Es soll die bestmögliche Kommunikation von Unsicherheiten am Beispiel von Wettervorhersagen untersucht werden. Das Medium zur Durchführung der Untersuchung ist ein Spiel, bei welchem die Teilnehmer anhand von unterschiedlich dargestellten Wettervorhersagen Entscheidungen treffen müssen. Damit im eigentlichen Spiel keine zu große Menge an Darstellungen auf den Teilnehmer trifft, wurde eine, dem Spiel vorausgehende, Online-Umfrage erstellt, um eine Vorauswahl aus möglichen Darstellungsformen für das eigentliche Experiment zu bestimmen.

### **4.1. Art der darzustellenden Daten**

Bei den, in dieser Untersuchung darzustellenden, unsicheren Daten, handelte es sich um diskret gespeicherte, aber eigentlich stetige Werte. In dieser Arbeit waren das Niederschlag, Temperatur und Windstärke. Diese diskreten Werte wurden zusammen mit ihrer Auftrittswahrscheinlichkeit und dem Vorhersagezeitraum gespeichert. Jedes Datum war also ein Tripel: (Wert  $x$ , Wahrscheinlichkeit  $y$ , Zeit  $z$ ). Ein Datensatz bestand aus sehr vielen dieser Tripel. Die Anzahl der Tripel und das Werteintervall zwischen denen sich die Werte befinden konnten, waren dabei theoretisch nicht begrenzt. Die Daten in dieser Arbeit waren immer normalverteilt. Wurden die Wert/Wahrscheinlichkeit-Paare, die zum selben Zeitraum gehören, zusammengefasst und gezeichnet, so ergab sich eine Gaußsche Glockenkurve der Wahrscheinlichkeitsverteilung für den jeweiligen Vorhersagezeitraum.

### **4.2. Identifikation und Klassifikation von Darstellungen**

Mit Hilfe der Umfrage sollten vier Darstellungen für das Spiel identifiziert werden. Es wurden sowohl konventionelle, als auch unkonventionelle Darstellungen in Betracht gezogen. Denn es sollte herausgefunden werden, wie Entscheidungen ausfallen, wenn sie auf Basis von Darstellungen getroffen werden, die Unsicherheiten gar nicht oder kaum darstellen. Im Vergleich dazu, sollte festgestellt werden, wie Entscheidungen ausfallen, die auf Basis von Darstellungen getroffen werden, die Unsicherheiten sehr differenziert und detailliert darstellen. Ziel war es, zu ermitteln, ob es signifikante Unterschiede in der Güte der Entscheidungen abhängig von der Darstellung gibt. Um dies herauszufinden, war es zunächst nötig mögliche Darstellungen zu identifizieren, die Unsicherheiten gar nicht oder nur sehr stark zusammengefasst darstellten und Darstellungen, die Unsicherheiten sehr detailliert veranschaulichten.

#### 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten

---

Dazu wurden gängige Darstellungen zur Visualisierung von Unsicherheiten untersucht. Außerdem wurden konventionelle sowie unkonventionelle Darstellungen aus früheren Untersuchungen [HI87] [FP13] näher betrachtet.

Häufig werden diskrete Wahrscheinlichkeitsverteilungen als Erwartungswert, Box-Plots, Quantile oder Histogramme dargestellt. Bei diesen Darstellungsformen, werden kontinuierliche Wahrscheinlichkeitsdichtefunktionen zu statistischen Kennwerten zusammengefasst oder vereinfacht als diskrete Funktion dargestellt. Wie in Kapitel 3.4 bereits näher beschrieben, ließen Florian Pappenberger et. al [FP13] Meteorologen bei einer Befragung mögliche Darstellungen für eine probabilistische 10-Tages-Prognose für Hochwasser skizzieren. Hierbei entschieden sich die Teilnehmer ebenfalls sehr häufig für das arithmetische Mittel, den Median, Quantile oder Box-Plots.

Harald Ibrek et. al [HI87] untersuchten ebenfalls verschiedene Visualisierungen zur Kommunikation von Unsicherheiten. Sie verwendeten bei ihrer Untersuchung allerdings zusätzlich zu den gängigen Darstellungen, auch unkonventionelle Darstellungsvarianten. Beispielsweise gab es Darstellungen, die die Wahrscheinlichkeit nicht über die Höhe, wie bei einem Histogramm, oder über die Fläche, wie bei einem Kreisdiagramm kommunizierten, sondern über die Dichte von Balken oder Punkten. Dadurch kommen sie völlig ohne Vertikalachse aus. Angelehnt an diese Darstellungen entstand die Darstellung 10 (Abbildung 4.2) der Untersuchung.

Charakteristisch für den Box-Plot ist, dass wichtige Daten der Wahrscheinlichkeitsverteilung, wie Median, Quantile und Extremwerte auf einen Blick erkennbar sind. Das 0,25-Quantil und das 0,75-Quantil markieren die Enden eines Rechteckes (Box), vertikale Linien zeigen die Extremwerte und den Median an.

Ein Quantil gibt den Wert an, den ein definierter Anteil der Beobachtungswerte nicht überschreitet. Alle Werte der Wahrscheinlichkeitsverteilung werden nach ihrer Größe geordnet und anschließend in gleichgroße Klassen, eingeteilt. Der Schwellenwert, also der Übergang zwischen zwei Klassen wird Quantil genannt. Beispielsweis gibt das 0,25-Quantil bzw. 25%-Quantil den Wert an, für den gilt, dass 25% aller Beobachtungswerte kleiner sind als dieser Wert. Wenn bei einer Wahrscheinlichkeitsverteilung, die die Regenmenge für den folgenden Tag angibt, das 0,25-Quantil den Wert 4mm hat, bedeutet dies, dass es in 25% der Fälle morgen höchstens 4 mm regnet. Allgemein wird der Anteil mit  $p$  bezeichnet, also  $p$ -Quantil. Hierbei kann  $p$  eine reelle Zahl zwischen 0 und 1 sein.

Diskret erfasste, aber eigentlich stetige Daten, wie beispielsweise Niederschlag, Temperatur und Windstärke, können als Häufigkeitsdichte repräsentiert und zum Beispiel in einem Histogramm dargestellt werden. Man verwendet Histogramme üblicherweise dann, wenn man den genauen Verlauf einer Häufigkeitsverteilung sehen möchte und nicht nur zusammenfassende Daten wie den Mittelwert oder die Standardabweichung, wie sie beispielsweise der Boxplot zeigt. Aus diesen Überlegungen heraus und mit der Absicht auch textuelle Darstellungen von unsicheren Daten zu berücksichtigen, wurden die folgenden zwölf Darstellungen für die vorgelagerte Umfrage gewählt. Die Darstellungen 1 bis 6 sind in Abbildung 4.1 und die Darstellungen 7 bis 12 in Abbildung 4.2 abgebildet.

- Darstellung 1 textueller Erwartungswert (deterministisch)
- Darstellung 2 textueller Erwartungswert mit Standardabweichung
- Darstellung 3 textuelle Quantilangaben (0.25-, 0.50-, 0.75-Quantile)

	<b>textuell</b>	<b>graphisch</b>
<b>Darstellung der Unsicherheit: keine</b>	Darstellung 1	keine Darstellung
<b>Darstellung der Unsicherheit: zusammengefasst</b>	Darstellung 2	Darstellung 4 Darstellung 5 Darstellung 6
<b>Darstellung der Unsicherheit: detailliert</b>	Darstellung 3	Darstellung 7 Darstellung 8 Darstellung 9 Darstellung 10 Darstellung 11 Darstellung 12

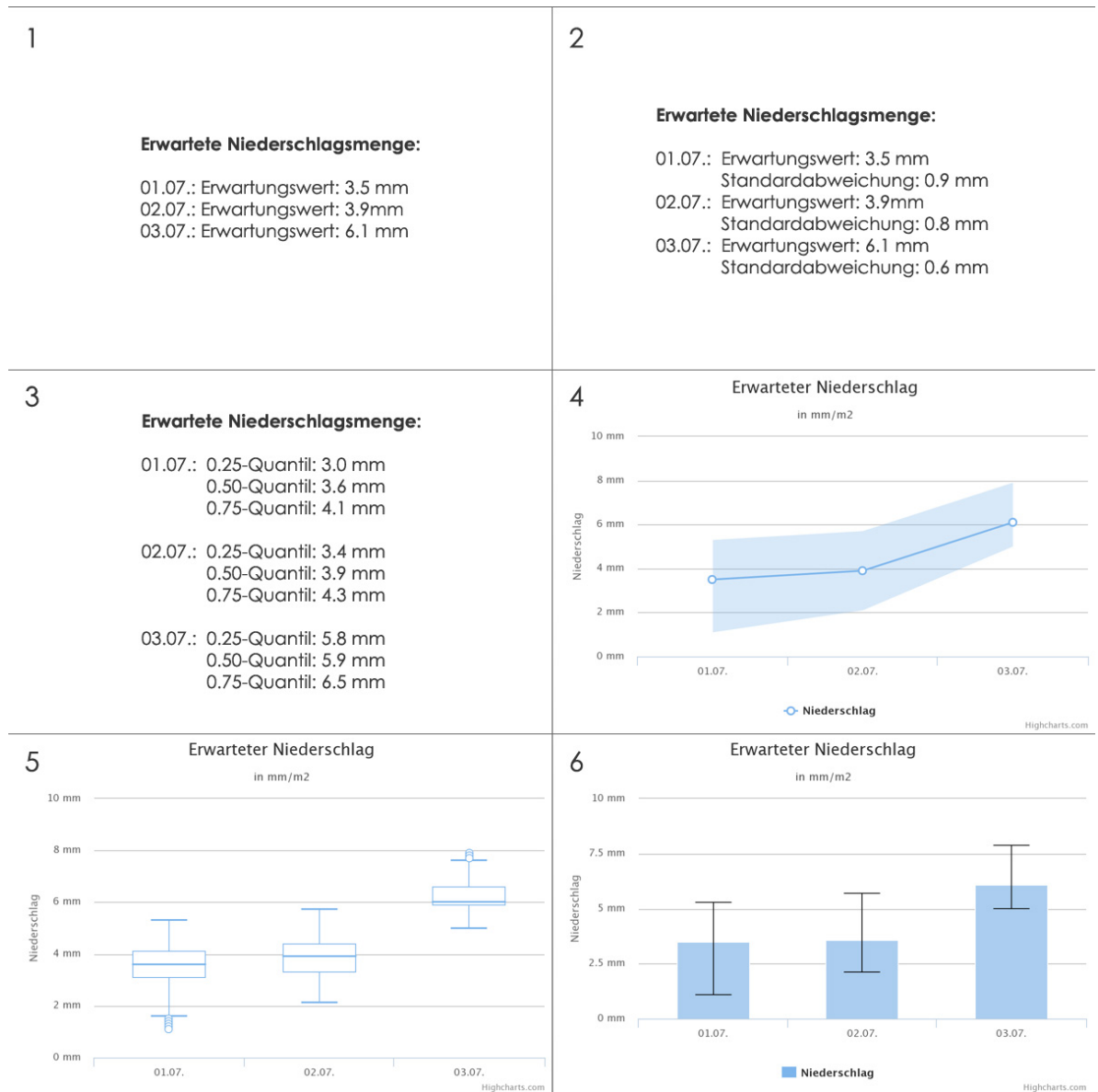
**Tabelle 4.1.:** Klassifikation der 12 Darstellungen nach dem Grad der Darstellung der Unsicherheit.

- Darstellung 4 Liniendiagramm des Erwartungswerts mit Spannweite als Flächendiagramm
- Darstellung 5 Box-Whisker-Plot
- Darstellung 6 Balkendiagramm des Erwartungswert mit Fehlerbalken
- Darstellung 7 Histogramm in Form eines Balkendiagramms
- Darstellung 8 kumulatives<sup>1</sup> Histogramm in Form eines gestapeltem Balkendiagramms
- Darstellung 9 kumulatives Histogramm in Form eines gestapeltem Flächendiagramms
- Darstellung 10 Wahrscheinlichkeitsdichtediagramm ohne Variation in der Höhe, bei dem die Unsicherheit stattdessen über transparente, gleich breite, horizontale Balken, die sich überlagern, kommuniziert wird (angelehnt an [HI87])
- Darstellung 11 kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen
- Darstellung 12 kontinuierliche Darstellung eines kumulativen Histogramms in Form eines Funktionsgraphen

Weiterhin wurde eine Klassifikation herausgearbeitet, in die die zwölf verschiedenen Darstellungen bereits vor Durchführung der Umfrage eingeteilt wurden. Die graphischen und textuellen Abbildungen wurden getrennt voneinander nach der Darstellung der Unsicherheit eingeteilt. Da vor der Durchführung der Umfrage nicht bekannt war, welche der Darstellungen sich am besten eignen, wurden möglichst vielfältige Darstellungen verwendet. Die Klassifikation ist in Tabelle 4.1 zu sehen.

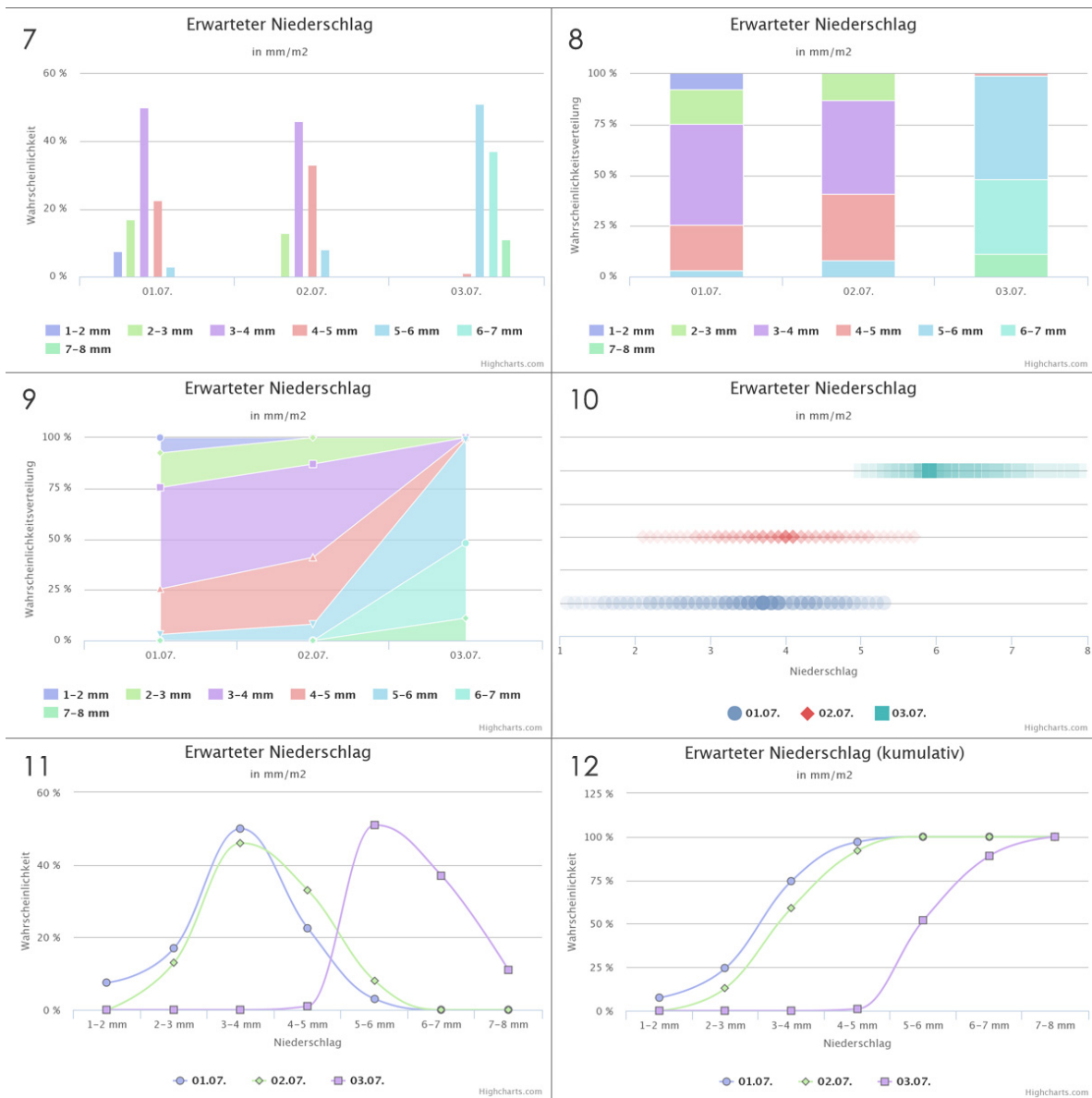
<sup>1</sup>kumulativ: die kumulative Häufigkeit wird aus der Summe der Häufigkeiten der Merkmalsausprägungen von der kleinsten Ausprägung bis hin zu der jeweils betrachteten Schranke berechnet.

## 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten



**Abbildung 4.1.:** Darstellungen 1-6 - Darstellung 1: textueller Erwartungswert, Darstellung 2: textueller Erwartungswert mit Standardabweichung, Darstellung 3: textuelle Quantilangaben (0.25-, 0.50-, 0.75-Quantile), Darstellung 4: Liniendiagramm des Erwartungswerts mit Spannweite als Flächendiagramm, Darstellung 5: Box-Whisker-Plot, Darstellung 6: Balkendiagramm des Erwartungswert mit Fehlerbalken.

## 4.2. Identifikation und Klassifikation von Darstellungen



**Abbildung 4.2.:** Darstellungen 7-12 - Darstellung 7: Histogramm in Form eines Balkendiagramms, Darstellung 8: kumulatives Histogramm in Form eines gestapeltem Balkendiagramms, Darstellung 9: kumulatives Histogramm in Form eines gestapeltem Flächen-diagramms, Darstellung 10: transparentes Wahrscheinlichkeitsdichtediagramm ohne Variation in der Höhe, Darstellung 11: kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen, Darstellung 12: kontinuierliche Darstellung eines kumulativen Histogramms in Form eines Funktionsgraphen.

## 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten

---

Die verschiedenen Abbildungen wurden drei Bereichen zugeordnet. Im ersten Bereich ist nur die Darstellung 1 zu finden. Sie bildet keine Unsicherheiten ab. Die Angabe ist auf einen einzigen deterministischen Wert reduziert.

Dem zweiten Bereich sind Darstellungen zugeordnet, die Unsicherheiten nur zusammengefasst darstellen. Hier sind die Abbildungen 2, 4, 5 und 6 zu finden. Die Daten sind stark zusammengefasst und auf wenige Kennwerte der Statistik, wie Mittelwert, Standardabweichung, Spannweite oder Quantile reduziert.

Der letzte Bereich enthält die Darstellungen 3, 7, 8, 9, 10, 11, und 12, die Unsicherheiten detailliert abbilden. Die enthaltenen Abbildungen lassen eine exaktere Einsicht in den Verlauf der diskreten Wahrscheinlichkeitsverteilung zu. Dabei zeigen die Darstellungen 8, 9 und 12 kumulative Häufigkeiten.

### 4.3. Methodik der Online-Umfrage

In diesem Unterkapitel werden die Befragung, die Vorbereitung und die Veröffentlichung näher beschrieben. Weiterhin werden die demographischen Daten der Umfrage-Teilnehmer zusammengefasst.

#### 4.3.1. Befragung

Zu Beginn der Umfrage wurden die Teilnehmer gebeten, Angaben zu ihrem demographischen Hintergrund zu machen, damit bei der späteren Auswertung aus diesen Merkmalen eventuell Muster oder Zusammenhänge erkennbar werden. Eine Zusammenfassung der demographischen Angaben ist in Kapitel 4.3.3 dargestellt.

Anschließend wurde den Teilnehmern die eigentliche Aufgabenstellung präsentiert: „Stellen Sie sich vor, Sie sind ein Bauer. Sie möchten Pflanzen anbauen und so viele wie möglich davon ernten. Wenn Ihre Pflanzen allerdings nicht genug Wasser bekommen, trocknen sie aus und Sie verlieren sie. Sie wissen für jede Pflanze wieviel Wasser sie mindestens täglich benötigt um zu überleben. Basierend auf einer Wettervorhersage müssen Sie entscheiden welche Pflanzen auf Ihrem Feld angepflanzt werden. Die Wettervorhersage soll Ihre Entscheidungsfindung unterstützen, aber ist nicht völlig zuverlässig. Jede Wettervorhersage ist nur eine grobe Schätzung. Es bleibt eine Restunsicherheit. Die nächsten Seiten werden Ihnen verschiedene textuelle und graphische Darstellungen der Wettervorhersage (nur Niederschlag) für die nächsten drei Tage zeigen. Bewerten Sie diese entsprechend der oben beschriebenen Situation.“

Daraufhin wurden den Teilnehmern, in einer zufälligen Reihenfolge, die in Kapitel 4.2 vorgestellten zwölf verschiedenen Darstellungen nacheinander gezeigt. Die zufällige Reihenfolge wurde gewählt, um Beeinflussungen der Beurteilung durch eine bestimmte Reihenfolge so gut wie möglich zu vermeiden. Die Teilnehmer hatten keine Möglichkeit zu einer schon beantworteten Frage zurückzukehren, dies sollte ausschließen, dass der Teilnehmer durch spätere Darstellungen neue Gesichtspunkte in seine Bewertung miteinbezieht und die Bewertung anschließend ändert.

Jede der zwölf gezeigten Darstellungen beinhaltete eine Wettervorhersage für den Niederschlag der nächsten drei Tage. Alle zwölf Darstellungen basierten auf dem gleichen Datensatz. Drei Darstellungen waren dabei rein textuell. Die restlichen neun Darstellungen waren graphisch aufbereitet. Die graphischen Darstellungen wurden alle mit dem gleichen graphischen System erstellt, damit sie sich in Stil und Qualität möglichst ähnlich waren (s. Kapitel 4.2). Dies sollte verhindern, dass die Teilnehmer subjektive Präferenzen bei der Gestaltung in die Gesamtbewertung einfließen lassen.

Die Teilnehmer sollten jede Darstellung nach folgenden vier Gesichtspunkten auf einer Likert-Skala (1 - trifft nicht zu, 2 - trifft eher nicht zu, 3 - teils-teils, 4 - trifft eher zu, 5 - trifft zu) bewerten:

- Die Darstellung hilft mir eine Entscheidung zu treffen
- Die Darstellung ist mir vertraut
- Die Darstellung ist einfach zu verstehen
- Die Darstellung ist optisch ansprechend

Die Fragen konnten in beliebiger Reihenfolge beantwortet werden und die Bewertung konnte, bis zum Anzeigen der folgenden Darstellung, beliebig häufig geändert werden. Im Anschluss an die Umfrage bekamen die Teilnehmer noch die Gelegenheit einen Kommentar oder eine Anregung zu hinterlassen.

#### 4.3.2. Vorbereitung und Veröffentlichung

Die Umfrage wurde mit dem webbasierten Umfragetool Lime Survey<sup>2</sup> vorbereitet und auf dem Lime Survey Server der HCI Gruppe veröffentlicht. Lime Survey stellt eine Umgebung zur Verfügung in der komplexe Umfragen erstellt und ausgewertet werden können.

Die Umfrage war öffentlich zugänglich, die Teilnehmer wurden nicht selektiert, da der Teilnehmerkreis möglichst groß und vielfältig sein sollte. Deshalb konnte jeder, der über den Link zur Umfrage verfügte teilnehmen. Die Umfrage wurde anonym und auf freiwilliger Basis durchgeführt. Der Link zur Umfrage wurde über soziale Netzwerke wie Facebook, per Email und mündlich verbreitet.

Um sicherzustellen, dass alle Teilnehmer alle Fragen und Antwortmöglichkeiten verstehen, wurde die Umfrage auf Deutsch und auf Englisch erstellt. Die Teilnehmer konnten jederzeit, auch während der Umfrage, die Sprache wechseln. Für die Auswertung der Daten wurden alle Antworten ins Englische übersetzt.

<sup>2</sup><http://www.limesurvey.org/>

#### 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten

Darstellung	Hilft bei Entscheidung	Vertraut	Einfach zu verstehen	Optisch ansprechend	Gesamt
1	3,63	4,29	4,28	2,3	3,63
2	3,68	3,9	3,34	2,13	3,26
3	2,86	2,79	2,37	1,7	2,43
4	4,12	3,74	4,12	4,01	4,00
5	3,22	2,73	2,52	2,48	2,74
6	3,48	3,11	3,08	2,98	3,16
7	3,93	4,21	3,67	3,67	3,87
8	3,47	3,49	3,16	3,52	3,41
9	3,16	2,84	2,74	3,42	3,04
10	3,73	2,31	3,59	3,59	3,31
11	3,89	3,88	3,46	3,6	3,71
12	3,44	3,5	2,88	3,52	3,34

**Abbildung 4.3.:** Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der 12 den Teilnehmern vorgestellten Darstellungen. Bewertungen kleiner als 2.4 wurden in rot, zwischen 2.4 und 3.0 in orange, zwischen 3.0 und 3.6 in hellgrün und größer als 3.6 in dunkelgrün dargestellt.

#### 4.3.3. Teilnehmer

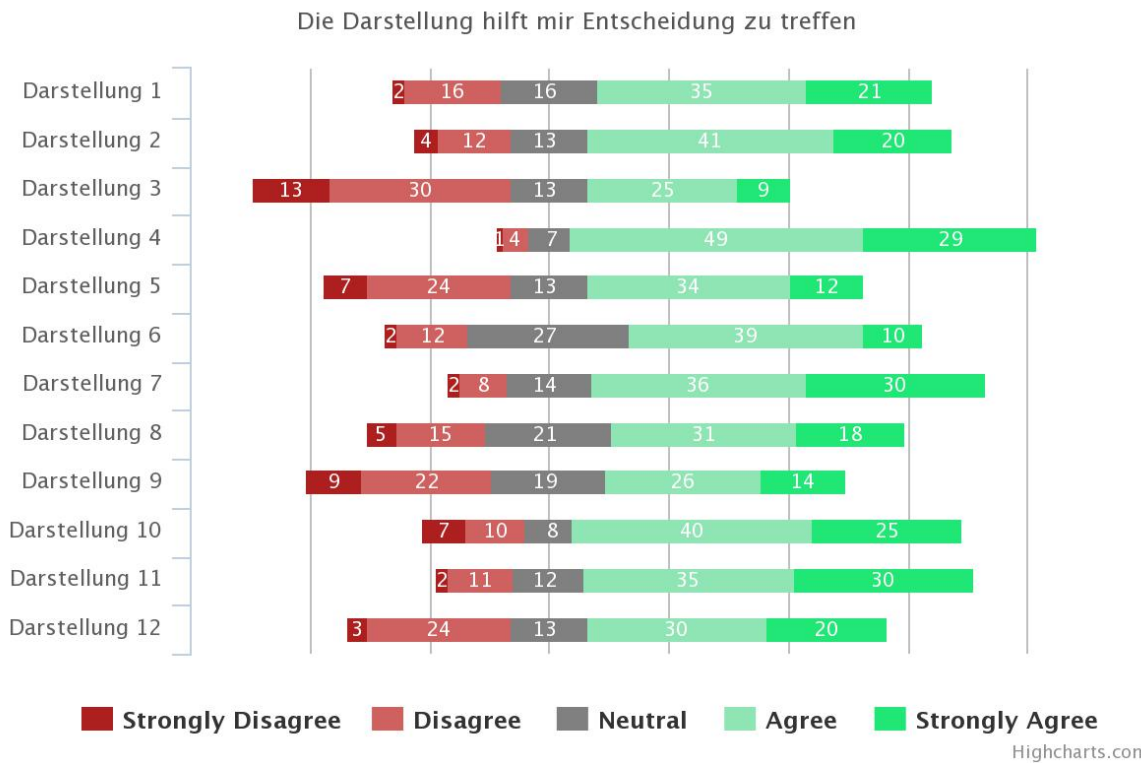
An der Umfrage beteiligten sich insgesamt 90 Probanden. Diese teilten sich auf in 36 Frauen und 54 Männer. Die Teilnehmer waren zwischen 18 und 82 Jahren alt, wobei das Durchschnittsalter bei 30,84 Jahren (SD: 12,61) lag. Als höchsten erreichten Abschluss hatten 5,56% der Teilnehmer die Promotion, 40% bereits ein abgeschlossenes Studium, 27,67% das Abitur, 20% eine abgeschlossene Berufsausbildung, 6,67% den Haupt- oder Realabschluss oder einen vergleichbaren Abschluss und nur 1,11% (noch) keinen Abschluss. Mit 42,22 % stellte die Informatik den am häufigsten angegebenen beruflichen beziehungsweise studentischen Hintergrund dar. Den zweiten Platz belegten Wirtschaftswissenschaften mit 20%. Danach folgten Handel mit knapp 9%, Lehramt mit 6,67%, Mechaniker mit 5,56%, Dienstleistung mit 4,44% und Sonstiges mit 7,78%. 4,44% der Teilnehmer machten keine Angabe zum beruflichen Hintergrund.

#### 4.4. Ergebnisse

Abbildung 4.3 stellt das arithmetische Mittel der Bewertungen für jede Frage und jede Darstellung dar. Eine dunkelgrüne Einfärbung der Zellen markiert einen Durchschnittswert von über 4,0. Bei einer durchschnittlichen Punktzahl von 3,6 bis 4,0 wird die Zelle hellgrün eingefärbt, eine weiße Zelle gibt einen Durchschnittswert von 3,0 bis 3,6 an und eine rote Einfärbung weist auf eine durchschnittliche Punktzahl von weniger als 3,0 hin.

Für die Daten in den folgenden vier Unterkapiteln, wurde jeweils zuerst ein Friedman Test und darauffolgend eine Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test durchgeführt und eine





**Abbildung 4.4.:** Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer half eine Entscheidung zu treffen.

Bonferroni-Korrektur mit  $p < 0.05$  angewendet, was zu einem Signifikanzniveau von  $p = 0.05/66 < 0,0008$  führte.

#### 4.4.1. Entscheidungshilfe

Abbildung 4.4 zeigt für jede der 12 Darstellungen die Verteilung der Antworten der 90 Teilnehmerstimmen bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer half eine Entscheidung zu treffen.

Es wurde ein Friedman Test durchgeführt, der einen statistisch signifikanten Unterschied für die verschiedenen Darstellungen bezüglich der Entscheidungshilfe zeigte,  $\chi^2(11) = 107,757$ ,  $p < 0.000016$ . Die exakten Ergebnisse der Post-hoc Analyse mit Wilcoxon-Vorzeichen-Rang-Test und Bonferroni-Korrektur sind in Tabelle 4.2 zu sehen.

Besonders erwähnenswert ist Darstellung 3. Diese unterscheidet sich zu fast allen Darstellungen signifikant mit der Ausnahme von Darstellung 5, 9 und 12. Sie schnitt in dieser Fragestellung mit Abstand am schlechtesten von allen Darstellungen ab (Durchschnitt: 2,86, SD: 1,26). Weiterhin fallen die Darstellungen 4 und 7 besonders auf. Beide schnitten in dieser Kategorie besonders gut ab

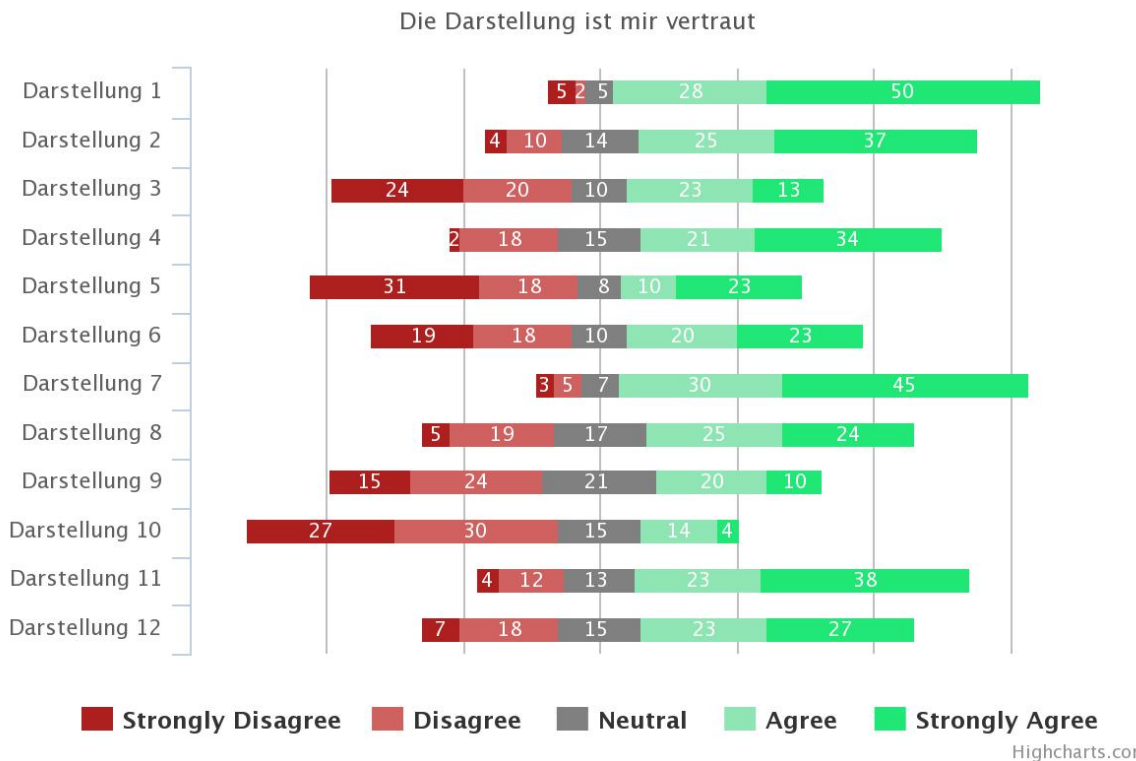
#### 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten

Darstell.	2	3	4	5	6	7	8	9	10	11	12
1	0,608	**	*	0,012	0,315	0,074	0,347	0,013	0,422	0,099	0,289
2	-	***	0,001	0,004	0,174	0,114	0,157	0,003	0,614	0,134	0,122
3	-	-	***	0,007	**	***	**	0,046	***	***	0,001
4	-	-	-	***	***	0,149	**	***	0,015	0,063	**
5	-	-	-	-	0,025	*	0,166	0,745	0,006	*	0,285
6	-	-	-	-	-	0,007	0,951	0,050	0,095	0,009	0,774
7	-	-	-	-	-	-	*	***	0,177	0,683	*
8	-	-	-	-	-	-	-	0,006	0,032	0,002	0,859
9	-	-	-	-	-	-	-	-	**	***	0,041
10	-	-	-	-	-	-	-	-	-	0,312	0,022
11	-	-	-	-	-	-	-	-	-	-	*

**Tabelle 4.2.:** Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer half eine Entscheidung zu treffen, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von  $p < 0,0008$  führte. Statistisch signifikante Werte sind mit einem (\*), zwei (\*\*) oder (\*\*\*) Sternen gekennzeichnet. \* $p < 0,0008$ , \*\* $p < 0,00016$ , \*\*\* $p < 0,000016$

Darstell.	2	3	4	5	6	7	8	9	10	11	12
1	*	***	**	***	***	0,417	***	***	***	0,009	***
2	-	***	0,352	***	***	0,009	0,012	***	***	0,940	0,021
3	-	-	***	0,728	0,072	***	***	0,672	0,003	***	**
4	-	-	-	***	0,001	0,006	0,184	***	***	0,392	0,239
5	-	-	-	-	0,015	***	*	0,431	0,036	***	*
6	-	-	-	-	-	***	0,043	0,147	**	*	0,056
7	-	-	-	-	-	-	***	***	***	0,008	***
8	-	-	-	-	-	-	-	***	***	0,015	0,660
9	-	-	-	-	-	-	-	-	*	***	**
10	-	-	-	-	-	-	-	-	-	***	***
11	-	-	-	-	-	-	-	-	-	-	0,001

**Tabelle 4.3.:** Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer vertraut war, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von  $p < 0,0008$  führte. Statistisch signifikante Werte sind mit einem (\*), zwei (\*\*) oder (\*\*\*) Sternen gekennzeichnet. \* $p < 0,0008$ , \*\* $p < 0,00016$ , \*\*\* $p < 0,000016$



**Abbildung 4.5.:** Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer vertraut war.

(Durchschnitt: 4,12, SD: 0,82 und Durchschnitt: 3,93, SD: 1,03). Darstellung 4 unterscheidet sich dabei signifikant von den Darstellungen 1, 3, 5, 6, 8, 9 und 12. Darstellung 7 unterscheidet sich signifikant von den Darstellungen 3, 5, 8, 9 und 12.

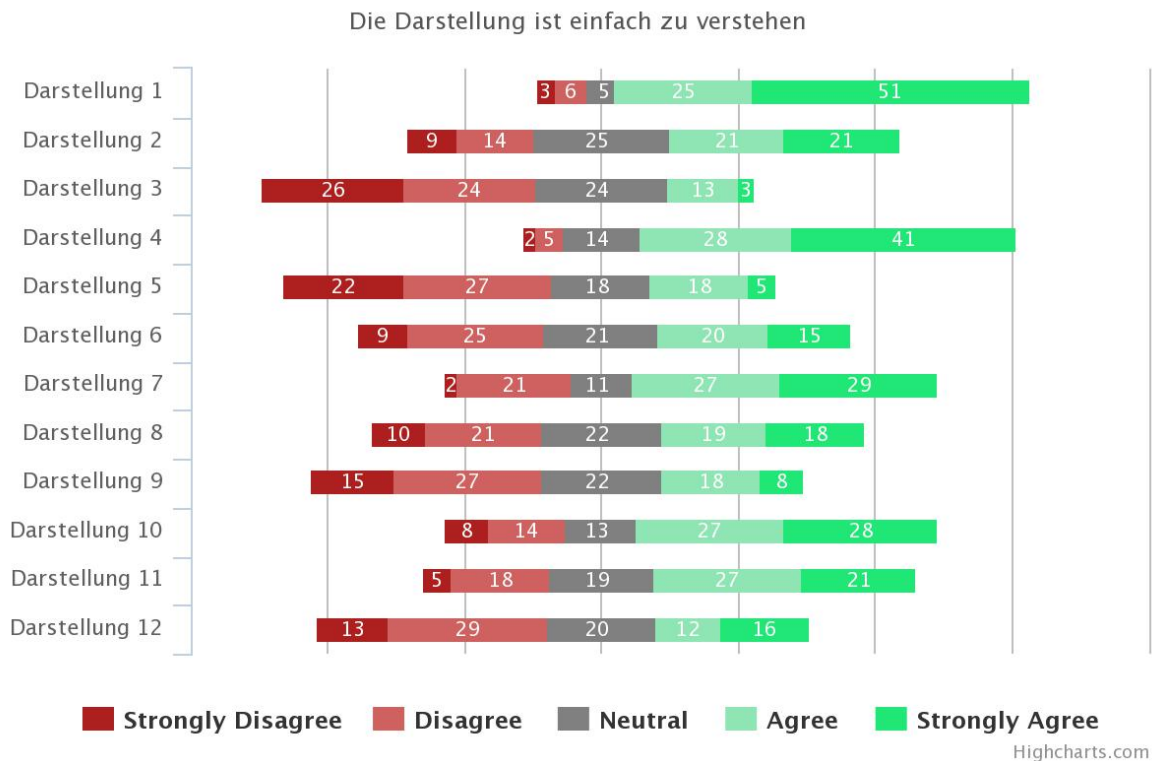
#### 4.4.2. Vertrautheit

Abbildung 4.5 zeigt für jede der 12 Darstellungen die Verteilung der Antworten der 90 Teilnehmerstimmen bezüglich der Frage, ob die jeweilige Darstellung dem Teilnehmer vertraut war.

Der Friedman Test zeigt einen statistisch signifikanten Unterschied für die verschiedenen Darstellungen bezüglich der Vertrautheit,  $\chi^2(11) = 269,915$ ,  $p < 0.000016$ . Die exakten Ergebnisse der Post-hoc Analyse mit Wilcoxon-Vorzeichen-Rang-Test und Bonferroni-Korrektur sind in Tabelle 4.3 zu sehen.

Darstellung 1 schnitt signifikant besser ab als alle anderen Darstellungen mit der Ausnahme von Darstellung 7 und 11 (Durchschnitt: 4,29, SD: 1,06). Darstellung 7 schnitt ebenfalls weit überdurchschnittlich ab (Durchschnitt: 4,21, SD: 1,03) und unterscheidet sich dabei signifikant von den Darstellungen 3, 5, 6, 8, 9, 10 und 12. Darstellung 10 (Durchschnitt: 2,31, SD: 1,19) schnitt bei der Vertrautheit

#### 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten



**Abbildung 4.6.:** Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer einfach zu verstehen war.

signifikant schlechter als die restlichen Darstellungen ab mit der Ausnahme von Darstellung 3 und 5.

#### 4.4.3. Einfaches Verständnis

Abbildung 4.6 zeigt für jede der 12 Darstellungen die Verteilung der Antworten der 90 Teilnehmerstimmen bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer einfach zu verstehen war.

Der Friedman Test zeigt einen statistisch signifikanten Unterschied für die verschiedenen Darstellungen bezüglich des einfachen Verständnis,  $\chi^2(11) = 235,675$ ,  $p < 0.000016$ . Die exakten Ergebnisse der Post-hoc Analyse mit Wilcoxon-Vorzeichen-Rang-Test und Bonferroni-Korrektur sind in Tabelle 4.4 zu sehen.

Darstellung 1 schnitt bei dieser Fragestellung wieder signifikant besser ab als alle anderen Darstellungen mit der Ausnahme von Darstellung 4 und 7 (Durchschnitt: 4,28, SD: 1,06). Ebenfalls weit überdurchschnittlich schnitt Darstellung 4 ab (Durchschnitt: 4,12, SD: 0,82) und unterscheidet sich dabei signifikant von allen anderen Darstellungen außer den Darstellungen 1, 7 und 10. Darstellung 3

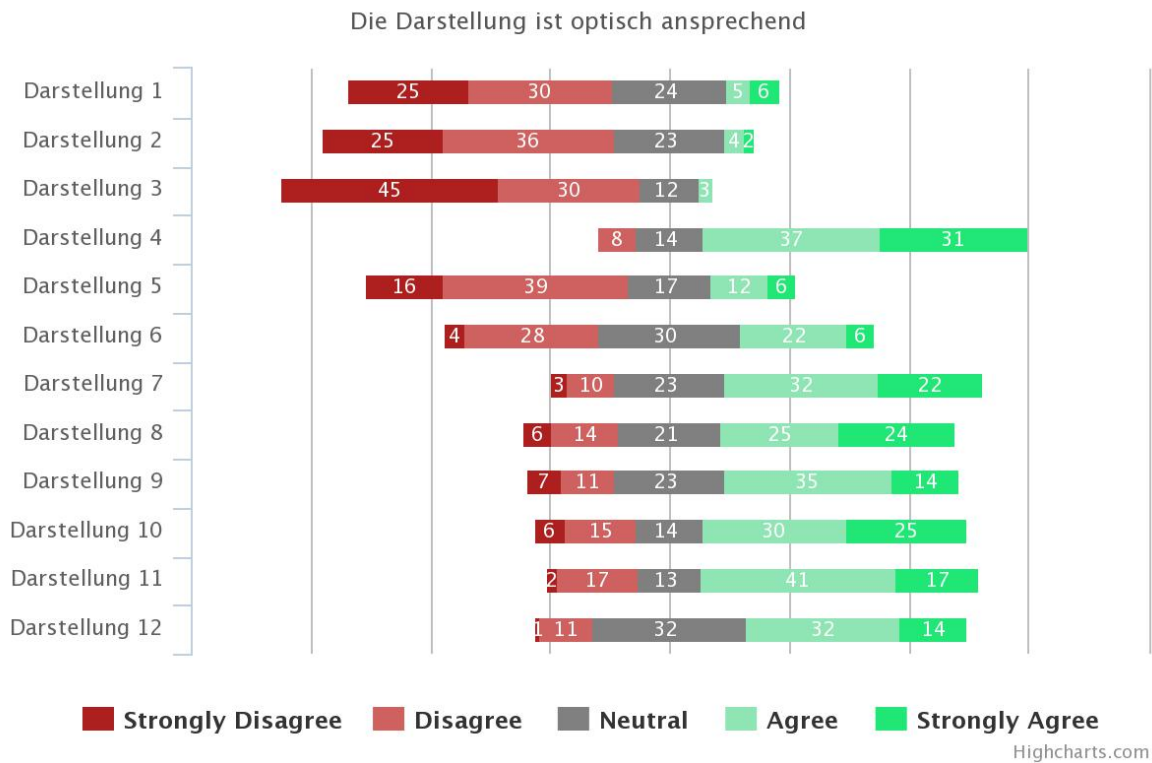
Darstell.	2	3	4	5	6	7	8	9	10	11	12
1	***	***	0,220	***	***	0,001	***	***	**	***	***
2	-	***	***	***	0,093	0,074	0,296	*	0,185	0,417	0,010
3	-	-	***	0,346	**	***	***	0,016	***	***	0,004
4	-	-	-	***	***	0,013	***	***	0,003	*	***
5	-	-	-	-	*	***	0,003	0,166	***	***	0,065
6	-	-	-	-	-	0,005	0,797	0,094	0,008	0,059	0,423
7	-	-	-	-	-	-	*	***	0,612	0,081	***
8	-	-	-	-	-	-	-	0,002	0,008	0,027	0,024
9	-	-	-	-	-	-	-	-	***	***	0,420
10	-	-	-	-	-	-	-	-	-	0,475	**
11	-	-	-	-	-	-	-	-	-	-	***

**Tabelle 4.4.:** Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer einfach zu verstehen war, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von  $p < 0,0008$  führte. Statistisch signifikante Werte sind mit einem (\*), zwei (\*\*) oder (\*\*\*) Sternen gekennzeichnet. \* $p < 0,0008$ , \*\* $p < 0,00016$ , \*\*\* $p < 0,000016$

Darstell.	2	3	4	5	6	7	8	9	10	11	12
1	0,132	***	***	0,218	***	***	***	***	***	***	***
2	-	**	***	0,019	***	***	***	***	***	***	***
3	-	-	***	***	***	***	***	***	***	***	***
4	-	-	-	***	***	0,029	0,011	*	0,008	0,007	0,001
5	-	-	-	-	*	***	***	***	***	***	***
6	-	-	-	-	-	**	0,001	0,003	*	**	*
7	-	-	-	-	-	-	0,279	0,087	0,427	0,708	0,215
8	-	-	-	-	-	-	-	0,398	0,573	0,477	0,777
9	-	-	-	-	-	-	-	-	0,112	0,148	0,470
10	-	-	-	-	-	-	-	-	-	0,840	0,469
11	-	-	-	-	-	-	-	-	-	-	0,232

**Tabelle 4.5.:** Ergebnisse der Post-hoc Analyse mit dem Wilcoxon-Vorzeichen-Rang-Test für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung optisch ansprechend ist, und eine Bonferroni-Korrektur, was zu einem Signifikanzniveau von  $p < 0,0008$  führte. Statistisch signifikante Werte sind mit einem (\*), zwei (\*\*) oder (\*\*\*) Sternen gekennzeichnet. \* $p < 0,0008$ , \*\* $p < 0,00016$ , \*\*\* $p < 0,000016$

#### 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten



**Abbildung 4.7.:** Verteilung der 90 Teilnehmerstimmen für jede Darstellung bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer optisch ansprechend war.

schnitt signifikant schlechter als alle anderen Darstellungen außer den Darstellungen 5, 9 und 12 ab (Durchschnitt: 2,37, SD: 1,15).

#### 4.4.4. Optisch ansprechend

Abbildung 4.7 zeigt für jede der 12 Darstellungen die Verteilung der Antworten der 90 Teilnehmerstimmen bezüglich der Frage, ob die jeweilige Darstellung für den Teilnehmer optisch ansprechend war.

Der Friedman Test zeigt einen statistisch signifikanten Unterschied für die verschiedenen Darstellungen bezüglich der Frage, ob die jeweilige Darstellung optisch ansprechend ist,  $\chi^2(11) = 368,885$ ,  $p < 0.000016$ . Die exakten Ergebnisse der Post-hoc Analyse mit Wilcoxon-Vorzeichen-Rang-Test und Bonferroni-Korrektur sind in Tabelle 4.5 zu sehen.

Bei der Frage, wie optisch ansprechend eine Darstellung sei, schnitt Darstellung 4 am besten ab (Durchschnitt: 4,28, SD: 0,93) und unterscheidet sich damit signifikant von allen anderen Darstellungen. Signifikant schlechter als alle anderen Darstellungen schnitt dahingegen Darstellung 3 ab (Durchschnitt: 1,7, SD: 0,83). Weiterhin schnitten auch die beiden anderen textuellen Darstellungen 1

Die Darstellung...	...ist mir vertraut	...ist einfach zu vers.	...ist opt. ansprechend
...hilft bei Entscheidung	0,51	0,68	0,53
...ist mir vertraut	-	0,61	0,28
...ist einfach zu verstehen	-	-	0,53

**Tabelle 4.6.:** Pearson-Korrelation zwischen den vier verschiedenen Fragestellungen. Alle Werte sind statistisch signifikant,  $p < 0,0005$ .

und 2 sehr schlecht ab (Durchschnitt: 2,3, SD: 1,14 und Durchschnitt: 2,13, SD: 0,95). Darstellung 1 unterscheiden sich dabei signifikant von allen Darstellungen mit Ausnahme von Darstellung 2 und 5. Darstellung 2 unterscheidet sich signifikant von allen Darstellungen außer der Darstellung 5.

#### 4.4.5. Korrelation zwischen den vier Fragestellungen

Es wurde die Pearson-Korrelation zwischen den vier verschiedenen Fragestellungen berechnet,  $p < 0,0005$ . Tabelle 4.6 zeigt die Korrelations-Matrix. Alle Werte sind statistisch signifikant. Die Ergebnisse lassen darauf schließen, dass die Fragenstellungen miteinander gekoppelt sind. Zwischen den Fragen, ob eine Darstellung einfach zu verstehen ist und ob sie bei einer Entscheidung hilft, war die positive Korrelation am höchsten ( $r = 0,68$ ). Ebenfalls recht hoch ( $r = 0,61$ ) war die positive Korrelation zwischen den Fragen, ob eine Darstellung einfach zu verstehen ist und ob sie dem Teilnehmer vertraut ist.

### 4.5. Kommentare der Teilnehmer

Sieben der 90 Teilnehmer fügten einen Kommentar hinzu. Ein Teilnehmer merkte an, dass es in Darstellung 10 hilfreich gewesen wäre, wenn die Vertikalachse eine Skala erhalten hätte. Eine Skalierung der Vertikalachse wurde bewusst nicht eingefügt, da die Wahrscheinlichkeit im Gegensatz zu den anderen Darstellungen nicht über die Höhe, sondern über die Überlappungen bzw. die Intensität der Einfärbung der transparenten Balken kommuniziert werden sollte.

Ein anderer Teilnehmer beschrieb die Studie als interessant, hätte sich aber gewünscht, dass es in den Darstellungen der Wahrscheinlichkeiten eine Übersicht über die Anforderungen der einzelnen Pflanzen geben würde. Das heißt, dass beispielsweise in den Darstellungen für den erwarteten Niederschlag die Mindestmenge an Niederschlag für jede der möglichen Pflanzen zusätzlich dargestellt wird. Diese Idee wurde beim Erstellen der Darstellungen ebenfalls bedacht, allerdings verworfen, da hierbei die Gefahr bestand, dass es einerseits die Entscheidungsfindung zu stark vereinfachen könnte und andererseits die Darstellungen sehr unübersichtlich und nicht mehr unabhängig von der Aufgabe wären.

In einem weiteren Kommentar eines Teilnehmers wurde darauf hingewiesen, dass manche Darstellungen besser verständlich gewesen wären, wenn man sie gespiegelt oder um  $90^\circ$  rotiert hätte. Dieser Teilnehmer nannte allerdings keine konkrete Darstellung.

Ein anderer Teilnehmer teilte mit, dass er teilweise Probleme damit hatte, die Darstellungen zu analysieren, da er farbenblind sei. Auf diesen Umstand wurde bei der Erstellung der Darstellungen tatsächlich keine Rücksicht genommen.

### 4.6. Diskussion

In allen vier Kategorien wurde Darstellung 4 von den Teilnehmern (Liniendiagramm des Erwartungswerts mit möglichem Wertebereich als Flächendiagramm) mit über 3,6 bewertet. Bei der Frage, ob der Teilnehmer der Meinung ist, dass die Darstellung ihm bei seiner Entscheidung helfen könnte, erreichte sie mit durchschnittlich 4,12 die höchste Bewertung, obwohl sie nur sehr undifferenziert die Unsicherheit kommuniziert, da die Wahrscheinlichkeitsverteilung nicht genauer eingesehen werden kann. Ähnlich verhält es sich auch bei Darstellung 1 (textueller Erwartungswert), die keine Unsicherheit kommuniziert und Darstellung 2 (textueller Erwartungswert mit Standardabweichung), die die Unsicherheit nur sehr undifferenziert kommuniziert. Auch diese beiden Darstellungen schnitten trotz dieser Tatsache überdurchschnittlich gut in dieser Kategorie ab (Durchschnitt: 3,63, SD 1,1 und Durchschnitt: 3,68, SD: 1,1). Aus diesem Resultat heraus, könnte man annehmen, dass Menschen deterministische Prognosen den unsicheren Prognosen vorziehen.

Wie erwartet waren die gängigen, aus dem Schulunterricht und den Medien bekannten Darstellungen, wie der einfache textuelle Erwartungswert (Darstellung 1), Erwartungswert mit Standardabweichung (Darstellung 2), Liniendiagramm (Darstellung 4), Balkendiagramm (Darstellung 7) und Funktionsgraph (Darstellung 11) im Gegensatz zu Quantilen und Box-Plots (Darstellung 3 und 5) den meisten Teilnehmern vertraut.

Es wurde die Pearson Korrelation berechnet. In Tabelle 4.6 sind die Ergebnisse zu sehen. Hierbei fiel auf, dass die Teilnehmer diejenigen Darstellungen am besten verstanden, die ihnen vertraut waren. Außerdem besteht ein Zusammenhang zwischen dem einfachen Verständnis einer Darstellung und wie sehr die Teilnehmer glaubten, dass diese Darstellung ihnen bei der Entscheidung helfen könnte. Die einzige Ausnahme ist Darstellung 10 (Wahrscheinlichkeitsdichtediagramm, bei der die Unsicherheit über transparente Überlagerungen kommuniziert wurde), die sehr vielen Teilnehmern nicht vertraut war. Trotzdem waren viele Teilnehmer der Meinung, dass sie einfach und intuitiv verständlich sei.

Die ersten drei Darstellungen waren allesamt textuell und schnitten wie erwartet im Gegensatz zu den graphischen Darstellungen optisch sehr schlecht ab. Sie belegten im Durchschnitt die letzten 3 Plätze. Darstellung 1 erreichte einen Durchschnitt von 2,3 (SD: 1,14). Darstellung 2 erreichte einen Durchschnitt von 2,13 (SD: 0,95) und Darstellung 3 erreichte einen Durchschnitt von 1,7 (SD: 0,83). Diese Ergebnisse könnten zur Annahme führen, dass Darstellungen die ausschließlich aus Text bestehen nicht ausreichen, um optisch ansprechend zu sein.

Die kumulativen Darstellungen 8,9 und 12 haben es, auf Grund der schlechten durchschnittlichen Punktzahlen, nicht ins Spiel geschafft, obwohl sie sich für die gegebene Problemstellung im Spiel sehr gut geeignet hätten. Im Spiel bzw. der Umfrage musste man sich die Frage stellen, welches relativ sichere Ereignis eintreten wird. Beispielsweise hätte man für sich als relativ sicher, die 90% Marke verwenden können. Dann war die Frage zu beantworten, wieviel Niederschlag mindestens mit



Darstellung	Hilft bei Entscheidung	Vertraut	Einfach zu verstehen	Optisch ansprechend	Gesamt
1	3,63	4,29	4,28	2,3	3,63
2	3,68	3,9	3,34	2,13	3,26
3	2,86	2,79	2,37	1,7	2,43
4	4,12	3,74	4,12	4,01	4,00
5	3,22	2,73	2,52	2,48	2,74
6	3,48	3,11	3,08	2,98	3,16
7	3,93	4,21	3,67	3,67	3,87
8	3,47	3,49	3,16	3,52	3,41
9	3,16	2,84	2,74	3,42	3,04
10	3,73	2,31	3,59	3,59	3,31
11	3,89	3,88	3,46	3,6	3,71
12	3,44	3,5	2,88	3,52	3,34

**Abbildung 4.8.:** Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der 12 Darstellungen, die den Teilnehmern vorgestellt wurden. Bewertungen kleiner als 2.4 wurden in rot, zwischen 2.4 und 3.0 in orange, zwischen 3.0 und 3.6 in hellgrün und größer als 3.6 in dunkelgrün dargestellt.

90% Wahrscheinlichkeit fallen wird. Diese Frage hätte mit Hilfe der kumulativen Darstellungen auf einen Blick ohne vorheriges Aufaddieren der einzelnen Wahrscheinlichkeiten, wie bei den anderen Darstellungen, beantwortet werden können.

## 4.7. Auswahl für das Experiment

Darstellung 1, Darstellung 4, Darstellung 7 und Darstellung 11 erreichten bei den Gesamtwertungen, die in Abbildung 4.8 zu sehen sind, die vier durchschnittlich höchsten Bewertungen. Weiterhin schnitten diese vier Darstellungen auch in den Einzelkategorien überdurchschnittlich gut ab ( $> 3,6$ ) und haben keine unterdurchschnittlichen Bewertungen ( $< 3,0$ ) erhalten. Einzige Ausnahme stellte Darstellung 1 dar. Sie erzielte in Kategorie „optisch ansprechend“ nur eine Bewertung von 2,3 (SD: 1,14). Andererseits erzielte sie in den den Kategorien „vertraut“ (4,29, SD: 1,06) und „einfach zu verstehen“ (4,28, SD: 1,06) die höchsten Bewertungen. Außerdem ist sie die einzige dieser vier Darstellungen, die rein textuell ist und steht somit im Kontrast zu den restlichen drei graphischen Darstellungen.

Weiterhin ist so mindestens jeweils eine Darstellung aus jeder der drei Klassen der Klassifikation in Tabelle 4.1, welche die Darstellungen nach dem Grad der Darstellung der Unsicherheit klassifizierte, vorhanden. Die vier Darstellungen 1, 4, 7 und 11 sind bezüglich ihres Detailgrads der Darstellung der Unsicherheit sehr unterschiedlich. Darstellung 1 stellte nur den Erwartungswert da, gab aber keinerlei Informationen zur Unsicherheit an. Darstellung 4 zeigte nur den Erwartungswert und die Spannweite der Werte, machte aber keinerlei Angaben zur Verteilung der Werte. Darstellung 7 stellte die Wahrscheinlichkeitsverteilung mit Hilfe eines Histogramms, in größeren Klassen eingeteilt, dar. Darstellung 11 stellte die Wahrscheinlichkeitsverteilung kontinuierlich in Form eines Funktionsgraphen

#### 4. Online-Umfrage über Visualisierungsalternativen für unsichere Daten

---

dar und fasste die Informationen zur Unsicherheit somit am wenigsten zusammen. Dieser Umstand ließ umfassendere Untersuchungen bezüglich der Frage zu, ob detaillierte Angaben zur Unsicherheit auch zu besseren Entscheidungen bei den Nutzern führen und ob Laien diese Informationen in realen Entscheidungen anwenden können.

Aus den genannten Gründen wurden Darstellung 1 (textueller Erwartungswert), Darstellung 4 (Liniendiagramm des Erwartungswerts mit Spannweite als Flächendiagramm), Darstellung 7 (Histogramm in Form eines Balkendiagramms) und Darstellung 11 (kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen) für das Spiel ausgewählt.

## 5. Anwendung Farm-Smart

Um zu testen, ob sich die vier, mit Hilfe der Online-Umfrage aus Kapitel 4 ausgewählten, Darstellungen tatsächlich für die Kommunikation von Unsicherheiten unter realen Bedingungen eignen, wurde ein Spiel implementiert. Dieses beruht auf Wahrscheinlichkeiten und den darauf basierenden Entscheidungen. Die für das Spiel benötigten Wahrscheinlichkeitsangaben wurden den Teilnehmern mit Hilfe von vier unterschiedlichen Darstellungen präsentiert. Um für die Teilnehmer einen Bezug zur Realität zu schaffen, wurde im Spiel die Führung eines Ackerbau-Betriebs thematisiert. Unter Zuhilfenahme eines wahrscheinlichkeitsbasierten Wetterberichts sollten die Teilnehmer möglichst profitable Anbauentscheidungen treffen. Für die graphische Darstellung der Wetterberichte wurden die vier identifizierten Darstellungen genutzt.

### 5.1. Spielablauf und Konzept

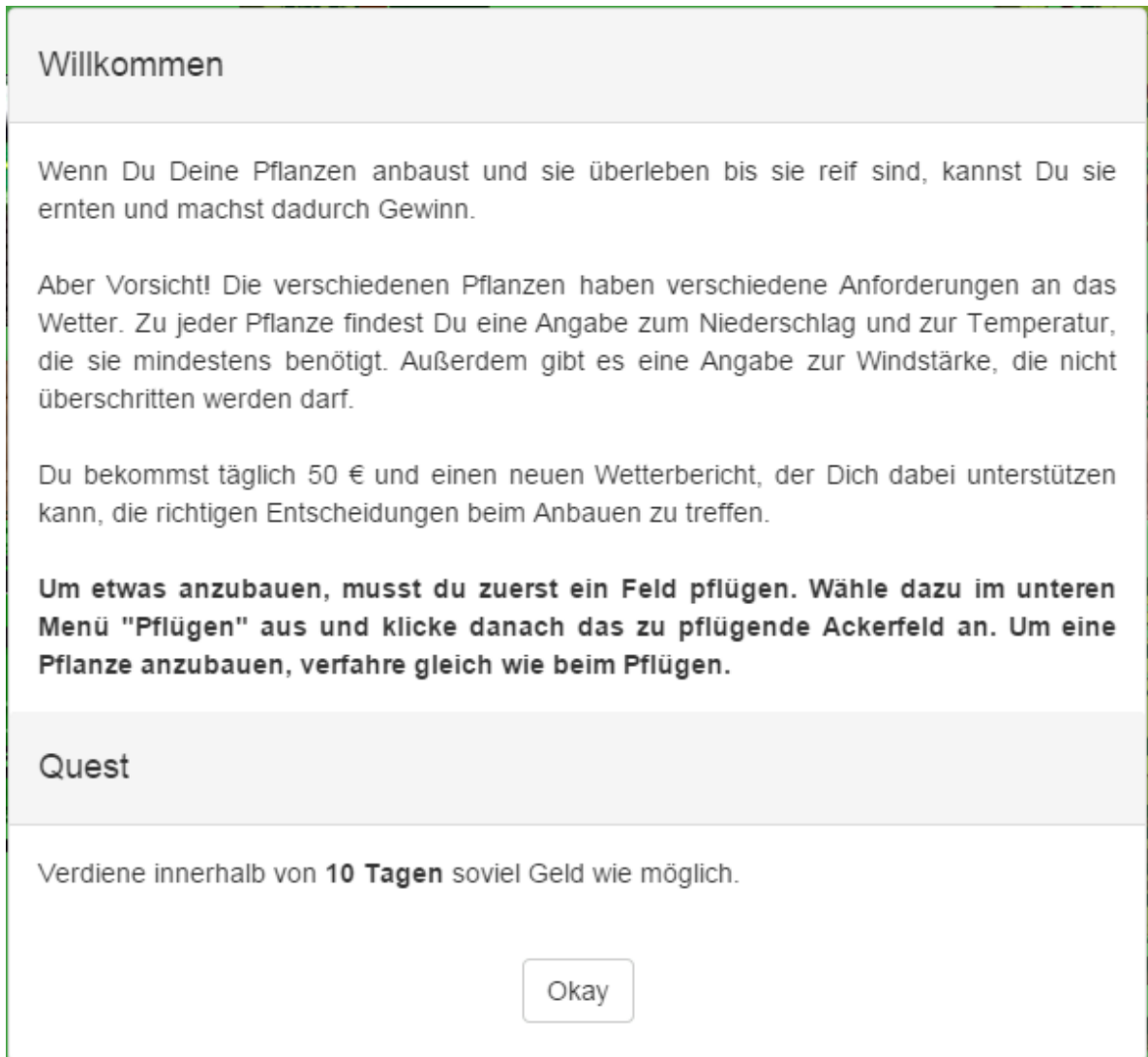
Der Spieler wird mit dem Text in Abbildung 5.1 begrüßt. Zu Beginn des Spieles erhalten alle Teilnehmer ein Startkapital von 200 € und bekommen den aktuellen Wetterbericht für die folgenden drei Tage vor dem ersten Anpflanzen gezeigt.

Der Spieler kann nun die ersten Felder pflügen und basierend auf dem Wetterbericht, verschiedene Pflanzen anbauen. Über einen Button steuert der Spieler selbst den Zeitpunkt des Tagwechsels und bekommt anschließend das tatsächlich eingetretene Wetter, sowie den Zustand seiner angebauten Pflanzen angezeigt. Außerdem werden unabhängig vom eingetretenen Wetter 50 € auf seinem Konto gutgeschrieben. Im weiteren Verlauf kann der Spieler seine erntereifen Pflanzen abbauen, bekommt den dazugehörigen Gewinn auf seinem Konto gutgeschrieben und kann unter Zuhilfenahme des neuen Wetterberichtes entscheiden ob und welche neuen Pflanzen er anbauen möchte.






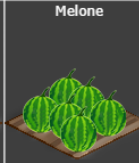


Wenn eine Pflanze stirbt bevor sie erntereif ist, bringt sie keinen Gewinn ein und kann stattdessen für zehn Euro beseitigt werden, um Platz für neue Pflanzen zu schaffen.

Das Spiel endet nach zehn Tagen. Ziel ist es, den Gewinn durch gute Entscheidungen beim Anbauen der Pflanzen in diesen zehn Tagen zu maximieren. Das Ranking der Spieler richtete sich nach dem Geldstand am Ende des Spiels.

Es gab insgesamt 8 verschiedene Pflanzen, die auf 36 potenziellen Ackerflächen angebaut werden konnten. Die Pflanzen hatten unterschiedliche Anbaukosten, welche sich je nach Art zwischen 20 Euro und 40 Euro bewegten. Die verschiedenen Werte beim Ernten der reifen Pflanzen lagen im Bereich von 40 Euro bis 160 Euro. Außerdem waren die Anbauzeiten ebenfalls von Pflanze zu Pflanze unterschiedlich, lagen jedoch immer zwischen 1 und 3 Tagen, damit der Anbauzeitraum mit der 3-Tages-Vorschau des Wetterberichts übereinstimmt und nicht über den prognostizierten Zeitraum



**Abbildung 5.1.:** Der Begrüßungstext im Spiel. Eine kleine Einweisung wird gegeben und die Aufgabe für den Spieler kurz beschrieben.

							
Kosten: 20 € Wert: 40 € Anbauzeit: 1 Tag	Kosten: 20 € Wert: 50 € Anbauzeit: 1 Tag	Kosten: 20 € Wert: 60 € Anbauzeit: 1 Tag	Kosten: 30 € Wert: 70 € Anbauzeit: 2 Tage	Kosten: 30 € Wert: 90 € Anbauzeit: 2 Tage	Kosten: 30 € Wert: 100 € Anbauzeit: 2 Tage	Kosten: 40 € Wert: 130 € Anbauzeit: 3 Tage	Kosten: 40 € Wert: 160 € Anbauzeit: 3 Tage
Niederschlag: > 1 mm Temperatur: > 0 °C Windst.: < 20 km/h	Niederschlag: > 6 mm Temperatur: > 9 °C Windst.: < 16 km/h	Niederschlag: > 5 mm Temperatur: > 11 °C Windst.: < 11 km/h	Niederschlag: > 2 mm Temperatur: > 3 °C Windst.: < 16 km/h	Niederschlag: > 3 mm Temperatur: > 7 °C Windst.: < 13 km/h	Niederschlag: > 7 mm Temperatur: > 10 °C Windst.: < 14 km/h	Niederschlag: > 3 mm Temperatur: > 4 °C Windst.: < 17 km/h	Niederschlag: > 4 mm Temperatur: > 4 °C Windst.: < 14 km/h

**Abbildung 5.2.:** Die Auswahl der Pflanzen im Menü des Spiels. Unter der jeweiligen Abbildung der Pflanze sind die Angaben zu den Kosten, Wert, Anbauzeit, benötigte Mindesttemperatur, Mindestniederschlagsmenge und maximale Windstärke abgebildet.

hinaus angebaut werden muss. Die genauen Werte der einzelnen Pflanzen sind in Abbildung 5.2 zu sehen.

Jede Pflanze hatte unterschiedliche Anforderungen an und Resistenzen gegen das Wetter. Alle pflanzenspezifischen Bedingungen mussten erfüllt sein, damit eine Pflanze einen Tag überlebt. Diese Bedingungen waren an die drei Faktoren Niederschlag, Temperatur und Windstärke gebunden. Um zu überleben musste die Pflanze eine Mindestmenge an Niederschlag bekommen, mindestens einer gewissen Temperatur und höchstens einer bestimmten Windstärke ausgesetzt sein.

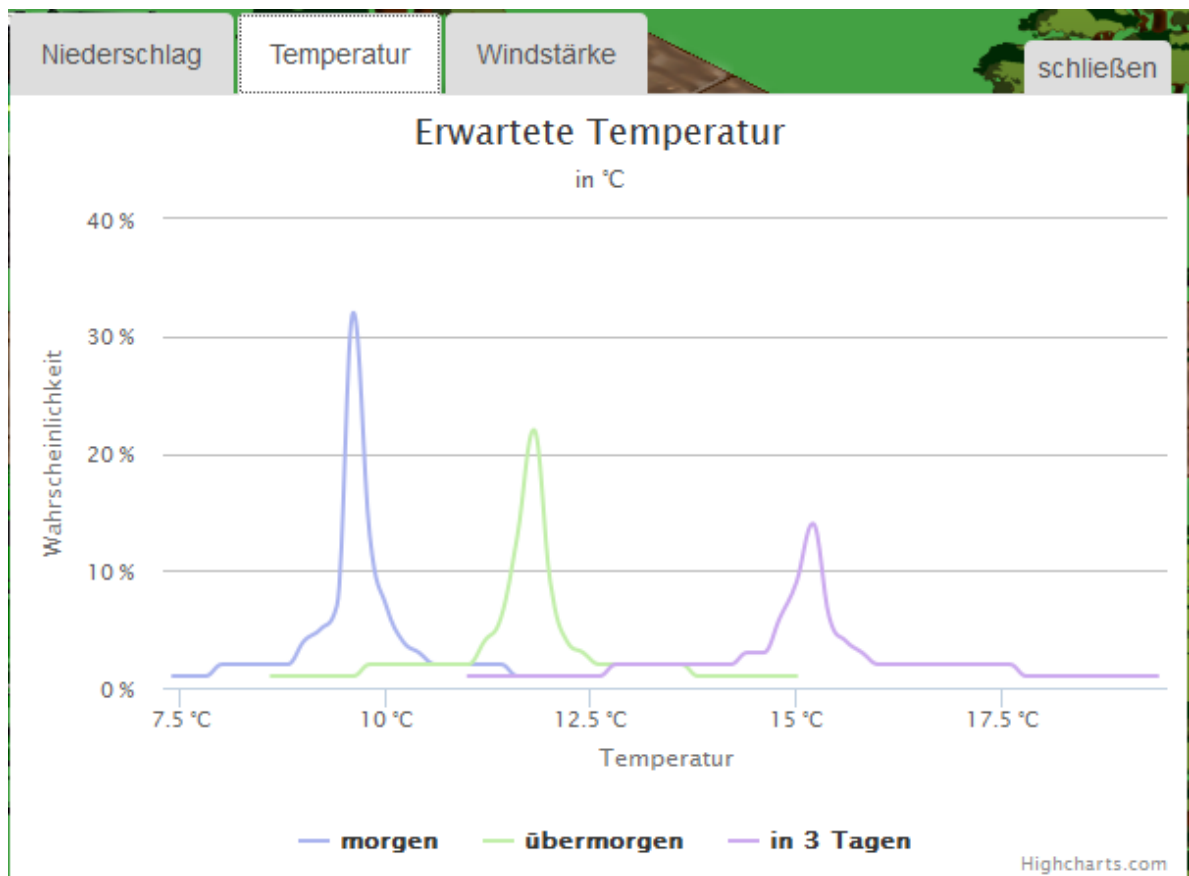
Um ein ausgeglichenes Spiel zu gewährleisten, wurden die Eigenschaften der Pflanzen im Vergleich zu ihren Kosten und dem erwarteten Gewinn gewählt. Deshalb gab es Pflanzen mit sehr geringen Anforderungen an das Wetter, also Pflanzen die sehr wetterunabhängig waren, wenig kosteten und eine kurze Anbauzeit hatten, dafür aber auch nur einen geringen Gewinn beim Ernten erzielten. Pflanzen, die hingegen sehr wetterempfindlich waren, relativ hohe Anbaukosten und eine lange Anbauzeit hatten, realisierten dementsprechend einen sehr viel höheren Wert bei der Ernte. Das Ziel dieser Differenzierung war es, dem Spieler einen Anreiz zu geben, taktischer zu spielen und durch strategisch kluge Spielzüge eine größere Diskrepanz beim finalen Geldstand zu erreichen. Dies sollte einen Aufschluss darüber geben, welche der zur Verfügung gestellten Darstellungen zu guten taktischen Entscheidungen und so zu erfolgreichen Ergebnissen führen.

Die für den Wetterbericht des Spiels ausgewählten Darstellungen 1, 4 und 7 aus Kapitel 4.2 wurden unverändert, wie in der Umfrage dargestellt (vgl. Abbildung 4.1 und Abbildung 4.2) in das Spiel implementiert. Darstellung 11 wurde leicht angepasst, indem die Anzahl der Klassen der kontinuierlichen Darstellung des Histogramms erhöht wurde, um den Detailgrad der Darstellung der Unsicherheit zu erhöhen und sich so stärker von Darstellung 7 zu unterscheiden. Die geänderte Darstellung 11 ist in Abbildung 5.3) zu sehen.

Der Wetterbericht wurde jeden Tag aufs Neue zufällig erstellt und enthielt eine Vorschau der nächsten 3 Tage. Er gab den erwarteten Niederschlag, die prognostizierte Temperatur und Windstärke an.

Die Anwendung verwaltete für den Wetterbericht intern jeweils drei Wahrscheinlichkeitsverteilungen für jedes der Wetterelemente. Die erste Wahrscheinlichkeitsverteilung enthielt die Daten für den folgenden Tag, die zweite für den zweiten Tag und die dritte Verteilung enthielt Daten für den dritten Tag.

## 5. Anwendung Farm-Smart



**Abbildung 5.3.:** Die geänderte Darstellung 11 des Wetterberichts im Spiel mit 3-Tages-Vorschau für Temperatur, Niederschlag und Windstärke, den der Spieler täglich einsehen konnte.

Der initiale Wetterbericht, der dem Teilnehmer in der ersten Runde, also bei Spielbeginn, für den folgenden Tag angezeigt wurde, wurde völlig zufällig erzeugt. Für die erzeugten Daten wurden lediglich Ober- beziehungsweise Untergrenzen festgelegt, die sicherstellten, dass sich die Ergebnisse in realistischen Bereichen bewegten. Die Grenzen legten fest, dass der Niederschlag zwischen 0 mm und 10 mm liegen musste, die Temperatur mindestens 0 °C und höchstens 20 °C betragen durfte. Die mögliche Windstärke war auf den Bereich zwischen 0 km/h und 20 km/h begrenzt.

Die Vorhersage für den zweiten Tag wurde aus der initialen Vorhersage für den ersten Tag abgeleitet, indem der Wertebereich der Verteilung für den ersten Tag zufällig, um höchstens 25% Prozent nach oben oder unten verschoben wurde. Außerdem wurde die Höhe und Breite der Wahrscheinlichkeitsverteilung variiert. Die Spannweite der Werte wurde breiter, das heißt, die Verteilung wurde flacher und unpräziser. Die Vorhersage für den dritten Tag wurde mit Hilfe des selben Prinzips erzeugt. Jedoch wurde die Vorhersage des dritten Tages aus der Vorhersage für den zweiten Tag erstellt.

Mit allen weiteren Tagen beziehungsweise Runden wurde folgendermaßen verfahren. Die Vorhersage für den ersten Tag wurde aus der Vorhersage für den zweiten Tag abgeleitet, indem die Spannweite

der Werte der Wahrscheinlichkeitsverteilung des zweiten Tages um eine zufällige, aber beschränkte Menge verringert wurde. Die Vorhersage für den zweiten Tag wurde auf die gleiche Weise aus der Vorhersage für den dritten Tag abgeleitet. Die Vorhersage für den dritten Tag einer neuen Runde wurden hingegen nicht aus dem Nachfolgetag, sondern aus dem Vorgängertag, also dem dritten Tag der vorhergegangenen Runde, durch eine geringe zufällige Verschiebung des Wertebereichs und eine Erhöhung der Wertespannweite der Wahrscheinlichkeitsverteilung berechnet.

Durch dieses Vorgehen wurde erreicht, dass die Vorhersage immer unpräziser und undifferenzierter wurde, je weiter die Prognose in die Zukunft reichte. Durch die zufälligen aber geringen Veränderungen von maximal 25% und Verschiebungen auf Basis des jeweiligen Vortags, unterlag das Wetter keinen zu großen unvorhersehbaren Schwankungen. Abbildung 5.3 zeigt beispielhaft einen Wetterbericht im Spiel.

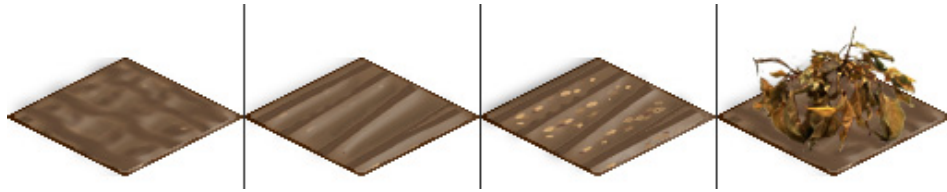
Um das Spiel möglichst intuitiv und leicht verständlich zu gestalten, wurde dem direkten Feedback durch das Spiel an den Spieler eine große Bedeutung beigemessen. Der Spieler sollte sich schnell einen Überblick über den Spielstand, den Zustand seiner Pflanzen und seiner Möglichkeiten verschaffen können. Verwirklicht wurde dies beispielsweise durch einen veränderlichen Mauszeiger. Wenn der Spieler seinen Mauszeiger über das Spielfeld bewegt, verändert sich der Cursor bei aktiver „Pflügen“-Auswahl im Menü in eine Schaufel. Ist eine Pflanze die momentan aktive Menüauswahl, so ändert sich der Cursor über einem gepflügten Ackerfeld in Samen und zeigt an, dass dort eine Pflanze angebaut werden kann. Ist eine Pflanze reif und kann geerntet werden, verändert sich der Mauszeiger in eine Hand, wenn er über der reifen Pflanze platziert wird.

Weiterhin erhielt der Spieler direktes Feedback, wenn er eine Pflanze anbauen wollte, für die er nicht mehr ausreichend Geld zur Verfügung hatte, indem die Geldanzeige beim missglückten Anbauversuch für zwei Sekunden rot aufleuchtete. Analog dazu leuchtete die Tages-Anzeige rot auf, wenn der Spieler versuchte eine Pflanze anzubauen, die erst zum Ernten bereit gewesen wäre, nachdem das Spiel bereits beendet war.

Außerdem erhielt der Spieler durch den Button für den Wetterbericht visuelles Feedback. Dieser blinkte rot, sobald ein neuer, vom Spieler noch nicht eingesehener Wetterbericht vorlag. Der Zustand seiner bebauten Ackerfelder und Pflanzen war für den Spieler zu jedem Zeitpunkt direkt ersichtlich. Es gab fünf verschiedene Zustände die ein Ackerfeld annehmen konnte, diese waren „ungepflügt“, „gepflügt“, „gesät“, „reif“ oder „gestorben“. Der Zustand „reif“ wurde mit der Grafik der jeweilig reifen Pflanze angezeigt, auf dem Ackerfeld war also beispielsweise ein reifer Kürbis zu sehen. Die Grafiken, die den vier anderen Zuständen zugeordnet waren sind in Abbildung 5.4 zu sehen. Bei der Wahl der Grafiken wurde darauf geachtet, dass diese den beschriebenen Zustand möglichst exakt und einfach verständlich abbildeten.

Genaue Informationen zu jedem einzelnen Ackerfeld erhielt der Spieler, wenn er den Mauszeiger über ein bepflanztes Feld bewegte. Per Mouseover wurden ihm je nach Zustand des Ackerfeldes Informationen zur Pflanzenart, der Anbaudauer, dem Wert beim Ernten der reifen Pflanze und den Kosten zum Entfernen bei einer gestorbenen Pflanze angezeigt.

Die zeitliche Einteilung des Spiels basierte auf Runden, während ein neuer Tag oder ein neuer Wetterbericht jeweils eine neue Runde darstellten. Das Spiel wurde bewusst nicht in Echtzeit durchgeführt, da ansonsten die Zeit, die für Bauentscheidungen, zur Verfügung gestanden hätte zu begrenzt wäre



**Abbildung 5.4.:** Die vier der fünf Zustände, die eine Ackerfläche annehmen konnte: „ungepflügt“, „gepflügt“, „gesät“ und „gestorben“ (von links nach rechts). Der fünfte Zustand „reif“ wurde mit der Grafik der jeweilig reifen Pflanze angezeigt und ist in Abbildung 5.2 zu sehen.

und eine kurze Reaktionszeit, Glück und schnelles Klicken zu wichtig gewesen wären. Die Intention war jedoch, dass gute, überlegte Entscheidungen zu einem besserem Ranking führen und nicht unüberlegte unter Zeitdruck getroffene Entscheidungen.

Am Ende erhielt der Spieler die Möglichkeit die Darstellung, die er in seiner gerade gespielten Partie verwendet hat, zu bewerten. Die Bewertungskriterien waren dieselben, wie in der vorangegangenen Umfrage. Es bestand außerdem die Möglichkeit einen persönlichen Kommentar zum Spiel abzugeben. Anschließend wurde dem Spieler die Bestenliste angezeigt, in der er die Punktzahlen aller bisherigen Spiele sämtlicher Teilnehmer sehen konnte. Dies sollte einen Anreiz schaffen, das Spiel nochmals zu spielen um das eigene Ergebnis zu verbessern.

Die Zuteilung der Darstellungen an die Spieler war nicht zufällig. Die Anwendung hat vor Spielbeginn den Spieler anhand der Facebook-ID identifiziert und ermittelt, mit welchen Darstellungen der Spieler bereits gespielt hatte. Sollte ein Spieler zum ersten mal spielen, hat die Anwendung dem Spieler die am wenigsten gespielte Darstellung zugeteilt. Bei einem wiederholten Spiel wurde jedoch sichergestellt, dass der Spieler eine neue, noch nicht verwendete Darstellung erhielt, so dass er einen neuen Anreiz hatte, das Spiel mehrmals zu spielen. Hat der Spieler mit jeder Darstellung bereits einmal gespielt, stellte die Anwendung trotzdem sicher, dass der Spieler in jedem weiteren Spiel eine andere Darstellung erhielt. Insgesamt wurden die Darstellungen also so verteilt, dass auf alle Spieler gesehen, jede Darstellung gleich oft verwendet wurde.

### 5.2. Implementierung

Vor der Implementierung des Spiels wurde ein Mockup erstellt, um den Spielablauf und die Benutzeroberfläche besser ermitteln und besprechen zu können. Es handelte sich um ein reines Grundgerüst der Bedienelemente ohne weitere Funktionalität. Die Erstellung erfolgte mit einem Bildbearbeitungsprogramm. Es fand keine Programmierung statt. Das Mockup befindet sich im Anhang A.

Das Spiel wurde als Webanwendung implementiert. Im Gegensatz zur Implementierung als mobile Anwendung wurde so gewährleistet, dass ausreichend Platz für das Spielfeld und die Darstellungen auf dem Bildschirm zur Verfügung stand. Um ein ansprechendes und übersichtliches Spielfeld zu erzeugen, wurde ein isometrisches Tile-basiertes 2D Spielfeld gewählt. Der Vorteil ist, dass das Spielfeld hierbei optisch wirkt, als wäre es dreidimensional, was realistischer, ansprechender und





**Abbildung 5.5.:** Das isometrisches Tile-basierte 2D Spielfeld mit den 36 Ackerflächen auf denen jeweils eine Pflanze angebaut werden kann. Damit eine Pflanze angebaut werden konnte, musste der Spieler zuerst das Feld für 10 Euro pflügen. Dazu musste er im unteren Menü das Feld „Pflügen“ auswählen und danach das zu pflügende Ackerfeld klicken. Daraufhin konnte er die Pflanze seiner Wahl anbauen, indem er wieder im Menü zuerst die Pflanze auswählte und danach auf ein bereits gepflühtes Ackerfeld klickte.

zeitgemäß ist. Im Gegensatz zu einer tatsächlich dreidimensionalen Darstellung ist es aber sehr viel einfacher zu erstellen und zu handhaben. Durch den Tile-basierten Aufbau war das Spielfeld beliebig in der Größe zu erweitern und anzupassen. In Abbildung 5.5 ist das isometrisches Tile-basiertes 2D Spielfeld zu sehen.

Um das Zeichnen des Spielfelds auf das HTML5 Canvas Element zu vereinfachen wurde CreateJS<sup>1</sup> verwendet. CreateJS ist eine Sammlung von Javascript Bibliotheken, die das Erstellen von interaktiven Inhalten mittels HTML5 erleichtern. Insbesondere wurde dadurch das Preloading der Spieleinhalte, die Erstellung der Animationen und das Erkennen von Mouseover- und Mouseclick-Aktionen des Spielers erleichtert.

<sup>1</sup><http://www.createjs.com/>

## 5. Anwendung Farm-Smart

---

Um die Erzeugung der Diagramme und Graphen aus den Simulations- und Spieldaten für die unterschiedlichen Darstellungen der Wetterberichte zu vereinfachen, wurde Highchart.js<sup>2</sup> verwendet. Highchart.js ist eine Bibliothek für Diagramme, die in Javascript geschrieben wurde. Sie ermöglicht ein vereinfachtes Hinzufügen von interaktiven Graphen und Diagrammen zu einer Webseite oder Webanwendung. Highchart.js wurde auch gewählt, weil sie eine Vielzahl an verschiedenen Diagrammarten unterstützt und bei weiteren Anforderungen beliebig angepasst werden kann. Highchart.js unterstützte nicht das direkte Zeichnen auf dem Spielfeld-Canvas. Highchart.js erstellte ein eigenes DOM-Element im Browser, welches dann mit Hilfe von CreateJS als Canvas-Element auf das Spielfeld gezeichnet und animiert werden musste.

Die Daten jeder einzelnen Runde aller gespielten Partien wurden erfasst und an eine MySQL Datenbank per AJAX/PHP asynchron übertragen. Unter anderem wurde hierbei die Facebook ID, der Facebook Name, das Geschlecht, das ungefähre Alter, die Uhrzeit, die Darstellung des Wetterberichts, die Runde, der momentane Geldstand, die Anzahl der Aufrufe des Wetterberichts, die Dauer der Ansicht des Wetterberichts, die Anbauentscheidungen und das tatsächlich eingetretene Wetter gespeichert.

Weiterhin wurden die in Kapitel 5.1 näher besprochenen Wahrscheinlichkeitsverteilungen der Wetterberichte jede Runde gemeinsam mit den anderen Informationen gespeichert. Pro Runde wurden neun verschiedene Wahrscheinlichkeitsverteilungen gespeichert. Für jeden der folgenden drei Tag wurde je eine Wahrscheinlichkeitsverteilung für den Niederschlag, die Temperatur und die Windstärke archiviert. Das Festhalten dieser Verteilungen war für die spätere Auswertung des Spiels von besonderer Bedeutung, denn dadurch konnte nachvollzogen werden, welche Darstellung und Verteilung einer bestimmten Anbauentscheidung eines Spielers zu Grunde lag.

<sup>2</sup><http://www.highcharts.com/>

## 6. Öffentliches Experiment

Das Spiel wurde bei diesem Experiment öffentlich durchgeführt, so sollten möglichst viele Spieler erreicht werden. Dadurch, dass das Spiel nicht moderiert wurde, konnte ein unbegrenzte Anzahl an Spielern gleichzeitig und zu jeder beliebigen Uhrzeit spielen.

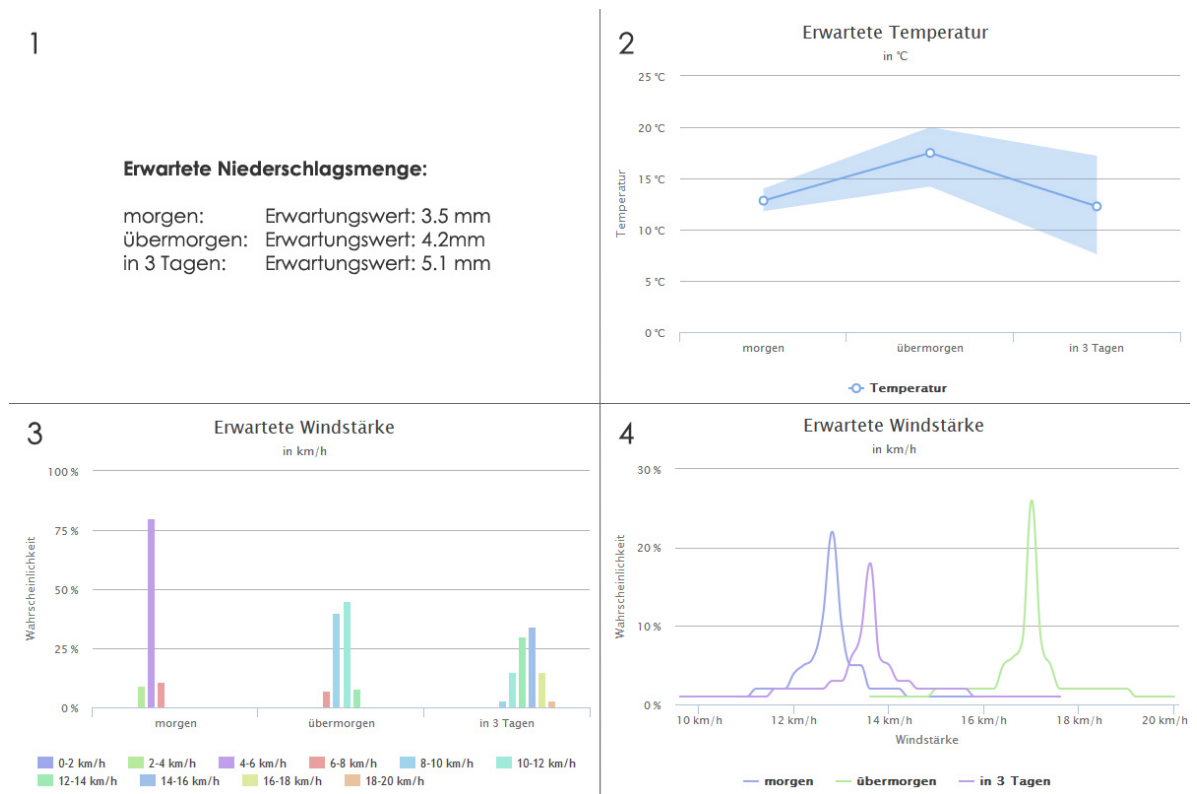
### 6.1. Auswertung der Daten

Ein Ziel des Experiments war es, herauszufinden, inwiefern sich die vier, in Kapitel 4.2 gewählten, Darstellungen für Unsicherheiten (siehe Abbildung 6.1) im Hinblick auf die Problemstellung, nämlich wie gut die Teilnehmer die Darstellungen verstehen und bei ihrer Entscheidungsfindung unter realen Bedingungen einsetzen können, unterscheiden. Es sollte also eine Aussage über die Qualität der Entscheidung abhängig von der Darstellung gemacht werden. Um diese Abhängigkeit feststellen zu können, muss ein messbarer Wert aus dem Spiel ermittelt werden, welcher im direkten Zusammenhang mit der Qualität der Entscheidungen des Spielers steht. Der Highscore, also der schlussendliche Geldstand des Spielers, sollte eigentlich sehr viel über die Qualität seiner Entscheidungen aussagen, schließlich ist er das direkte Resultat seiner Handlungen und der Spieler war immer darauf Bedacht diesen zu maximieren.

Allerdings hatte das Glück in diesem Experiment, einen überaus großen Einfluss auf den endgültigen Geldstand eines Spielers. Der Spieler konnte die meiste Zeit des Spiels, solange ihm der Anbauplatz auf dem Ackerfeld nicht ausging, seinen Geldstand exponentiell erhöhen, indem er jede Runde all sein Vermögen in weitere Anpflanzungen reinvestierte. Dies barg aber gleichzeitig ein großes Risiko und hatte zur Folge, dass wenn ein Spieler gegen Ende des Spiels sein gesamtes Geld investierte und eine einzige Fehlentscheidung traf, all sein Geld verloren war. Bis zum Ende des Spiels konnte er sich dann häufig nicht mehr von diesem Fehler erholen.

Beispielsweise könnte ein Spieler in Runde 7 von 10 bewusst das Risiko eingehen bei einer 60% Gewinnchance sein ganzes Vermögen in Anbauten zu investieren. Wenn dann aber die 40% Chance des schlechten Wetters eintritt verliert der Spieler alles und kann sich in diesem Spiel nicht mehr davon erholen. Der Spieler hat die Darstellung richtig interpretiert und statistisch gesehen trifft der Spieler keine falsche Entscheidung. Trotzdem erspielt der Teilnehmer nur einen geringen Geldstand, was bei einer Bewertung, die rein auf den Highscore bezogen ist, zu einer schlechten Bewertung der Wahrscheinlichkeitsdarstellung führen würde. Die Darstellung wäre also schlecht bewertet, obwohl die eigentliche Entscheidung des Spielers gut war. Aufgrund der geringen Teilnehmerzahl und dem hohen Einfluss des Glücks war eine Bewertung basierend auf dem Highscore nicht praktikabel.

## 6. Öffentliches Experiment



**Abbildung 6.1.:** Die untersuchten Darstellungen 1-4 - Darstellung 1: textueller Erwartungswert, Darstellung 2: Liniendiagramm des Erwartungswerts mit Spannweite als Flächenmodell, Darstellung 3: Histogramm in Form eines Balkendiagramms, Darstellung 4: kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen.

Bei einer höheren Teilnehmerzahl wäre eine Bewertung anhand des Highscores möglich gewesen, da sich das Rauschen im Highscore, welches durch die Unsicherheit bei vorhergesagtem und eingetretenem Wetter entsteht, mit der Menge der Daten ausgleichen würde.

Aus den genannten Gründen war es daher sinnvoll, die Qualität der Entscheidungen der Spieler anhand eines, vom Glück unabhängigen Kriteriums zu bewerten. Anstatt das Endergebnis des gesamten Spiels in Summe zu bewerten, wurden die einzelnen Anbauentscheidungen jeder Runde ausgewertet.

In jeder Runde wurde berechnet, welche Entscheidung auf Grund der angezeigten Darstellung erfolversprechend war und den maximalen Gewinn einbringen würde. Die so ermittelte, statistisch gute Entscheidung wurde mit der tatsächlich vom Spieler getroffenen Anbauentscheidung verglichen. Durch diese Vorgehensweise war die Bewertung komplett unabhängig vom Eintritt des Wetters und vom Glück des Spielers entkoppelt. Anstatt des wahren, tatsächlich erzielten Gewinns wurde der erwartete Gewinn einer Anbauentscheidung eines Spielers ermittelt, um eine Aussage über die Qualität seiner Entscheidung zu treffen. Roulston et. al [MSR06] sind in ihren Entscheidungs-Experimenten ähnlich vorgegangen, indem sie ebenfalls den erwarteten Gewinn an Stelle des tatsächlichen Ergebnisses beurteilten.

Pflanze	Gewinn pro Anbautag
Sonnenblume	36,67
Karotte/Melone	30
Salat	26,67
Mais	25
Weintrauben	20
Kürbis	15
Erdbeere	10

**Tabelle 6.1.:** Gewinne pro Anbautag für jede Pflanze. Wenn das Wetter es zuließ verschiedene Pflanzen anzubauen, war es sinnvoll, sich nach dieser Tabelle zu richten, um den höchsten Gewinn zu erzielen.

Das getrennte Auswerten jeder Runde hatte zusätzlich den positiven Nebeneffekt, dass sich die Datenmenge beinahe verzehnfachte. Dies war möglich, da nicht nur die Geldstände, angezeigten Darstellungen und Anbauentscheidungen für jede Runde gespeichert wurden, sondern auch die Rohdaten und Wahrscheinlichkeitsverteilungen aus denen der Wetterbericht erstellt wurde.

Um die optimale Anbauentscheidung für die jeweilige Runde zu ermitteln, wurde für jede Pflanze überprüft, ob sie mit dem vorhergesagten Wetter vereinbar ist. Als Referenzgröße wurde der Erwartungswert, des zu diesem Zeitpunkt dargestellten Wetters, gewählt, da dieser im Mittel richtig liegt.

Bei der Auswertung wurden aus den gespeicherten Vorhersagedaten der folgenden drei Tage der jeweiligen Runde, die Erwartungswerte errechnet und daraufhin ermittelt, welche Pflanzen potentiell erfolgreich wären. Auch der Spieler im Spiel hätte auf diese Weise seine Entscheidungen treffen können. In Darstellung 1 und 2 war der Erwartungswert direkt aus der Darstellung abzulesen. In Darstellung 3 und 4 war er zwar nicht direkt angegeben, jedoch ungefähr abzuschätzen, da die dargestellten Verteilungen normalverteilt waren. Aus den so ermittelten erfolgversprechenden Pflanzen wurde dann die optimale Entscheidung ermittelt, indem die Pflanze mit dem höchsten Gewinn pro Anbautag gewählt wurde. Der Gewinn pro Anbautag für eine Pflanze ermittelt sich dabei wie folgt:

$$x = \frac{a - b - c}{d}$$

Wobei  $x$  = [Gewinn pro Anbautag],  $a$  = [Wert der Pflanze],  $b$  = [Anbaukosten der Pflanze],  $c$  = [Kosten für das Pflügen],  $d$  = [Anbauzeit der Pflanze] und die Kosten für das Pflügen mit 10€ für jede Pflanze gleich war. Tabelle 6.1 zeigt die sortierten Gewinne pro Anbautag für jede Pflanze.

Nun wurde die Pflanzen ermittelt, die der Spieler in dieser Runde angebaut hatte und mit der optimalen Anbauentscheidung verglichen. Stimmt die Anbauentscheidung des Spielers und die optimale Entscheidung überein, so lag eine optimale Entscheidung des Spielers für diese Runde vor. Lag keine Übereinstimmung vor, kam es zu einer weiteren Auswertung. Bei dieser wurde überprüft, ob der Spieler eine suboptimale, aber trotzdem sinnvolle Entscheidung getroffen hat.

Theorien in der Ökonomie und Psychologie, wie beispielsweise die „Expected utility theory“ [Neu44] oder die „Prospect Theory“ [DK79] [AT92], legen nahe, dass die Menschen sich in der Regel nicht streng an den Erwartungswert halten, um Entscheidungen zu treffen. Außerdem neigen die Menschen dazu bei moderaten Wahrscheinlichkeiten eines Verlustes sich eher risikofreudig zu verhalten. Um suboptimale, aber sinnvolle Entscheidungen zu erkennen, wurde daher überprüft, ob eine Anbauentscheidung, je nach Risikoverhalten, sinnvoll war. Dazu wurden die Grenzwerte für Niederschlag, Temperatur und Windstärke der nächsten 3 Tage bestimmt, die unter dem gegebenen Risikoverhalten noch vertretbar gewesen wären. Anschließend wurden die errechneten Grenzwerte mit den Anforderungen an Niederschlag, Temperatur und Windstärke der vom Spieler angebauten Pflanze verglichen. Lagen die Anforderungen der vom Spieler angebauten Pflanze im gültigen Bereich, war also der Niederschlag und die Temperatur kleiner und die Windstärke größer als der Grenzwert, so war die Anbauentscheidung sinnvoll. Stimmt die Gleichung nicht überein, so wurde die Entscheidung als nicht sinnvoll gewertet.

Bei der Bewertung von suboptimalen, aber sinnvollen Entscheidungen wurde nach risikoaversen, risikoneutralem und risikofreudigem Verhalten unterschieden. Die exakte Berechnung der Grenzwerte nach denen entschieden wurde, ob ein Anbau suboptimal, aber sinnvoll war, ist in Tabelle 6.2 zu sehen. Die Berechnungen wurden abhängig von den Darstellungen und dem Risikoverhalten festgelegt. Wo es möglich war, wurde für das risikoaverse Verhalten ein Risiko von 0%, für das risikoneutrale Verhalten ein Risiko von 50% und für das risikofreudige Verhalten ein Risiko von 80% zu Grunde gelegt.

Da dieses Bewertungsverfahren jedoch differenzierte Wahrscheinlichkeitsangaben benötigt, musste für Darstellung 1, die nur den Erwartungswert zeigte, ein Näherungsverfahren verwendet werden. Morss et. al [REM08] stellten fest, dass die meisten Menschen vermuten, dass hinter deterministischen Wettervorhersagen ein möglicher Wetterbereich liegt. Für die deterministische Vorhersage der Temperatur von 75°F vermuteten die meisten Teilnehmer in den Befragungen einen Bereich von +/- 2°F. Eine weitere große Gruppe von über 20% vermuteten sogar einen Bereich von +/- 5°F. Daher wurde in Darstellung 1 (textueller Erwartungswert) für einen risikoaversen Spieler eine Schwankung von 2°C angenommen, welcher umgerechnet in Grad Celsius ungefähr dem Mittelwert der vermuteten Schwankungen der Teilnehmer aus den Befragungen von Morss et. al entspricht. Da die Temperaturerfordernis der Pflanzen immer eine Mindestangabe war, wurde der Grenzwert für die Temperatur bei risikoaversen Verhalten aus dem Erwartungswert abzüglich 2°C errechnet. Analog dazu, wurde der Niederschlag mit -1 mm berechnet, da die Spannweite des Niederschlags in mm genau die Hälfte der Spannweite der Temperatur in °C betrug. Die Windstärke war eine Maximalangabe und wurde mit +2km/h berechnet, da die Spannweite des Niederschlags in mm genau gleichgroß war, wie die Spannweite der Temperatur in km/h. Die Grenzwerte für das risikoneutrale Verhalten deckten sich mit den Erwartungswerten, die der Spieler angezeigt bekam. Die Grenzwerte für das risikofreudige Verhalten verhielten sich spiegelverkehrt zum risikoaversen Verhalten, da Morss et. al [REM08] zeigten, dass der Bereich, den die Teilnehmer hinter einer deterministischen Prognose vermuten, symmetrisch ist.

Darstellung 2 (Liniendiagramm des Erwartungswerts mit Spannweite als Flächendiagramm) zeigte zum einen den Erwartungswert direkt durch ein Liniendiagramm und zusätzlich die Spannweite der möglichen Werte als Flächendiagramm. Als Grenzwerte für Darstellung 2 bei risikoaversen Verhalten wurden für den Niederschlag und die Temperatur die kleinsten Werte der dargestellten

Darstellung	Risikoverhalten	Berechnung	
1	risikoavers	Niederschlag	Erwartungswert - 1 mm
		Temperatur	Erwartungswert - 2 °C
		Windstärke	Erwartungswert + 2 km/h
	risikoneutral	Niederschlag	Erwartungswert
		Temperatur	Erwartungswert
		Windstärke	Erwartungswert
	risikofreudig	Niederschlag	Erwartungswert + 1 mm
		Temperatur	Erwartungswert + 2 °C
		Windstärke	Erwartungswert - 2 km/h
2	risikoavers	Niederschlag	kleinster Wert der Spannweite
		Temperatur	kleinster Wert der Spannweite
		Windstärke	größter Wert der Spannweite
	risikoneutral	Niederschlag	Erwartungswert
		Temperatur	Erwartungswert
		Windstärke	Erwartungswert
	risikofreudig	Niederschlag	80% der Spannweite vom Erwartungswert bis zum größten Wert
		Temperatur	80% der Spannweite vom Erwartungswert bis zum größten Wert
		Windstärke	20% der Spannweite vom kleinsten Wert bis zum Erwartungswert
3	risikoavers	Niederschlag	kleinster Wert
		Temperatur	kleinster Wert
		Windstärke	größter Wert
	risikoneutral	Niederschlag	Median
		Temperatur	Median
		Windstärke	Median
	risikofreudig	Niederschlag	0.8-Quantil
		Temperatur	0.8-Quantil
		Windstärke	0.2-Quantil
4	risikoavers	Niederschlag	kleinster Wert
		Temperatur	kleinster Wert
		Windstärke	größter Wert
	risikoneutral	Niederschlag	Median
		Temperatur	Median
		Windstärke	Median
	risikofreudig	Niederschlag	0.8-Quantil
		Temperatur	0.8-Quantil
		Windstärke	0.2-Quantil

**Tabelle 6.2.:** Die Berechnungen der Grenzwerte für Niederschlag, Temperatur und Windstärke für jede Darstellung und jedes Risikoverhalten.

## 6. Öffentliches Experiment

---

Spannbreite gewählt. Es wird davon ausgegangen, dass der risikoaverse Spieler seine Entscheidung an den schlechtesten Fall anpasst, weil er den Verlust einer Pflanze unter allen Umständen vermeiden will. Analog dazu wird der Grenzwert für die Windstärke aus dem größten Wert der Spannweite berechnet. Die Grenzwerte für das risikoneutrale Verhalten wurden wieder direkt aus dem Erwartungswert, den der Spieler bei dieser Darstellung direkt angezeigt bekam, berechnet. Die Grenzwerte für das risikofreudige Verhalten wurden ähnlich, wie beim risikoaversen Verhalten ermittelt. Bei diesem Verhalten wurde ein Risiko von 80% als akzeptabel unterstellt. Da aber die exakte Verteilung bei dieser Darstellungsform nicht ersichtlich war, wurde der Grenzwert für Temperatur und Niederschlag bei 80% der Spannweite zwischen Erwartungswert und größtem Wert gesetzt. Analog dazu wurde der Grenzwert bei der Windstärke bei 20% der Spannweite vom kleinsten Wert bis zum Erwartungswert gesetzt.

Darstellung 3 (Histogramm in Form eines Balkendiagramms) und Darstellung 4 (kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen) vermitteln dem Spieler annähernd gleich viele Informationen über die Wahrscheinlichkeitsverteilung. Deshalb wurden die Grenzwerte für die zwei Darstellungen auf die gleiche Weise berechnet. Für den risikoaversen Spieler wurde der kleinste Wert, für den Niederschlag und die Temperatur, und der größte Wert für die Windstärke als Grenzwert gewählt. Diese Werte waren in beiden Darstellungen von den Spielern direkt abzulesen. Um den Grenzwert für den risikoneutralen Spieler zu bestimmen wurde der Median ermittelt. Dieser Grenzwert wurde für beide Darstellungen und alle drei Wetterarten gleich berechnet. Die Grenzwerte des risikofreudigen Spielers wurden für den Niederschlag und die Temperatur mit Hilfe des 0,8-Quantils bestimmt, da dieses in beiden Darstellungen grob abgeschätzt werden kann und dem zu Grunde gelegten 80% Risiko entsprach. Analog dazu wurde für die Berechnung des Grenzwertes für die Windstärke das 0,2-Quantil gewählt.

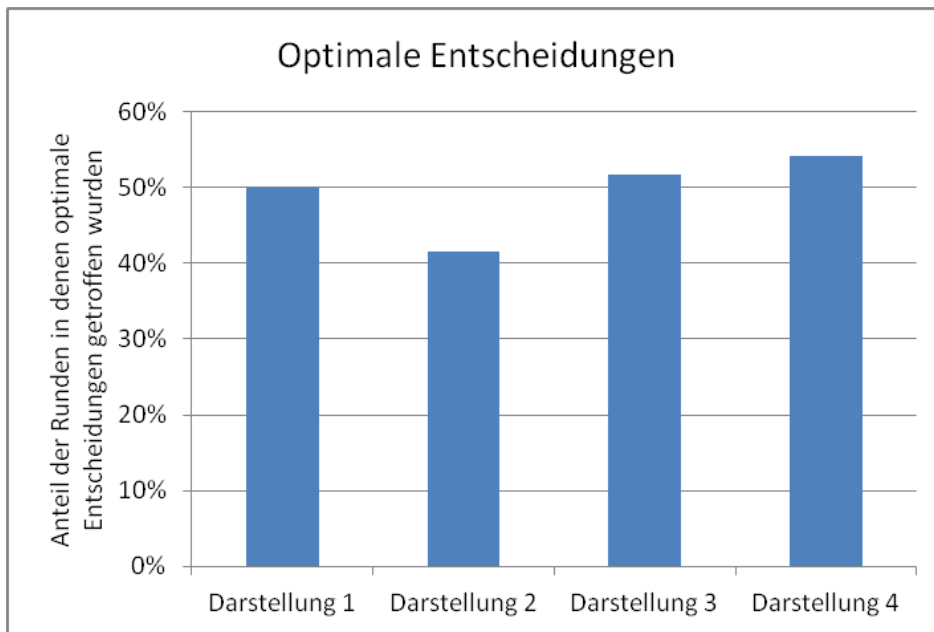
### 6.2. Veröffentlichung

Das Spiel wurde auf dem Server der HCI Gruppe gehostet und als frei zugängliche Facebook Anwendung veröffentlicht. Das Spiel war somit für jeden Facebook-Nutzer zugänglich. Die Spieler wurden nicht selektiert, so konnte jeder, der über den Link zum Spiel verfügte teilnehmen. Die einzige Voraussetzung war, sich mit einem gültigen Facebook Account vor Beginn des Spiels anzumelden. Der Link zum Spiel wurde über soziale Netzwerke wie Facebook, per Email und mündlich verbreitet.

### 6.3. Teilnehmer

Im Laufe des öffentlichen Experiments wurden insgesamt 518 Runden von 23 verschiedenen Teilnehmern gespielt. Davon waren 39,1% Frauen und 60,9% Männer. Das exakte Alter konnte bei dieser Umfrage nicht ermittelt werden, da Facebook nur den Zugriff auf Altersbereiche erlaubt. Es konnte lediglich festgestellt werden, dass 82,6% Prozent der Teilnehmer älter als 21 Jahre sind, 14% der Teilnehmer zwischen 18 und 21 Jahren alt sind und nur 4,4% jünger als 18 Jahre sind.





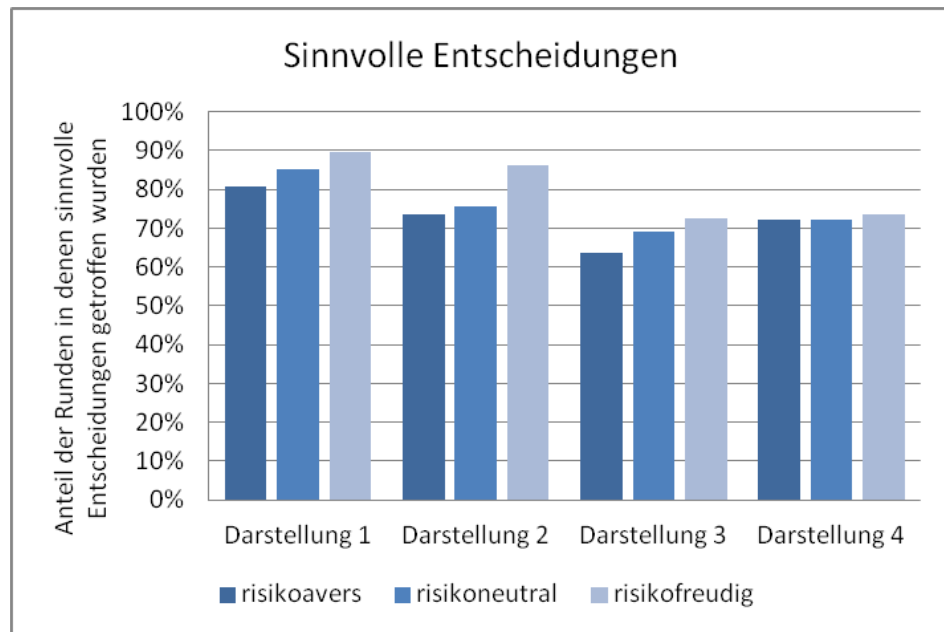
**Abbildung 6.2.:** Optimale Entscheidungen abhängig von der Darstellung im öffentlichen Experiment.

## 6.4. Ergebnisse

Nachfolgend werden die Ergebnisse des öffentlichen Experiments dargestellt. Hierbei wird die Qualität der Entscheidungen, die die Teilnehmer im Spiel getroffen haben, sowie die Dauer und Häufigkeit der Einsicht des Wetterberichts der Teilnehmer im Spiel analysiert. Weiterhin werden die Ergebnisse der kurzen Befragung der Teilnehmer am Ende des Spiels präsentiert. Sie erhielten dieselben vier Fragen, die schon bei der vorausgegangenen Online-Umfrage gestellt wurden. In die Auswertung wurden nur Runden einbezogen, in denen der Teilnehmer mindestens einmal den Wetterbericht eingeblendet hat.

### 6.4.1. Optimale Entscheidungen

In Abbildung 6.2 werden abhängig von der Darstellung der Anteil der Runden gezeigt, in denen die Spieler, die in Kapitel 6.1 definierten, optimalen Entscheidungen getroffen haben. Die Teilnehmer trafen die meisten optimalen Entscheidungen, wenn sie Darstellung 4 angezeigt bekommen hatten. Mit dieser Darstellung entschieden die Teilnehmer in 54,2% der Runden optimal. Mit Darstellung 3 wurde in 51,7% der Runden eine optimale Entscheidung gefällt. Mit der deterministischen Darstellung 1 trafen die Teilnehmer in exakt 50% der Runden eine optimale Entscheidung. Das Schlusslicht bildet Darstellung 2 bei diesem Experiment. Hierbei wurde in nur 41,5% der Runden eine optimale Entscheidung getroffen.



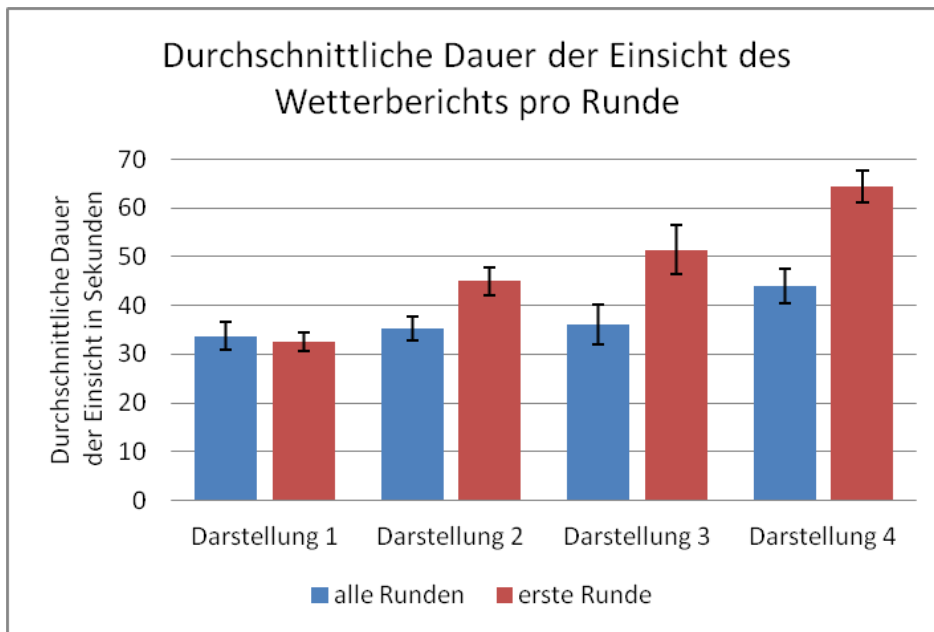
**Abbildung 6.3.:** Sinnvolle Entscheidungen abhängig von der Darstellung und dem Risikoverhalten im öffentlichen Experiment.

### 6.4.2. Suboptimale Entscheidungen

In Abbildung 6.3 werden, abhängig von der Darstellung und dem Risikoverhalten des Spielers der Anteil der Runden gezeigt, in denen die Spieler, die in Kapitel 6.1 definierten, suboptimalen, aber sinnvollen Entscheidungen getroffen haben. Natürlich sind diese Anteile sehr viel höher, da es meistens mehrere sinnvolle, aber immer nur eine optimale Entscheidung gibt. Darstellung 1 schneidet bei den suboptimalen Entscheidungen, sowohl bei risikoaversem, als auch bei risikofreudigem Verhalten, am besten ab. Bei risikoaversem Verhalten trafen die Teilnehmer in 80,7% der Runden sinnvolle Entscheidungen, bei risikoneutralem Verhalten in 85,2% der Fälle und beim risikofreudigem Verhalten entschieden sich sogar 89,9% richtig. Das Schlusslicht bildete Darstellung 3. Mit dieser Darstellung trafen die Teilnehmer, die risikoaverses Verhalten zeigten, in nur noch 72,2% der Fälle sinnvolle Entscheidungen, bei risikoneutralem Verhalten ebenfalls 72,2% und bei risikofreudigem Verhalten in 73,6% der Fälle.

### 6.4.3. Dauer der Einsicht des Wetterberichts

Die Spieler konnten sich während des Spiels in jeder Runde den Wetterbericht beliebig häufig über einen Button einblenden lassen. Um etwas anbauen zu können, mussten sie ihn manuell wieder ausblenden. Die Zeitspanne zwischen Einblenden und Ausblenden des Wetterberichtes wurde als Anzeigedauer gespeichert. Dadurch war es möglich, die durchschnittliche Dauer der Einsicht des Wetterberichts pro Runde zu ermitteln.



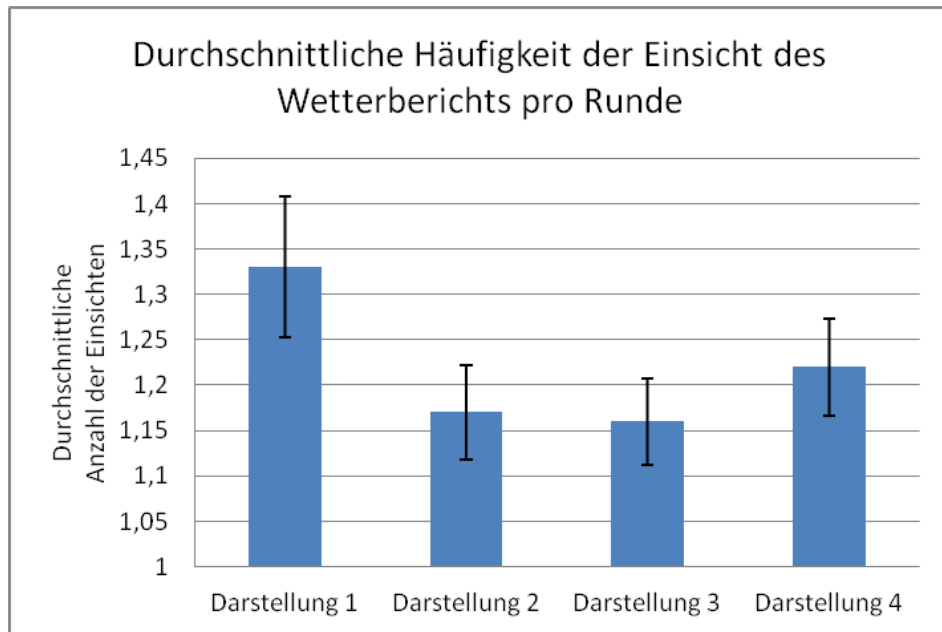
**Abbildung 6.4.:** Durchschnittliche Dauer der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im öffentlichen Experiment. Fehlerbalken zeigen den Standardfehler.

Abbildung 6.4 zeigt die durchschnittliche Anzeigedauer des Wetterberichts in Abhängigkeit von der Darstellung. Mit der sehr einfachen, deterministischen und textuellen Darstellung 1 benötigten die Spieler im Durchschnitt am wenigstens Zeit, um den Wetterbericht zu betrachten und eine Anbauentscheidung ableiten zu können. Sie wurde durchschnittlich nur 33,7 (SD: 26,3) Sekunden lang eingeblendet. Am zweitschnellsten konnten die Teilnehmer nach Einsicht von Darstellung 2 eine Entscheidung fällen. Hier benötigten die Teilnehmer durchschnittlich 35,2 (SD: 23,3) Sekunden. Darstellung 3 wurde im Durchschnitt 36 (SD: 31,9) Sekunden lang betrachtet und mit Darstellung 4 benötigten die Teilnehmer durchschnittlich 44 (SD: 30,7) Sekunden bis sie eine Entscheidung fällen konnten. Weiterhin wurde jeweils für die erste Runde die durchschnittliche Anzeigedauer ermittelt. Diese war mit allen Darstellungen, außer mit Darstellung 1, höher als die durchschnittliche Dauer aller Runden, was darauf zurück zu führen ist, dass die Teilnehmer sich in der ersten Runde mit der Darstellung vertraut machten und sich auf die Darstellung einstellten. Das Verhältnis der Darstellungen untereinander hat sich aber nicht verändert. Der Wetterbericht in der ersten Runde wurde mit Darstellung 1 durchschnittlich 32,6 (SD: 18,8) Sekunden angezeigt, mit Darstellung 2 durchschnittlich 45 (SD: 27,4) Sekunden, mit Darstellung 3 durchschnittlich 51,4 (SD: 38,7) Sekunden und mit Darstellung 4 durchschnittlich 64,5 (SD: 27,9) Sekunden.

## 6. Öffentliches Experiment

Darstellung 1	93%	93%	77%	77%	85%	75%	73%	60%	50%	27%
Darstellung 2	77%	78%	71%	69%	69%	75%	80%	70%	70%	20%
Darstellung 3	73%	57%	55%	64%	64%	45%	36%	50%	27%	18%
Darstellung 4	81%	69%	57%	60%	50%	67%	50%	42%	50%	9%
	Runde 1	Runde 2	Runde 3	Runde 4	Runde 5	Runde 6	Runde 7	Runde 8	Runde 9	Runde 10

**Abbildung 6.5.:** Anteil der Runden in denen der Wetterbericht mindestens einmal angezeigt wurde in Abhängigkeit der Darstellung und Runde im öffentlichen Experiment.

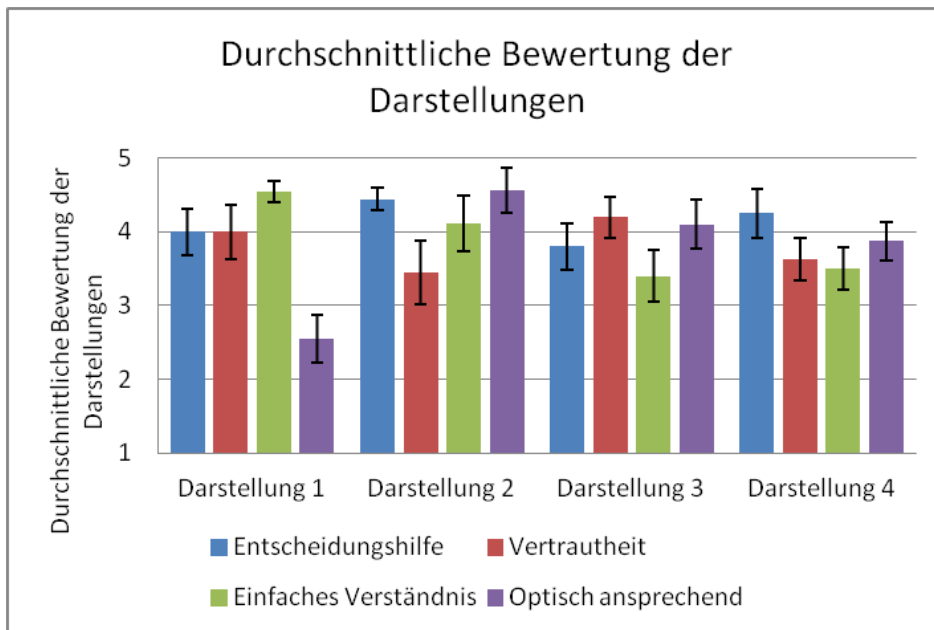


**Abbildung 6.6.:** Durchschnittliche Häufigkeit der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im öffentlichen Experiment. Fehlerbalken zeigen den Standardfehler.

### 6.4.4. Häufigkeit der Einsicht des Wetterberichts

Abbildung 6.5 zeigt die Anteile der Runden in denen der Wetterbericht mindestens einmal angezeigt wurde in Abhängigkeit der Darstellung und Runde. In Runde 1 werden alle Darstellungen in mindestens 73% der Runden angeschaut. In der letzten Runde 10 waren es abhängig von der Darstellung nur noch 9% bis 27% der Runden. Dies lag daran, dass in der letzten Runde nur noch Pflanzen gerntet werden konnten, aber nicht mehr angebaut, weshalb eine Einsicht des Wetterberichtes unnötig wurde. Darstellung 3 wurde speziell in den letzten Runden im Vergleich zu den anderen Darstellungen weniger eingeblendet.

Weiterhin wurde gemessen und gespeichert, wie häufig sich die Teilnehmer den Wetterbericht einblenden ließen. Es wurden nur Runden ausgewertet wurden, in denen der Wetterbericht mindestens einmal angezeigt wurde, wodurch die Mindestanzahl eine Einsicht ist. In Abbildung 6.6 wird abhängig



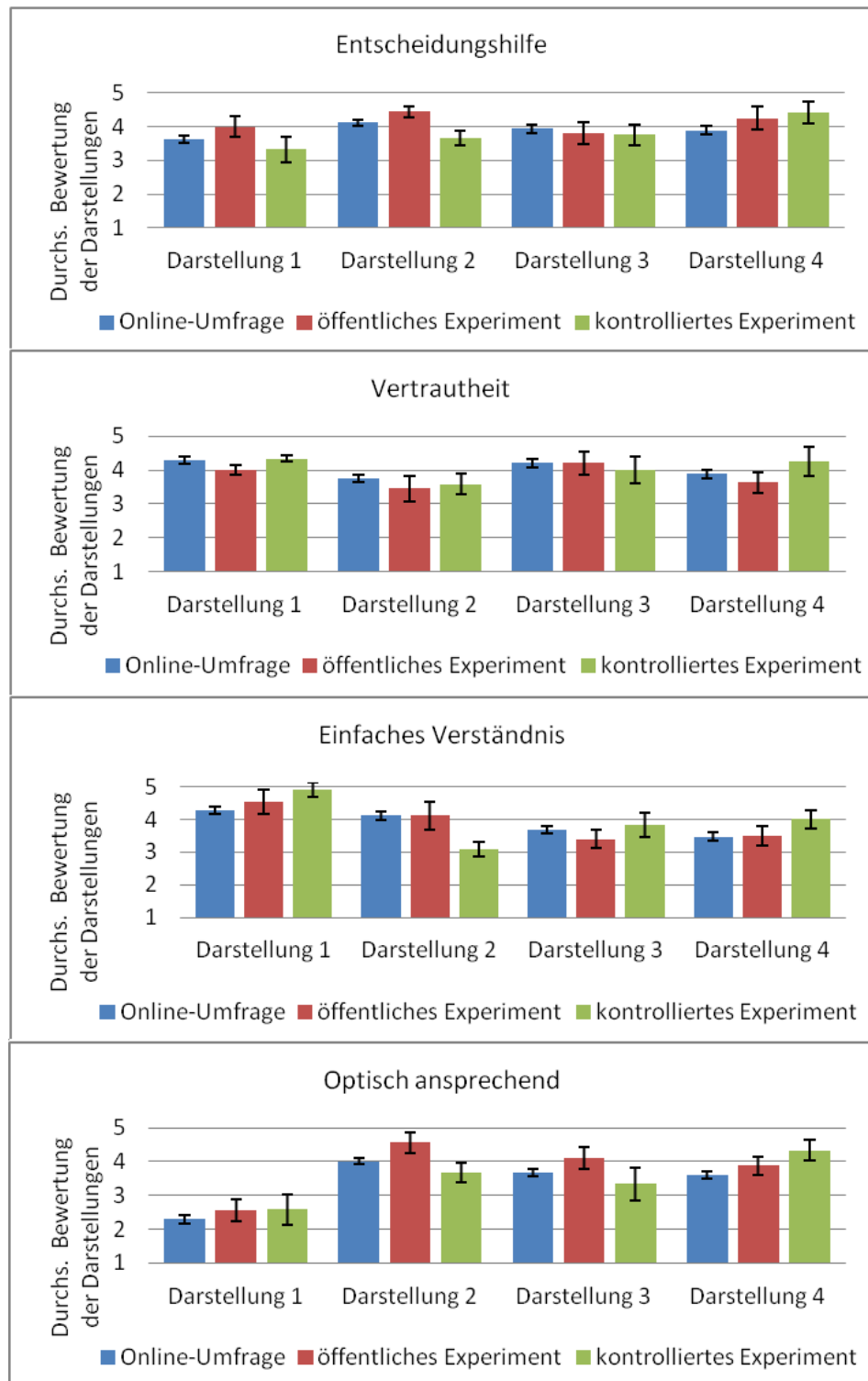
**Abbildung 6.7.:** Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der vier Darstellungen, die die Teilnehmer abgegeben hatten, nachdem sie ein Spiel vollständig mit der jeweiligen Darstellung gespielt hatten. Die Teilnehmer wurden gefragt, ob die Darstellung ihnen geholfen hatte eine Entscheidung zu treffen, ob die Darstellung ihnen bereits vertraut war, ob die Darstellung für sie einfach zu verstehen war und ob die Darstellung optisch ansprechend war. Fehlerbalken zeigen den Standardfehler.

von der Darstellung die durchschnittliche Häufigkeit der Einsicht des Wetterberichts pro Runde dargestellt. Darstellung 2 wurde mit durchschnittlich nur 1,16 (SD: 0,5) mal am seltensten eingeblendet. Darstellung 3 wurde nur knapp öfter nämlich durchschnittlich 1,17 (SD: 0,36) mal eingeblendet. Darstellung 4 wurde durchschnittlich 1,22 (SD: 0,45) mal eingeblendet und Darstellung 1 wurde durchschnittlich 1,33 (SD: 0,73) mal eingeblendet.

#### 6.4.5. Fragestellungen aus der Online-Umfrage

In Abbildung 6.7 werden die durchschnittlichen Bewertungen der vier Darstellungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) gezeigt, die die Teilnehmer abgegeben hatten, nachdem sie ein Spiel vollständig mit der jeweiligen Darstellung gespielt hatten. Die Fragen waren die selben wie in der Online-Umfrage aus Kapitel 4. Die Teilnehmer wurden gefragt, ob die Darstellung ihnen bei der Entscheidungsfindung geholfen hat, ob ihnen die Darstellung bereits vertraut war, ob die Darstellung für sie einfach zu verstehen war und ob sie die Darstellung optisch ansprechend fanden. Abbildung 6.8 zeigt eine Gegenüberstellung der durchschnittlichen Bewertungen der Teilnehmer

## 6. Öffentliches Experiment



**Abbildung 6.8.:** Gegenüberstellung der durchschnittlichen Bewertungen aus der vorausgegangen Online-Umfrage, dem öffentlichen Experiment und dem kontrollierten Experiment bezüglich der vier Fragestellungen: Entscheidungshilfe, Vertrautheit, Einfaches Verständnis, Optisch ansprechend. Fehlerbalken zeigen den Standardfehler.

aus der vorausgegangenen Online-Umfrage, dem öffentlichen Experiment und dem kontrollierten Experiment.

## 6.5. Kommentare der Teilnehmer

Lediglich zwei konstruktive Teilnehmer haben einen Kommentar hinzugefügt. Ein Teilnehmer merkte an, dass die Blautöne in Darstellung 3 kaum zu unterscheiden seien. Der zweite Teilnehmer fand, dass das Wetter zu häufig schlecht sei und er so nur Erdbeeren anbauen konnte.

## 6.6. Diskussion

Die Analyse der Daten hat gezeigt, dass die Teilnehmer abhängig von der Darstellung unterschiedlich viele und gute Entscheidungen getroffen haben.

Abbildung 6.3, vergleicht die vier Darstellungen in Bezug auf das Risikoverhalten des Spielers. Hier fällt auf, dass der Anteil der sinnvollen Entscheidungen pro Runde mit der Risikobereitschaft ansteigt. Dieses Ergebnis ist gut nachvollziehbar, da bei einem risikofreudigen Verhalten die absolute Anzahl an möglichen sinnvollen Entscheidungen am größten ist. Weiterhin wird aus der Analyse dieser Darstellung aber auch ersichtlich, dass das Risikoverhalten des Spielers keinen Einfluss auf das relative Verhältnis der Darstellungen zueinander hatte. Die Rangordnung der Darstellungen bleibt, in Bezug auf den Anteil an sinnvollen Entscheidungen, unabhängig vom Risikoverhalten immer dieselbe, insgesamt am besten schnitt Darstellung 1 ab.

Die Teilnehmer, die mit Darstellung 1 spielten, trafen unabhängig vom Risikoverhalten die meisten sinnvollen Entscheidungen. Das lässt für diesen Kontext, dieses Experiment und dieses Entscheidungsszenario den Schluss zu, dass es für die Entscheidungsfindung ausreichend ist, die Daten nur deterministisch, ohne Angabe von Wahrscheinlichkeiten darzustellen. Im Falle dieses Experimentes enthielt die deterministische Darstellung den Erwartungswert für den Niederschlag, die Temperatur oder die Windstärke.

Dass die meisten Menschen mit Unsicherheiten umgehen und darauf basierend Entscheidungen treffen können, stellten Roulston et. al [MSR06] und Joslyn et. al [JL12] in ihren Arbeiten fest. Die Ergebnisse ihrer Arbeiten, auf die in Kapitel 3.1 bereits näher eingegangen wurde, konnten von den Erkenntnissen, die in der vorliegenden Arbeit gewonnen wurden, bestätigt werden. Schließlich trafen die Teilnehmer, bei risikoneutralem Verhalten, unabhängig von der Darstellung, in knapp 70% der Runden eine sinnvolle Entscheidung auf Basis von unsicheren Daten. Bei Darstellung 1 entschieden sich die Teilnehmer sogar in über 85% der Runden sinnvoll.

Morss et. al [REM08] fanden bei einer Befragung heraus, dass die Teilnehmer probabilistische Wettervorhersagen den deterministischen Wettervorhersagen vorziehen. In dem hier vorgestellten Experiment, mussten die Teilnehmer sowohl aufgrund von deterministischen als auch aufgrund von probabilistischen Wettervorhersagen echte Entscheidungen treffen. Es stellte sich heraus, dass anteilig die meisten optimalen Entscheidungen basierend auf probabilistischen Darstellungen getroffen

## 6. Öffentliches Experiment

---

wurden. Optimale Entscheidungen werden also durch möglichst exakte und aufschlussreiche Informationen über die Unsicherheit, wie sie Darstellung 3 und 4 enthielten, unterstützt. Die Einschätzung der Teilnehmer der Umfrage von Morss et. al wurde demnach von den Ergebnissen der tatsächlichen Entscheidungen in diesem Experiment bestätigt. Diese Aussage gilt jedoch nur für optimale Entscheidungen.

Der Anteil an lediglich sinnvollen Entscheidungen war dagegen am größten, wenn die Teilnehmer mit Darstellung 1 spielten. Basierend auf deterministischen Darstellungen wurden also insgesamt mehr sinnvolle Entscheidungen getroffen als auf Basis von probabilistischen Darstellungen. Nur um optimale Entscheidungen, wie sie in Abbildung 6.2 gezeigt werden, zu treffen war es von Vorteil den Teilnehmern exaktere und aufschlussreichere Informationen über die Unsicherheit zu präsentieren.

Außerdem wurden die verschiedenen Darstellungen in Hinblick auf die Dauer der Wettereinsicht analysiert. Hier wurde wie erwartet festgestellt, dass die Teilnehmer die deterministische Darstellung 1 im Durchschnitt am kürzesten betrachteten, um eine Anbauentcheidung abzuleiten. Sie enthielt die wenigsten Informationen und für jeden der drei vorhergesagten Tage war jeweils nur ein Wert in Textform abzulesen. Diesen konnten die Teilnehmer effizient und ohne Umrechnen, Überschlagen oder Interpretationen mit den Anbaubedingungen der Pflanzen vergleichen. Ähnlich einfach ist auch Darstellung 2 aufgebaut, die ermöglichte den Teilnehmern vermutlich ebenfalls relativ schnell eine Entscheidung zu fällen. Bei den Darstellungen 3 und 4, die sehr viel komplexer aufgebaut sind und mehr Informationen enthalten, war die durchschnittliche Betrachtungsdauer etwas höher. Trotzdem wurde der Wetterbericht bei Spielen die mit Darstellung 4 gespielt wurden durchschnittlich nur ca. 30,7% länger eingeblendet als bei den Spielen, die mit Darstellung 1 gespielt wurden.

Man kann also davon ausgehen, dass wenn keine optimale Entscheidung getroffen werden muss, die Menschen mit deterministischen Angaben schneller und besser zurecht kommen. Wenn aber optimale Entscheidungen getroffen werden müssen, in denen die Unsicherheit eine größere Rolle spielt, treffen die Menschen bessere Entscheidungen, wenn ihnen detaillierte Informationen über die Unsicherheiten zur Verfügung gestellt werden. In diesem Experiment waren das die genauer aufgeschlüsselten Wahrscheinlichkeitsverteilungen für die Niederschlagsmenge, die Temperatur und die Windstärke der nächsten drei Tage.

In der ersten Runde des Spiels sehen die Teilnehmer eine Darstellung zum ersten Mal. Sie müssen die gezeigten Informationen ordnen, interpretieren und bestmöglich in ihren Entscheidungsprozess einbeziehen. Die durchschnittlich Betrachtungsdauer in der ersten Runde lässt Rückschlüsse darauf zu, wie viel Zeit die Teilnehmer benötigen um eine Darstellung zu verstehen. Vergleicht man die durchschnittliche Dauer der Einblendung des Wetterberichts der unterschiedlichen Darstellungen in der ersten Runde, zeigt sich, dass Darstellung 1 durchschnittlich am wenigsten lange betrachtet wurde. Man kann deshalb darauf schließen, dass Darstellung 1 auch bei der erstmaligen Ansicht eines Spielers schneller und intuitiver interpretiert werden kann, als die anderen Darstellungen. Dies ist vermutlich darauf zurück zu führen, dass erstens sehr wenige Informationen dargestellt werden und der Erwartungswert zweitens ein sehr gebräuchliche Angabe ist. Je komplexer die Darstellungen waren, desto länger wurden sie in der ersten Runde eingeblendet.

Weiterhin wurde ermittelt, in wie viel Prozent der Runden der Wetterbericht überhaupt eingeblendet wurde. Das Ergebnis wurde in Abbildung 6.5 dargestellt. Hierbei fällt auf, dass Darstellung 3 und 4 besonders in den letzten Runden seltener überhaupt eingeblendet wurden. Dies lässt vermuten,



dass die Teilnehmer sich am Anfang des Spiels noch Mühe gegeben haben und die Darstellung öfter eingeblendet haben, aber später das Interesse an dieser Darstellung verloren, da sie ihnen eventuell nicht zusagte oder zu mühsam zu interpretieren waren. Im Gegensatz dazu schien das Interesse an den einfacheren Darstellungen, die die Unsicherheit gar nicht oder nur sehr vereinfacht darstellten, größer zu sein.

Zusätzlich wurde auch die Häufigkeit der Einblendung des Wetterberichtes im Spiel erfasst. Bei Spielen mit Darstellung 1 wurde der Wetterbericht am häufigsten eingeblendet. Die Darstellungen 2 und 3 blendeten die Teilnehmer am seltensten ein. Eine Erklärung für dieses Phänomen wurde nicht gefunden.

Im Anschluss an das Spiel beantworteten die Spieler dieselben Fragen wie die Teilnehmer der Online-Umfrage. Die Bewertungen der Teilnehmer der Online-Umfrage und des Spiels waren sehr ähnlich. Mit mindestens 3,8 bewerteten die Teilnehmer die Unterstützung bei der Entscheidungsfindung aller vier Darstellungen überdurchschnittlich gut. Darstellung 2 schnitt hier, wie bei der Online-Umfrage bereits, am besten ab (vgl. Abbildung 6.8). Interessanterweise deckt sich diese Einschätzung nicht mit den Ergebnissen bezüglich der tatsächlichen Anzahl an getroffenen optimalen, als auch suboptimalen, Entscheidungen der Teilnehmer im Spiel. Im Hinblick darauf, wie vertraut die verschiedenen Darstellungen den Teilnehmern waren, unterschieden sich die Bewertungen für die vier Darstellungen kaum. Das Histogramm in Form eines Balkendiagrammes, also Darstellung 3 war den Spieler am vertrautesten. Bei der Frage nach dem Verständnis schnitt Darstellung 1 am besten ab, wie erwartet war diese am einfachsten zu verstehen. Gleichzeitig fanden die Teilnehmer Darstellung 1 am wenigsten ansprechend. Auch dies war zu erwarten, da die Darstellung keine graphischen Elemente sondern nur Text enthielt. Die Bewertung der graphischen Darstellungen unterschied sich dagegen kaum.



## 7. Kontrolliertes Experiment

Bei diesem Teilbereich der Untersuchung wurde das Experiment kontrolliert und moderiert durchgeführt. Dies sollte sicherstellen, dass jeder Spieler mit jeder der vier möglichen Darstellung spielte. Die Abfolge der Darstellungen war bei jedem Teilnehmer unterschiedlich, wobei die Reihenfolge der Darstellungen jedoch gleichmäßig verteilt war. Außerdem war es so möglich noch bessere demographische Informationen und klareres Feedback zu erhalten.

### 7.1. Auswertung der Daten

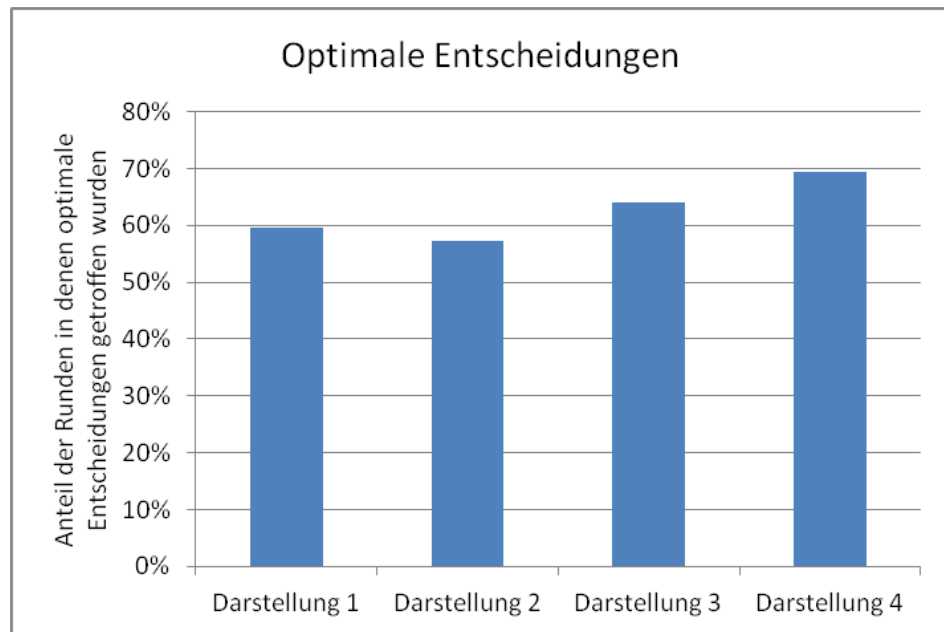
Die Auswertung fand auf genau die gleiche Weise wie beim öffentlichen Experiment in Kapitel 6.1 statt. Die optimalen und suboptimalen Entscheidungen wurden nach den selben Richtlinien ermittelt.

### 7.2. Veröffentlichung

Das Spiel war nicht öffentlich zugänglich. Die Teilnehmer wurden zufällig ausgewählt und um Teilnahme gebeten. Als kleinen Anreiz für die Teilnahme wurden zwei Gutscheine verlost. Ein Gutschein sollte unter allen Teilnehmern verlost werden und der andere sollte der Teilnehmer mit dem höchsten Geldstand bekommen. Für jeden Teilnehmer wurde ein eigener Link mit einer eindeutigen Teilnehmer-ID erstellt, der für die Teilnahme zwingend benötigt wurde, die Teilnahme an sich war jedoch freiwillig. Der Link zur Teilnahme am Spiel wurde den betroffenen Personen per E-Mail mitgeteilt. Weiterhin wurden die Teilnehmer in dieser E-Mail gebeten, die Spiele in einer vorhergesehenen Reihenfolge zu spielen und im Anschluss an die Spiele in einer Antwort-Email ihr Alter, Geschlecht und den beruflichen Hintergrund mitzuteilen und zu berichten, welche Darstellung ihnen persönlich am meisten zugesagt hat.

### 7.3. Teilnehmer

An dem kontrollierten Experiment beteiligten sich insgesamt zwölf Menschen, darunter waren vier Frauen und acht Männer. Die Teilnehmer waren zwischen 23 und 54 Jahren alt, wobei das Durchschnittsalter bei 32,3 Jahren (SD: 12,4) lag. Der berufliche Hintergrund der Teilnehmer war breit gefächert. Unter den Teilnehmern befanden sich ein Informatiker, ein Einzelhändler, drei Personen mit betriebswirtschaftlichem Hintergrund, drei Personen mit naturwissenschaftlichem Hintergrund und vier Personen, die im Dienstleistungssektor tätig sind.



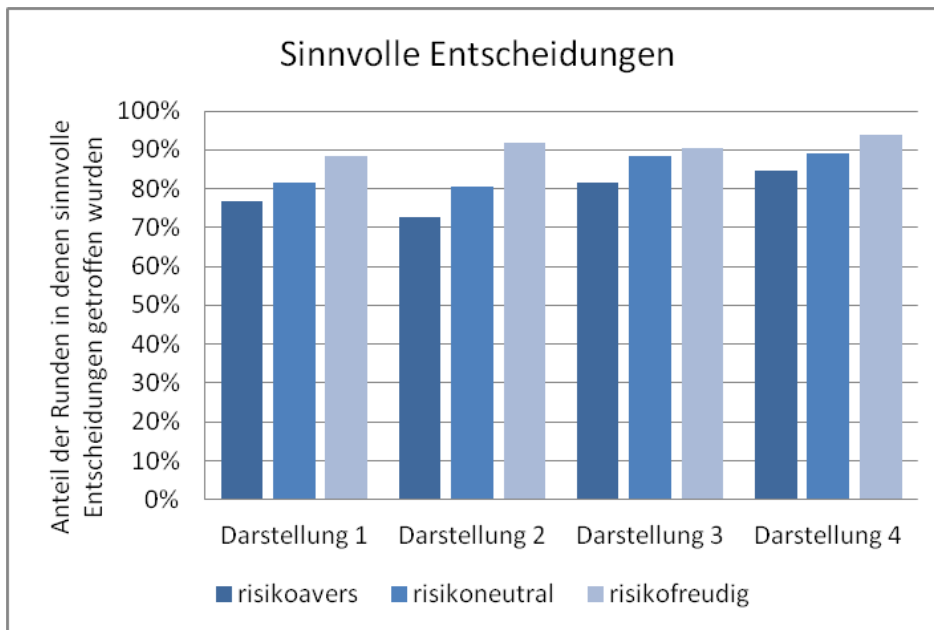
**Abbildung 7.1.:** Optimale Entscheidungen abhängig von der Darstellung im kontrollierten Experiment.

### 7.4. Ergebnisse

Im Folgenden werden die Ergebnisse des kontrollierten Experiments dargestellt. Hierbei wird die Qualität der Entscheidungen, die die Teilnehmer im Spiel getroffen haben, sowie die Dauer und Häufigkeit der Einsicht des Wetterberichts der Teilnehmer im Spiel analysiert. Weiterhin werden die Ergebnisse der kurzen Befragung der Teilnehmer am Ende des Spiels präsentiert. Diese erhielten dieselben vier Fragen, die schon bei der vorausgegangenen Online-Umfrage gestellt wurden. In die Auswertung wurden nur Runden einbezogen, in denen der Teilnehmer mindestens einmal den Wetterbericht eingeblendet hat. Dies ergibt eine Anzahl von 442 gewerteten Runden in diesem Experiment. Weiterhin wurde darauf geachtet, dass jeder Spieler mit jeder Darstellung genau einmal spielt. Die Reihenfolge mit der die Spieler auf die Darstellungen trafen wurde im Voraus bestimmt, um eine gleichmäßige Verteilung der Darstellungen an die Spieler zu gewährleisten. Das heißt, dass auf alle Spieler gesehen, jede Darstellung gleich häufig an der selben Position auf einen Spieler traf.

#### 7.4.1. Optimale Entscheidungen

In Abbildung 7.1 wird abhängig von der Darstellung der Anteil der Runden gezeigt, in denen die Spieler, die in Kapitel 6.1 definierten, optimalen Entscheidungen getroffen haben. Darstellung 4 schnitt bei diesem Vergleich am besten ab, mit dieser Darstellung entschieden die Teilnehmer in 69,4% der Runden optimal. Mit Darstellung 3 waren es 64,1% der Runden. Mit der deterministischen Darstellung

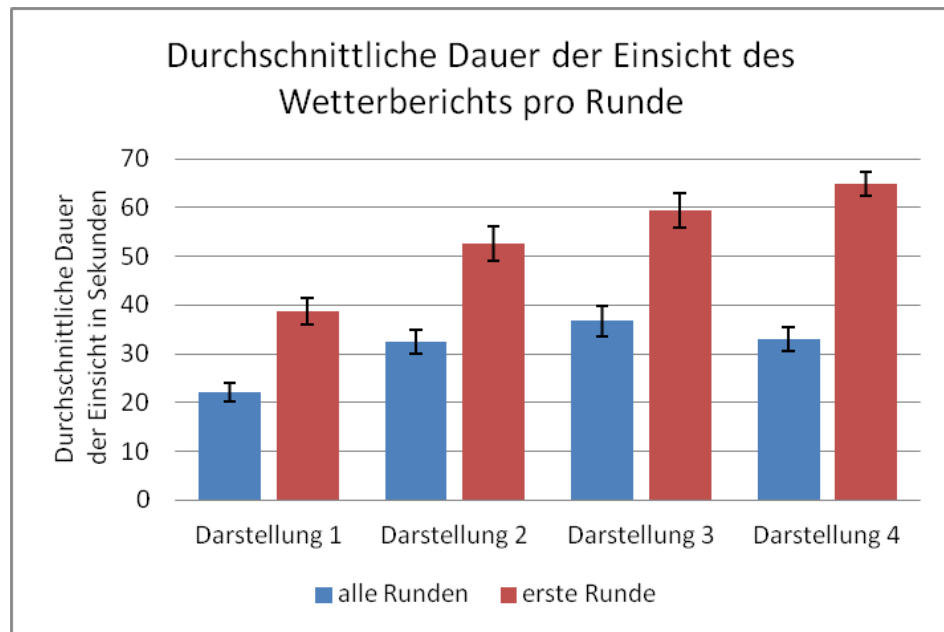


**Abbildung 7.2.:** Sinnvolle Entscheidungen abhängig von der Darstellung und dem Risikoverhalten im kontrollierten Experiment.

1 trafen die Teilnehmer immer noch in 59,6% der Runden eine optimale Entscheidung. Am wenigsten optimale Entscheidungen, nämlich nur 57,3%, haben die Teilnehmer mit Darstellung 2 getroffen.

#### 7.4.2. Suboptimale Entscheidungen

In Abbildung 7.2 wird abhängig von der Darstellung und dem Risikoverhalten des Spielers der Anteil der Runden gezeigt, in denen die Spieler, die in Kapitel 6.1 definierten, suboptimalen, aber sinnvollen Entscheidungen getroffen haben. Darstellung 4 schnitt bezüglich der Häufigkeit der suboptimalen Entscheidungen unabhängig vom Risikoverhalten am besten ab. Risikoaverse Teilnehmer trafen mit Darstellung 4 in 84,7% der Runden sinnvolle Entscheidungen, bei risikoneutralem Verhalten in 89,2% und bei risikofreudigem Verhalten sogar in 93,7% der Fälle. Mit Darstellung 3 trafen die Teilnehmer bei risikoavertem Verhalten in 81,6% der Fälle sinnvolle Entscheidungen, bei risikoneutralem Verhalten in 88,3% und bei risikofreudigem Verhalten in 90,3% der Fälle. Wenn die Spieler mit Hilfe von Darstellung 2 Entscheidungen treffen sollten, so trafen sie bei risikoavertem Verhalten in 72,6% der Fälle sinnvolle Entscheidungen, bei risikoneutralem Verhalten in 80,6% und bei risikofreudigem Verhalten in 91,9% der Fälle. Mit Darstellung 1 entschieden sich die Teilnehmer bei risikoavertem Verhalten in 76,9% der Fälle sinnvoll, bei risikoneutralem Verhalten in 81,7% und bei risikofreudigem Verhalten in 88,5% der Fälle.



**Abbildung 7.3.:** Durchschnittliche Dauer der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im kontrollierten Experiment. Fehlerbalken zeigen den Standardfehler.

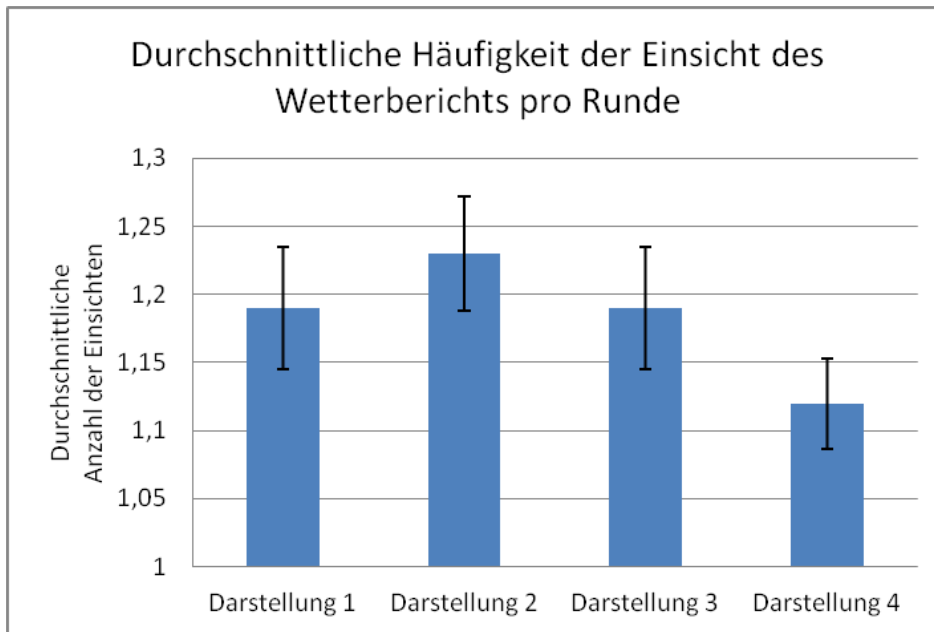
### 7.4.3. Dauer der Einsicht des Wetterberichts

Die Spieler konnten sich während des Spiels in jeder Runde den Wetterbericht beliebig häufig über einen Button einblenden lassen. Um etwas anbauen zu können, mussten sie ihn manuell wieder ausblenden. Die Zeitspanne zwischen Einblenden und Ausblenden des Wetterberichtes wurde als Anzeigedauer gespeichert. Dadurch war es möglich, die durchschnittliche Dauer der Einsicht des Wetterberichts pro Runde zu ermitteln.

In Abbildung 7.3 wird die durchschnittliche Dauer in Abhängigkeit von der Darstellung gezeigt. Die deterministische und textuelle Darstellung 1 wurde mit 22,1 (SD: 20,4) Sekunden im Mittel am kürzesten von den Spielern betrachtet, um eine Anbauentscheidung abzuleiten. Am zweitschnellsten konnten die Teilnehmer nach Einsicht von Darstellung 2 eine Entscheidung fällen. Hier benötigten die Teilnehmer durchschnittlich 32,4 (SD: 27,7) Sekunden. Darstellung 4 wurde im Durchschnitt 33,0 (SD: 25,6) Sekunden lang betrachtet und mit Darstellung 3 benötigten die Teilnehmer durchschnittlich 36,7 (SD: 31,1) Sekunden bis sie eine Entscheidung fällen konnten. Weiterhin wurde die durchschnittliche Dauer der Einsicht des Wetterberichtes in der ersten Runde ermittelt. Diese war bei allen Darstellungen deutlich höher als die durchschnittliche Dauer der Einsicht aller Runden. Der Wetterbericht in der ersten Runde wurde mit Darstellung 1 durchschnittlich 38,8 (SD: 27,7) Sekunden angezeigt, mit Darstellung 2 durchschnittlich 52,5 (SD: 39,5) Sekunden, mit Darstellung 3 durchschnittlich 59,4 (SD: 35,3) Sekunden und mit Darstellung 4 durchschnittlich 64,9 (SD: 25,7) Sekunden.

Darstellung 1	100%	75%	92%	83%	100%	92%	92%	92%	83%	58%
Darstellung 2	100%	100%	93%	100%	86%	79%	100%	92%	92%	75%
Darstellung 3	92%	100%	83%	83%	92%	92%	100%	100%	75%	42%
Darstellung 4	100%	100%	100%	100%	92%	83%	92%	100%	92%	67%
	Runde 1	Runde 2	Runde 3	Runde 4	Runde 5	Runde 6	Runde 7	Runde 8	Runde 9	Runde 10

**Abbildung 7.4.:** Anteil der Runden in denen der Wetterbericht mindestens einmal angezeigt wurde in Abhängigkeit der Darstellung und Runde im kontrollierten Experiment.

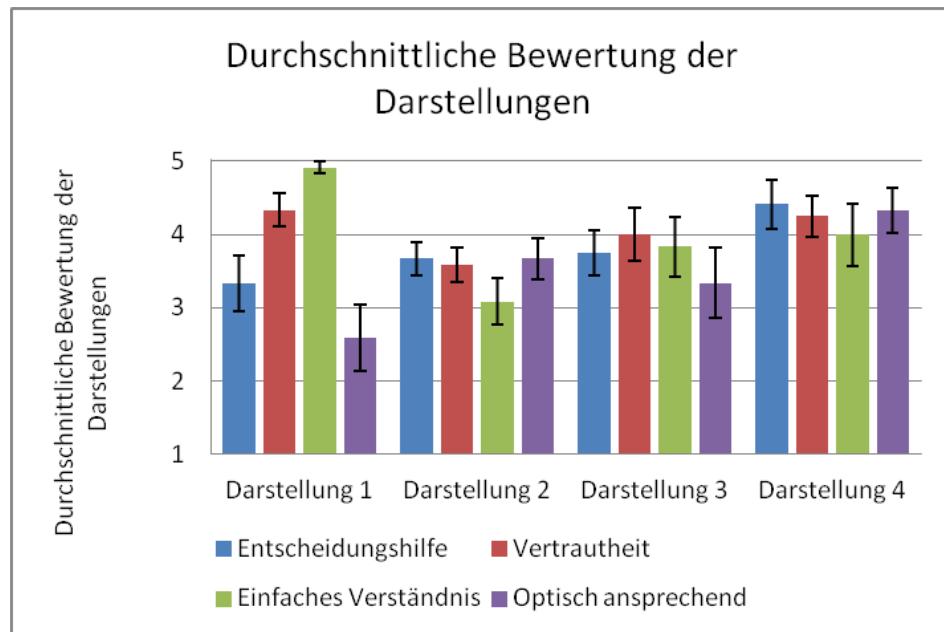


**Abbildung 7.5.:** Durchschnittliche Häufigkeit der Einsicht des Wetterberichts pro Runde abhängig von der Darstellung im kontrollierten Experiment. Fehlerbalken zeigen den Standardfehler.

#### 7.4.4. Häufigkeit der Einsicht des Wetterberichts

Abbildung 7.4 zeigt, in Abhängigkeit der Darstellung und Runde, die Anteile der Runden in denen der Wetterbericht mindestens einmal angezeigt wurde. In den Runden 1 bis 9 wurde der Wetterbericht durchweg in mindestens 75% der Runden eingeblendet, häufig sogar in jeder Runde. In der letzten Runde 10 lagen die Einblendungen des Wetterberichts abhängig von der Darstellung nur noch zwischen 42% und 75% der Runden.

Weiterhin wurde gemessen und gespeichert, wie häufig sich die Teilnehmer den Wetterbericht einblenden ließen. Da in die Auswertung nur Runden einbezogen wurden, in denen der Wetterbericht mindestens einmal angezeigt wurde, ist die Mindestanzahl eine Einsicht. In Abbildung 7.5 wird abhängig von der Darstellung die durchschnittliche Häufigkeit der Einsicht des Wetterberichts pro Runde dargestellt. Darstellung 4 wurde mit durchschnittlich nur 1,12 (SD: 0,35) Mal am seltensten



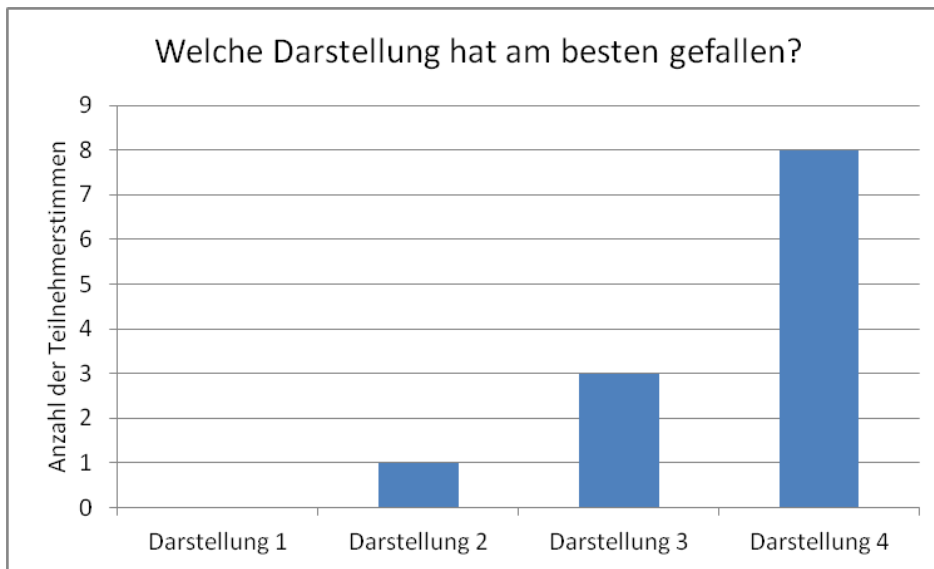
**Abbildung 7.6.:** Durchschnittliche Bewertungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) der vier Darstellungen, die die Teilnehmer abgegeben hatten, nachdem sie ein Spiel vollständig mit der jeweiligen Darstellung gespielt hatten. Die Teilnehmer wurden gefragt, ob die Darstellung ihnen geholfen hatte eine Entscheidung zu treffen, ob die Darstellung ihnen bereits vertraut war, ob die Darstellung für sie einfach zu verstehen war und ob die Darstellung optisch ansprechend war. Fehlerbalken zeigen den Standardfehler.

eingebildet. Darstellung 1 und 3 wurden gleich häufig, nämlich durchschnittlich 1,19 (SD: 0,46) Mal, eingebildet. Darstellung 2 wurde durchschnittlich 1,23 (SD: 0,47) Mal eingebildet.

### 7.4.5. Fragestellungen aus der Online-Umfrage

Nach dem die Teilnehmer ein Spiel vollständig mit der jeweiligen Darstellung gespielt hatten mussten sie die Darstellung in vier Fragen bewerten. In Abbildung 7.6 werden die durchschnittlichen Bewertungen der vier Darstellungen auf einer 5-Level Likert Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) gezeigt. Die Fragen waren die selben wie in der Online-Umfrage aus Kapitel 4. Die Teilnehmer wurden gefragt, ob die Darstellung ihnen half eine Entscheidung zu treffen, ob die Darstellung ihnen bereits vertraut war, ob die Darstellung für sie einfach zu verstehen war und ob sie die Darstellung optisch ansprechend fanden. Darstellung 2 wurde von den Teilnehmern als größte Entscheidungshilfe bewertet. Darstellung 3 bewerteten die Teilnehmer in Sachen Vertrautheit am höchsten. Die Teilnehmer fanden, dass Darstellung 1 am einfachsten zu verstehen ist. Darstellung 4 wurde von den Teilnehmern am optisch ansprechendsten bewertet.





**Abbildung 7.7.:** Die Verteilung der zwölf Teilnehmerstimmen im kontrollierten Experiment bezüglich der Frage, welche Darstellung ihnen am besten gefallen hat.

#### 7.4.6. Befragung nach der besten Darstellung

Die Teilnehmer des kontrollierten Experiments haben, im Gegensatz zu den Teilnehmern des öffentlichen Experiments, mit jeder der vier Darstellungen gespielt. Nach sie mit allen vier Darstellungen gespielt hatten, wurden sie gefragt, welche Darstellung ihnen allgemein am besten gefallen hat. Die Verteilung der zwölf Teilnehmerstimmen ist in Abbildung 7.7 zu sehen. Darstellung 4 erhielt bei dieser Umfrage mit Abstand den meisten Zuspruch. Neun von zwölf Teilnehmern gefiel der Wetterbericht im Spiel mit dieser Darstellung am besten. Darstellung 3 erhielt drei Teilnehmerstimmen. Darstellung 2 gefiel immerhin noch einem Teilnehmer am besten. Die einfache, textuelle Darstellung, die keinerlei Unsicherheit kommunizierte, erhielt keinerlei Zuspruch.

### 7.5. Kommentare der Teilnehmer

Es wurden zwei konstruktive Kommentare von den Teilnehmern hinterlassen. In einem Kommentar merkte der Teilnehmer an, dass er die Unterteilung der Intervalle des Histogramms in Darstellung 3 zu grob fand. In dem anderen Kommentar wurde mitgeteilt, dass dem Teilnehmer Darstellung 4 mehr zusagte als Darstellung 2.

### 7.6. Diskussion

In diesem kontrollierten Experiment ist sowohl der Anteil optimaler, als auch suboptimaler Entscheidungen höher als beim öffentlichen Experiment. Die Spieler trafen unabhängig von Darstellung und Risikoverhalten in über 70% der Runden sinnvolle Entscheidungen. Legt man ein risikofreudiges Verhalten bei der Interpretation der Entscheidungen zugrunde, waren mit Darstellung 4 sogar über 90% der Entscheidungen sinnvoll. Die Ergebnisse dieses Experimentes unterstützen die Resultate von Roulston et. al [MSR06] und Joslyn et. al [JL12], die feststellten, dass die meisten Menschen mit Unsicherheiten umgehen können und auf ihnen basierend Entscheidungen treffen können.

Im öffentlichen Experiment trafen die Teilnehmer mit den Darstellungen 3 und 4 mehr optimale Entscheidungen als mit den Darstellungen 1 und 2. Der Anteil an suboptimalen Entscheidungen war jedoch mit den Darstellungen 1 und 2 höher. Im Gegensatz dazu haben die Teilnehmer des kontrollierten Experimentes in den Spielen mit den Darstellungen 3 und 4 nicht nur mehr optimale Entscheidungen, sondern auch mehr suboptimale, aber sinnvolle Entscheidungen getroffen als mit den Darstellungen 1 und 2. Die Darstellungen 3 und 4, die Informationen zu den Unsicherheiten detailliert und ausgeprägt darstellten, schnitten also bei beiden Entscheidungskategorien am besten ab.

Die Ergebnisse weisen darauf hin, dass Laien bessere Entscheidungen treffen, wenn ihnen Informationen zu Unsicherheiten zur Verfügung gestellt werden. Mit Darstellung 4, die die Wahrscheinlichkeitsverteilungen am exaktesten aufschlüsselte, trafen die Teilnehmer sowohl am meisten optimale als auch am meisten suboptimale Entscheidungen. Sie war im Kontext dieses Experimentes die Darstellung, mit der die Teilnehmer die meisten richtigen Entscheidungen getroffen haben und stellte sich somit als beste Visualisierungsalternative zur Vermittlung von unsicheren Daten an Laien heraus. Diese Schlussfolgerung stimmt ebenfalls mit der eigenen Einschätzung der Teilnehmer nach dem Spiel überein (s. Abbildung 7.6). Dort erhielt Darstellung 4 von allen Darstellungen die durchschnittlich höchste Bewertung bezüglich der Frage, ob die Darstellung eine Entscheidungshilfe sei. Weiterhin erhielt Darstellung 4 auch bei der Frage nach der allgemein besten Darstellung (s. Abbildung 7.7) den meisten Zuspruch.

Die durchschnittliche Dauer, die ein Spieler benötigte um den Wetterbericht in der jeweiligen Darstellungsform zu betrachten, war ähnlich lange wie beim öffentlichen Experiment. Die schlichte Darstellung 1, die sehr wenige Informationen enthielt, wurde auch in diesem Experiment am kürzesten von den Spieler eingeblendet. Darstellung 2, die ebenfalls relativ wenige Informationen abbildet, ließen sich die Spieler am zweitkürzesten anzeigen. Im Gegensatz zum öffentlichen Experiment betrachteten die Spieler im kontrollierten Experiment allerdings Darstellung 4 kürzer als Darstellung 3. Dies könnte daran liegen, dass man die Entwicklung des Vorhersagezeitraums bei Darstellung 4 sehr leicht interpretieren kann, indem man einfach die Lage der Kurven zueinander vergleicht. Liegen beispielsweise die Kurven der Vorhersagen für den Niederschlag in drei Tagen und in zwei Tagen weiter rechts der Kurve für den folgenden Tag, so ist auf einen Blick erkennbar, dass der Niederschlag in den folgenden drei Tagen kontinuierlich zunimmt. Die Legende in Darstellung 4 wurde hauptsächlich benötigt um die verschiedenen Tage der Prognose unterscheiden zu können. Die restlichen Informationen konnten direkt aus dem Graphen abgelesen werden. In Darstellung 3 dagegen musste man zuerst das Werteintervall für jeden dargestellten Balken in der Legende ablesen, um die Wahrscheinlichkeitsverteilung interpretieren zu können.

Die durchschnittliche Dauer der Einblendung des Wetterberichts in der ersten Runde war für jede Darstellung höher als beim öffentlichen Experiment. Die Reihenfolge änderte sich im Vergleich zum öffentlichen Experiment jedoch nicht. Darstellung 1 wurde auch in diesem Experiment am wenigsten lange betrachtet. Mit zunehmender Komplexität der Darstellungen stieg die Dauer der Einblendung in der ersten Runde. Man kann daher vermuten, dass Darstellung 1 auch bei der erstmaligen Ansicht durch die Teilnehmer schneller und intuitiver interpretiert werden kann, als die graphischen Darstellungen. Dies ist vermutlich darauf zurück zu führen, dass mit Darstellung 1 zum einen sehr wenige Informationen dargestellt werden und zum anderen der Erwartungswert eine sehr gebräuchliche Angabe ist. Je komplexer die Darstellungen waren, desto länger wurden sie in der ersten Runde eingeblendet. Außerdem fällt auf, dass mit Darstellung 4 die durchschnittliche Dauer der Einblendung der ersten Runde über doppelt so hoch war wie die Dauer der Einblendung in allen Runden. Dies lässt vermuten, dass es bei Darstellung 4 den größten Lerneffekt bei den Teilnehmern gegeben haben könnte.

Weiterhin wurde ermittelt, in wievielen Runden der Wetterbericht überhaupt eingeblendet wurde. Das Ergebnis wurde prozentual angegeben und in Abbildung 7.4 dargestellt. Hierbei fällt auf, dass im Gegensatz zu den Ergebnissen im öffentlichen Experiment die Einsichten des Wetterberichts durchweg, unabhängig von Darstellung und Runde, sehr viel höher lagen. Dies ist vermutlich darauf zurückzuführen, dass dieses Experiment unter kontrollierten Bedingungen stattfand und ein finanzieller Anreiz für gutes Spiel in Form eines Gutscheines vorhanden war. Aus diesem Grund sahen sich die Teilnehmer die Darstellung vermutlich auch dann an, wenn ihre Interpretation hohen Aufwand erforderte oder sie dem Teilnehmer nicht zusagte, da nur so ein gutes Ergebnis erzielt werden konnte. In der letzten Runde ging die Anzahl der Einblendungen des Wetterberichtes zurück. Dies kann mutmaßlich dadurch begründet werden, dass in der letzten Runde nur noch Pflanzen geerntet, aber nicht mehr angebaut, werden konnten, weshalb die Notwendigkeit den Wetterbericht einzusehen entfiel. Ansonsten konnten in den unterschiedlichen Runden keine erheblichen Unterschiede bei der Dauer der Einblendung festgestellt werden.

Die durchschnittliche Bewertung der vier Fragen fiel in diesem Experiment ähnlich aus, wie bei der Online-Umfrage und dem öffentlichen Experiment. Bei der Frage, wie gut eine Darstellung die Entscheidung unterstützt, ergab sich sowohl im Vergleich zu den Ergebnissen der Online-Umfrage, als auch zum öffentlichen Experiment (vgl. Abbildung 6.8) ein Unterschied. Bei dieser Frage schnitt Darstellung 2 im öffentlichen Experiment und in der Online-Umfrage am besten ab. Beim kontrollierten Experiment bewerteten die Teilnehmer Darstellung 4 am höchsten. Das Ergebnis der Befragung deckt sich auch mit den Ergebnissen aus der Analyse der Daten des Spiels. Die tatsächliche Anzahl an optimalen und suboptimalen Entscheidungen, war mit Darstellung 4 am höchsten. Das bedeutet, dass die Teilnehmer in diesem Experiment besser einschätzen konnten, ob eine Darstellung tatsächlich eine gute Entscheidungshilfe war.



## 8. Zusammenfassung und Fazit

Obwohl wir sehr häufig Entscheidungen aufgrund von Risiken und Unsicherheiten, wie sie beispielsweise im täglichen Wetterbericht vorkommen, treffen müssen, werden uns hierbei die Angaben zur Unsicherheit entweder auf sehr verschiedene Arten oder überhaupt nicht dargestellt. Dies liegt einerseits daran, dass nicht ausreichend erforscht ist, ob Darstellungen, die Angaben zur Unsicherheit enthalten, für die Menschen einen Mehrwert bei ihren Entscheidungen bringen und andererseits daran, dass unklar ist, welche Darstellungen und Visualisierungen sich am besten für die Vermittlung von unsicheren Daten an Laien eignen. Da die Forschung, die bisher bereits zu dieser Problemstellung betrieben wurde, meist sehr theoretisch blieb und nur mit Hilfe von Umfragen untermauert wurde, fehlen empirischen Daten aus echten Entscheidungen, die unter realen Bedingungen gewonnen wurden. Um diese Lücke zu schließen, wurde in dieser Arbeit ein Spiel entwickelt, bei dem die Teilnehmer auf Basis von unsicheren Daten Entscheidungen fällen mussten. Beruhend auf der Analyse der Daten und den Entscheidungen aus den gespielten Partien, sollten Erkenntnisse bezüglich der Kommunikation von unsicheren Daten als Lösungsansatz für die Problemstellung gewonnen werden.

Es wurden zunächst zwölf verschiedene Darstellungen zur Kommunikation von Unsicherheiten identifiziert und mit Hilfe einer Online-Umfrage evaluiert. Vier der Darstellungen wurden ausgewählt und im Spiel umgesetzt. Um die erforderlichen empirischen Daten zu erhalten, wurde das Spiel darauf hin zum einen öffentlich und unmoderiert, zum anderen unter kontrollierten Bedingungen von den Teilnehmern gespielt. Die dadurch gewonnenen Daten wurden analysiert, um herauszufinden, ob Menschen Darstellungen, die Angaben zur Unsicherheit enthalten, zielführend einsetzen können. Außerdem sollte beantwortet werden, ob es einen Vorteil für die Entscheidungsträger darstellt, wenn sie Prognosen mit Informationen zur Unsicherheit erhalten oder ob deterministische Prognosen zu den gleichen Entscheidungen führen. Weiterhin sollte ermittelt werden, welche Darstellungen geeignet sind, um unsichere Daten an Laien zu vermitteln, wie viel Zeit sie benötigen, um die unterschiedlichen Darstellungen bei erstmaliger Betrachtung zu interpretieren und eine Entscheidung abzuleiten.

Die Ergebnisse dieser ersten Experimente mit realen Entscheidungen können nicht alle Fragen abschließend beantworten, jedoch erlauben sie, einige vorläufige Schlussfolgerungen zu ziehen. Die Analyse der Entscheidungsdaten hat gezeigt, dass die meisten Menschen mit Unsicherheiten umgehen und darauf basierend Entscheidungen treffen können. Außerdem war zu erkennen, dass die Angabe von Informationen zur Unsicherheit für die meisten Menschen einen Mehrwert für ihre Entscheidungen bedeutet, sogar, dass die Genauigkeit der Angaben direkt die Qualität der Entscheidungen beeinflusst. Umso detaillierter die Angaben zur Unsicherheit sind, desto bessere Entscheidungen werden von Laien getroffen. Die kontinuierliche Darstellung eines Histogramms in Form eines Funktionsgraphen stellte sich im Kontext der tatsächlichen Entscheidungen unter Einbeziehung der Unsicherheiten des Wetters als beste Visualisierungsalternative zur Vermittlung von unsicheren Daten an Laien heraus. Ebenfalls gut eignet sich ein Histogramm in Form eines Balkendiagramms als Visualisierung zur Vermittlung von unsicheren Daten an Laien. Die Selbsteinschätzung von Laien, welche Darstellung

## 8. Zusammenfassung und Fazit

---

für sie persönlich hilfreich ist, um eine gute Entscheidung basierend auf unsicheren Daten zu treffen, deckt sich sehr häufig nicht mit der Realität. Die Teilnehmer der Studie waren nie unter Zeitdruck und nur die Qualität der Entscheidung war entscheidend. In einem Szenario, bei welchem die Qualität der Entscheidung nicht ausschlaggebend, aber die Zeit für die Entscheidung kritisch ist, kann eine deterministische Darstellung sinnvoll sein.

Eine große Einschränkung dieses Experiments ist, dass die Teilnehmerzahl relativ gering war und die Ergebnisse deshalb nicht repräsentativ sind. Die geringe Teilnehmerzahl ist unter anderem vermutlich darauf zurückzuführen, dass die Teilnahme am Spiel mit allen vier Darstellungen durchschnittlich ungefähr eine Stunde benötigte, was vielen potentiellen Teilnehmern zu lange dauerte. Weiterhin besteht eine Einschränkung darin, dass die Entscheidungsszenarien im Spiel sehr viel simpler als komplexe reale Entscheidungsprobleme sind. Der Einsatz von Entscheidungsszenarien in kontrollierter Umgebung lies sich jedoch nur auf diese Weise umsetzen, da die Ergebnisse durch die starke Vereinfachung reproduzierbar waren und eine fokussiertere Untersuchung zuließen. Aus einer experimentell-ökonomischen Sicht war für das öffentliche Experiment eine weitere Einschränkung, dass die Teilnehmer keine realen finanziellen Auszahlungen, abhängig von ihren Entscheidungen, erhalten haben. Durch einen finanziellen Anreiz hätten die Teilnehmer gegebenenfalls größere Anstrengungen in die Wahl ihrer Anbauentscheidungen gelegt. Diese Einschränkung wurde im kontrollierten Experiment, zumindest teilweise, überwunden. Hier erhielten die Teilnehmer einen finanziellen Anreiz die bestmöglichen Entscheidungen zu treffen in Form eines Gutscheins, der an den Spieler mit dem höchsten Geldstand ausgegeben wurde. Dieser Umstand könnte begründen, warum im kontrollierten Experiment die Anzahl und Qualität der Entscheidungen, gegenüber dem öffentlichen Experiment, höher ausgefallen sind.

Auch wenn diese Erkenntnisse im Kontext der Wettervorhersage gewonnen wurden, gibt es viele weitere Bereiche, in denen es erforderlich ist Unsicherheiten zu kommunizieren und in denen die Erkenntnisse weiterhelfen und Anwendung finden können. Viele Entscheidungen, die auf Basis von unsicheren Daten getroffen werden, haben weitreichende Folgen. Beispiele dafür sind der Vergleich von kurzfristigen und langfristigen Risiken einer medizinischen Behandlung oder die Kommunikation von Unsicherheiten für die Reise- und Ankunftszeiten mit öffentlichen Verkehrsmitteln.

Die in dieser Arbeit gewonnenen Erkenntnisse tragen dazu bei zu erkennen, ob unsichere Daten an Laien vermittelt werden sollen und welche Darstellungen oder Visualisierungen dafür zu bevorzugen sind. Die Erkenntnisse unterstützen und ergänzen weitere Entwicklungen und zukünftige Arbeiten auf diesem Gebiet. Sie lassen sich beispielsweise einsetzen, wenn es um die Erforschung von Visualisierungen für andere Aufgaben und Anwendungen geht, bei denen unsichere Daten an Laien kommuniziert werden sollen. Das Verständnis über die Fähigkeit zur Kommunikation von unsicheren Daten mittels verschiedener Visualisierungsalternativen und Darstellungsformen kann weiterhin dazu beitragen, dass unsichere Daten seltener deterministisch dargestellt werden und stattdessen die probabilistische Kommunikation von unsicheren Daten weitere Verbreitung findet. Die probabilistische Darstellung von unsicheren Daten kann, wie gezeigt, einen großen Mehrwert für die Entscheidungen der Menschen bedeuten.

# A. Anhang

## A.1. Mockup der Anwendung Farm-Smart

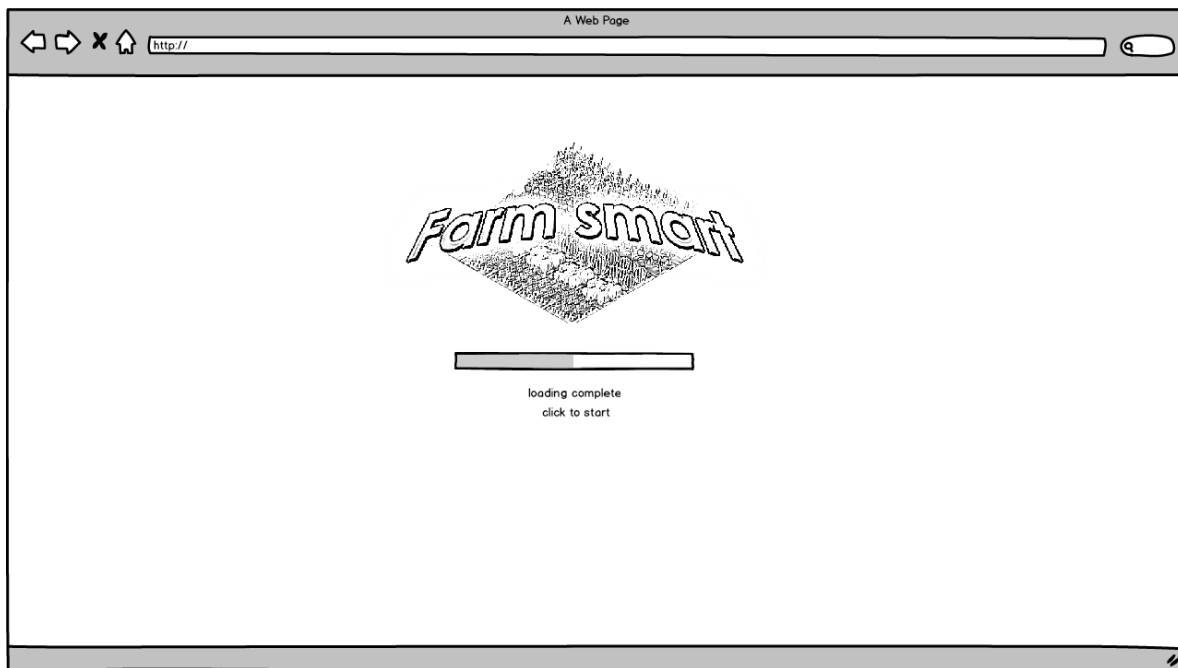


Abbildung A.1.: Logo und Fortschrittsbalken (Preloading der Bilder)

## A. Anhang

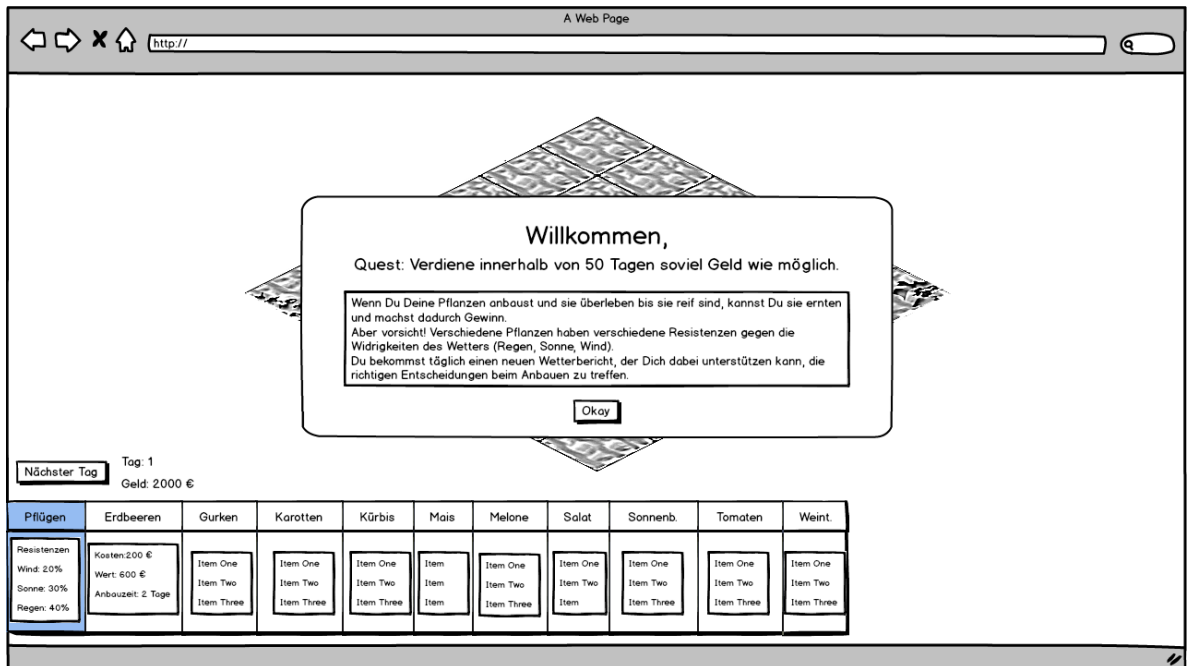


Abbildung A.2.: Willkommensbildschirm, Aufgabenbeschreibung und kleine Einweisung

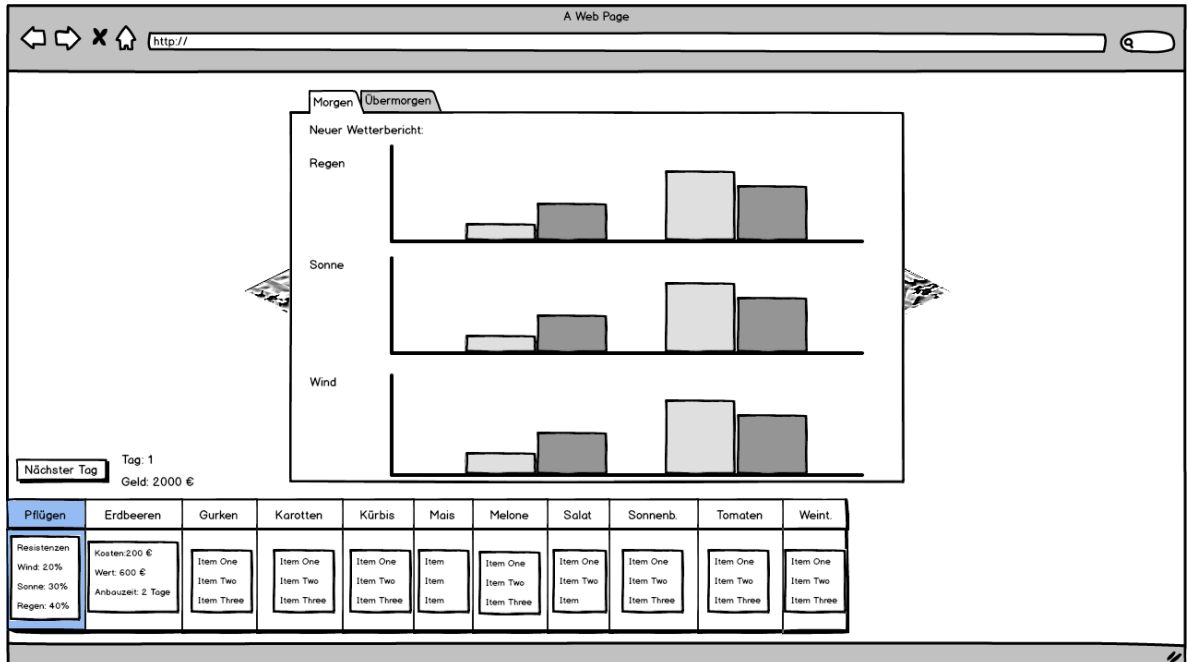
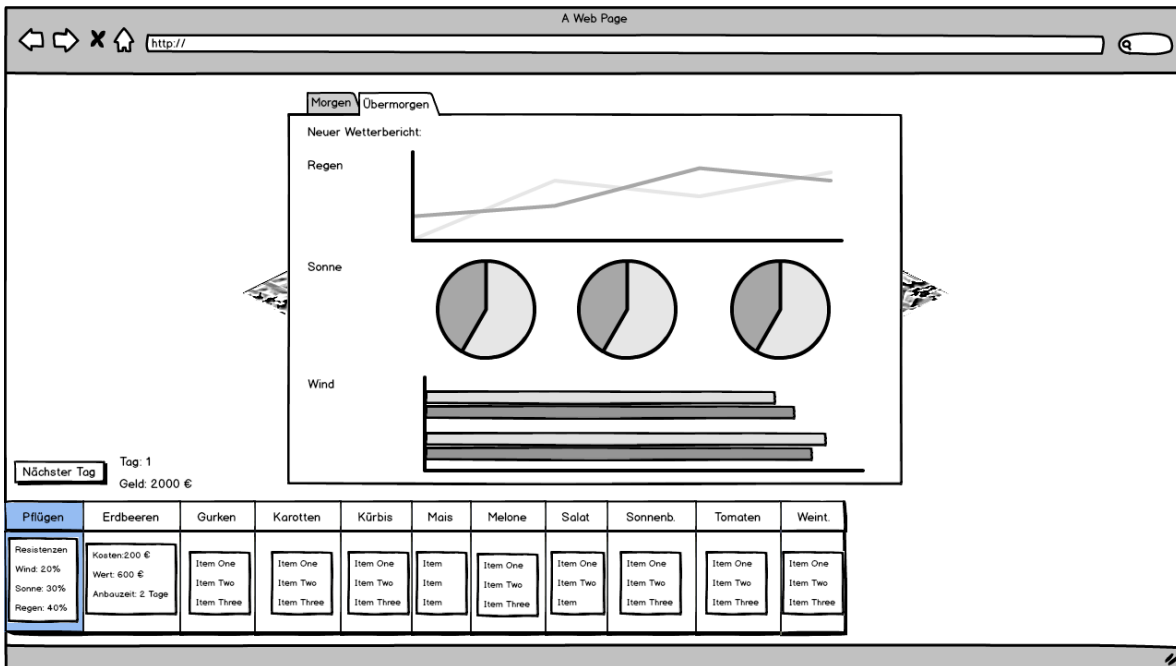


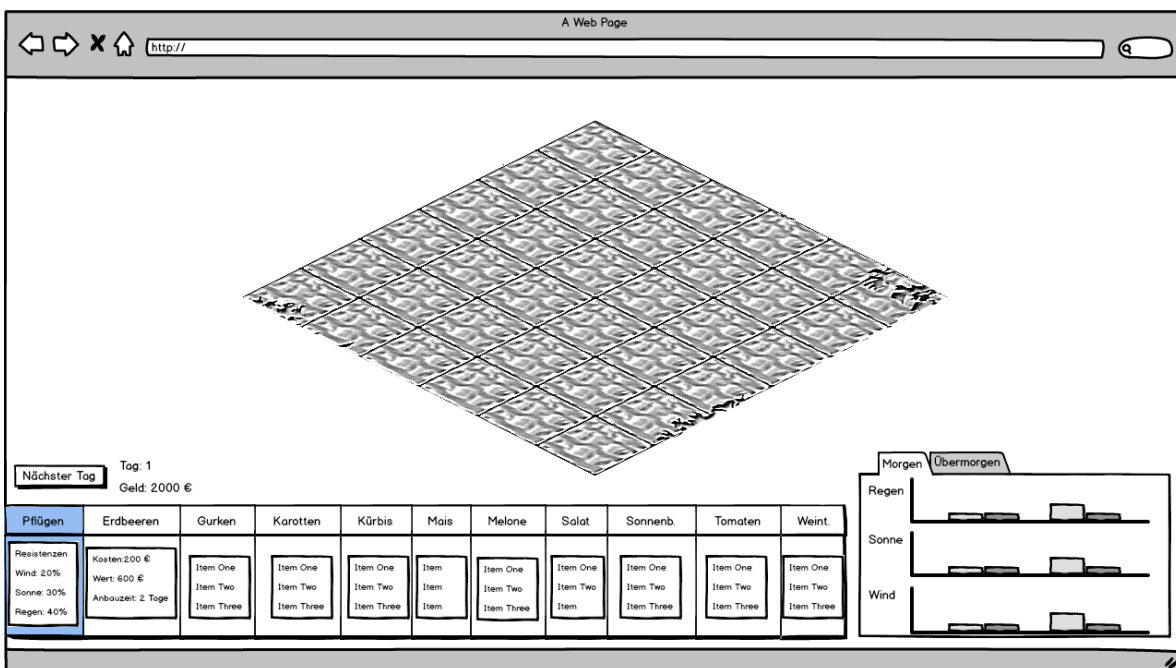
Abbildung A.3.: Dem Spieler wird der erste Wetterbericht für den morgigen Tag angezeigt



## A.1. Mockup der Anwendung Farm-Smart



**Abbildung A.4.:** Ansicht des ersten Wetterberichts für den übermorgigen Tag. Es werden verschiedene Visualisierungen der Wahrscheinlichkeitsverteilungen angezeigt.



**Abbildung A.5.:** Der Wetterbericht ist vom Spieler unten rechts immer noch einzusehen. Der Spieler kann nun darauf basierend entscheiden, welche Pflanzen er anbauen möchte.

## A. Anhang

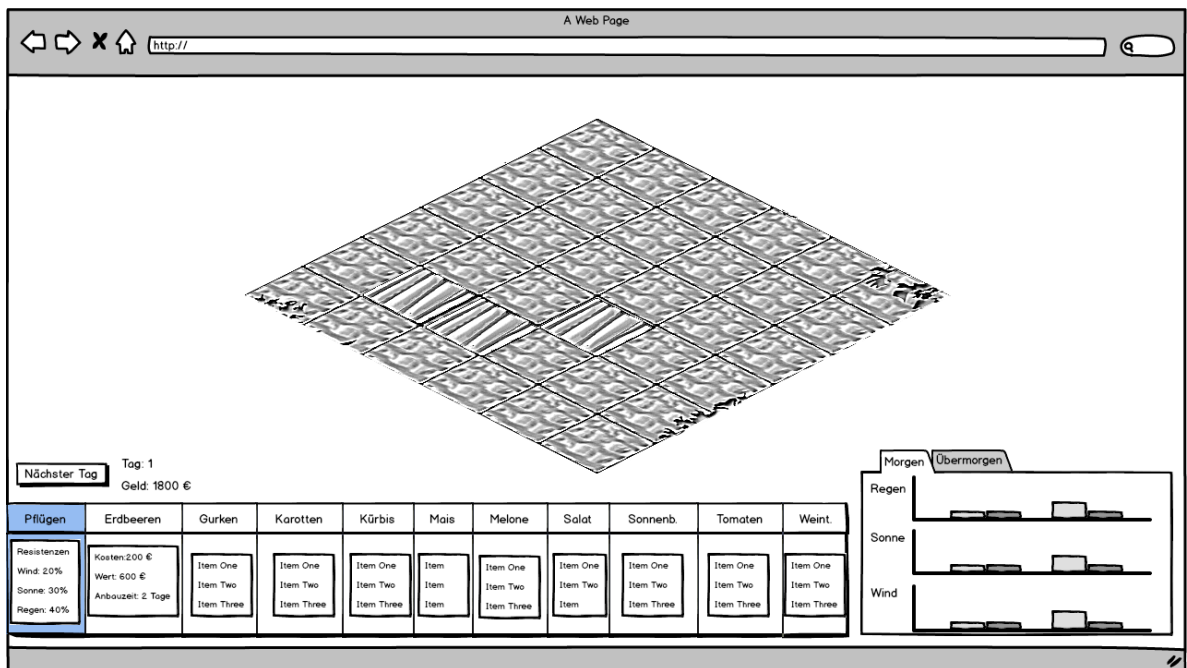


Abbildung A.6.: Bevor der Spieler etwas anpflanzen kann, muss er zuerst ein Ackerfeld pflügen.

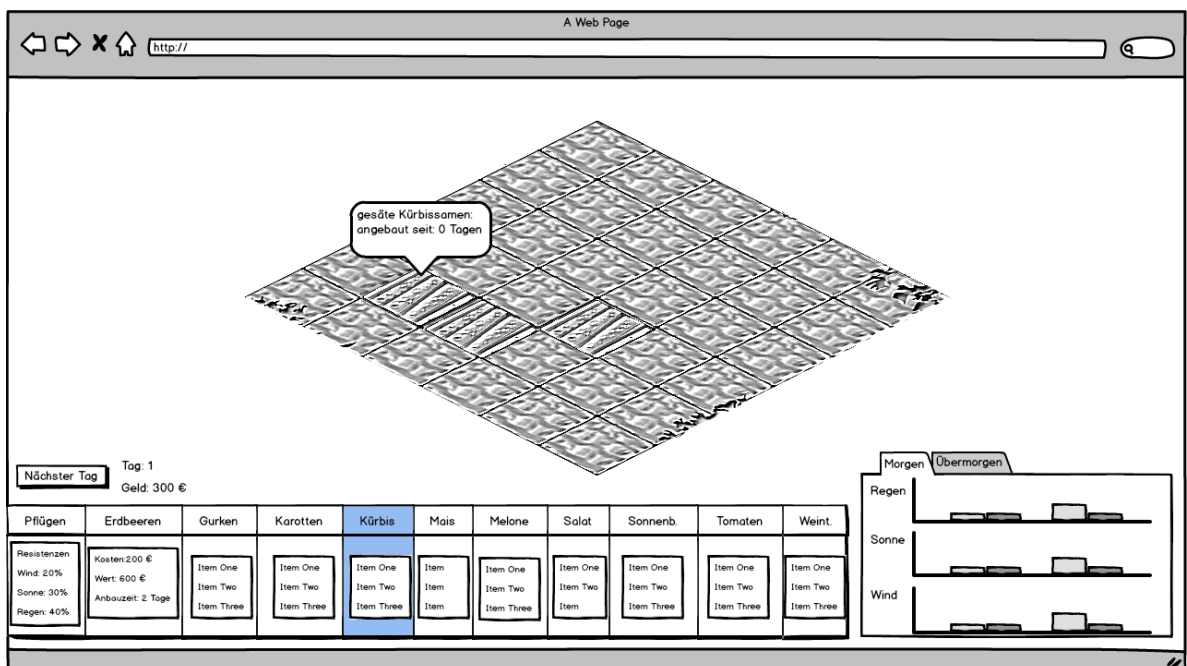
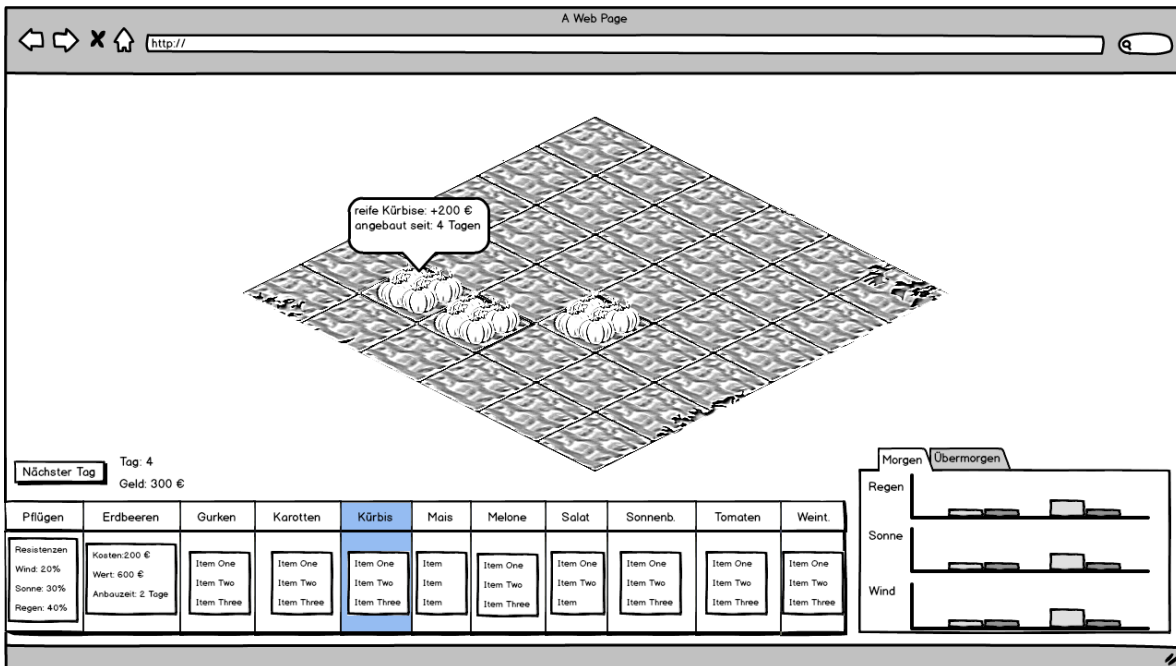
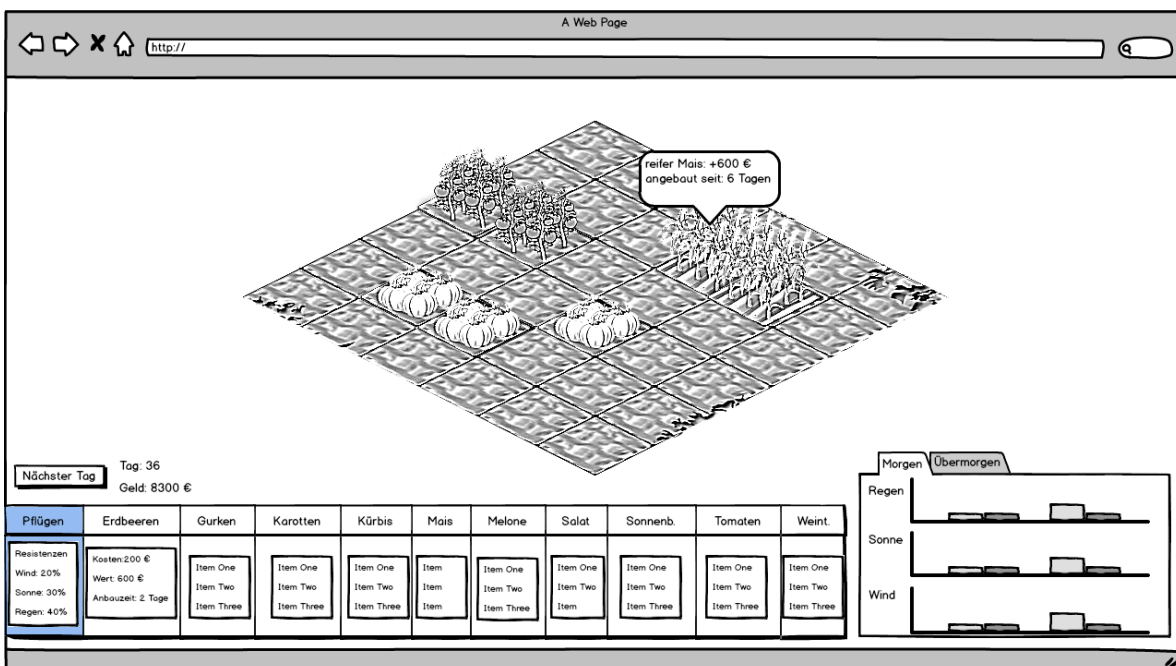


Abbildung A.7.: Die Pflanzen haben verschiedene Anbaukosten, Werte beim Ernten, Anbauzeiten und Anforderungen an Regen, Sonne und Wind. Der Spieler entscheidet sich aufgrund des Wetterberichts Kürbisse anzubauen.

## A.1. Mockup der Anwendung Farm-Smart



**Abbildung A.8.:** Vier Tage sind vergangen. Die Kürbisse sind reif, da sie eine Anbauzeit von 4 Tagen haben. Der Spieler kann sie nun ernten.



**Abbildung A.9.:** Es vergehen weitere Tage. Der Spieler bekommt täglich neue Wetterberichte und baut darauf basierend neue Pflanzen an und erntet reife Pflanzen.

## A. Anhang

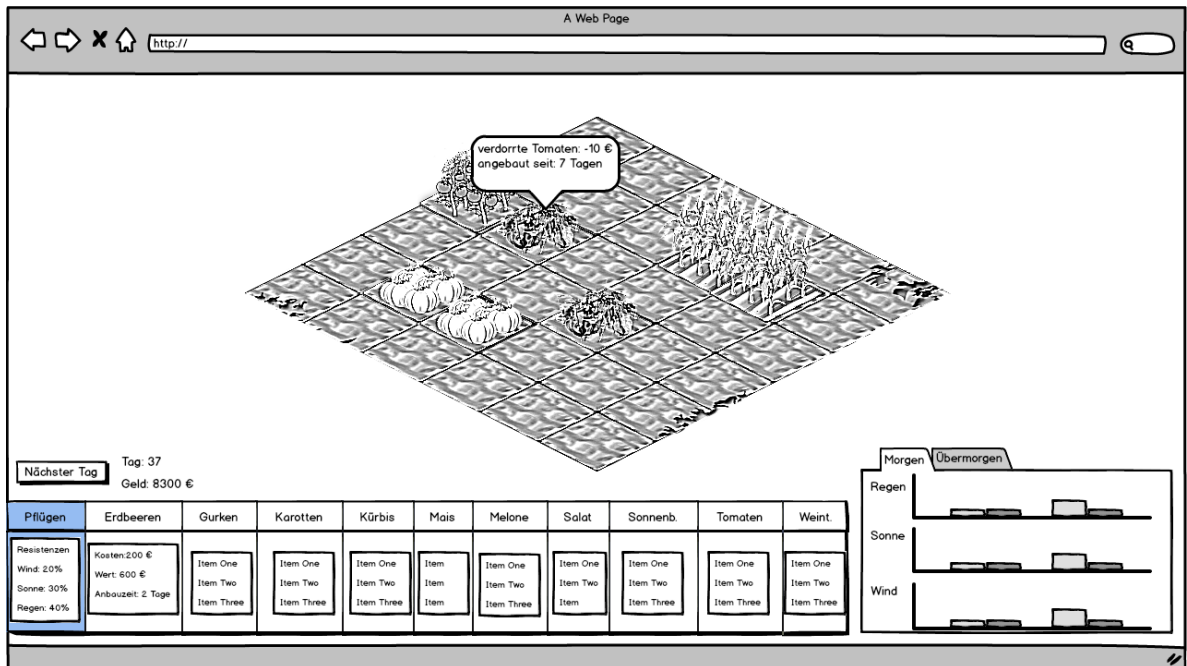


Abbildung A.10.: Tritt das falsche Wetter ein, können die Pflanzen kaputt gehen und er macht Verlust. Die kaputten Pflanzen muss der Spieler entgeltlich entfernen.

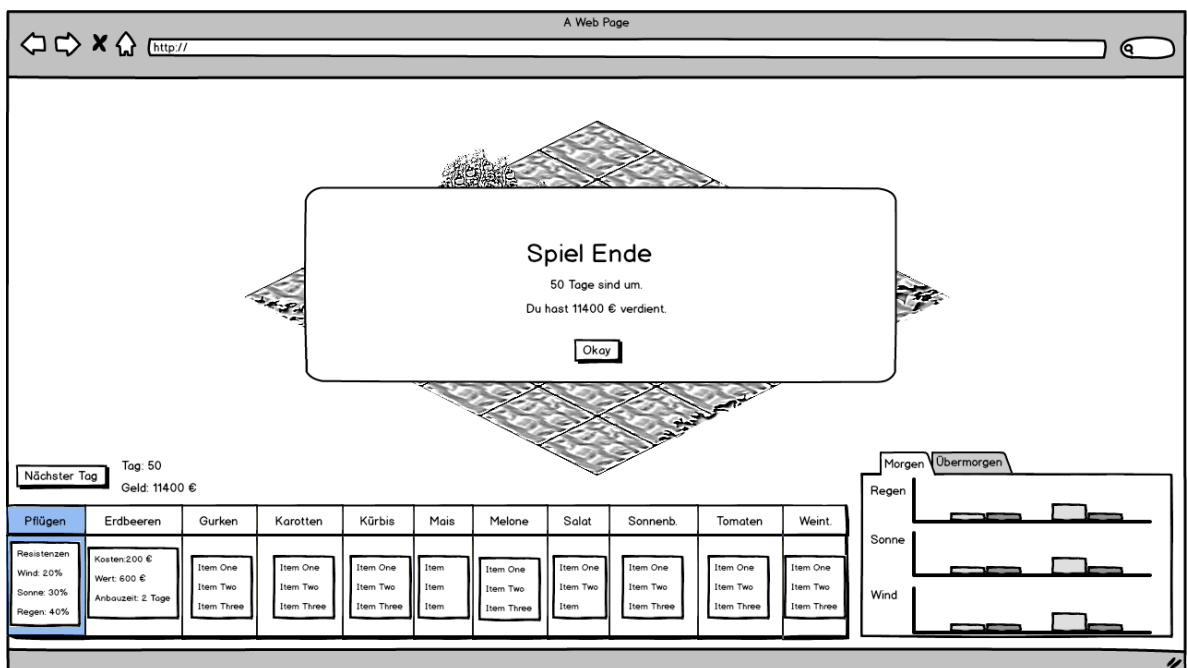


Abbildung A.11.: Die Spieldauer ist um. Das Spiel endet und der endgültige Geldstand wird angezeigt und gespeichert.

## A.1. Mockup der Anwendung Farm-Smart

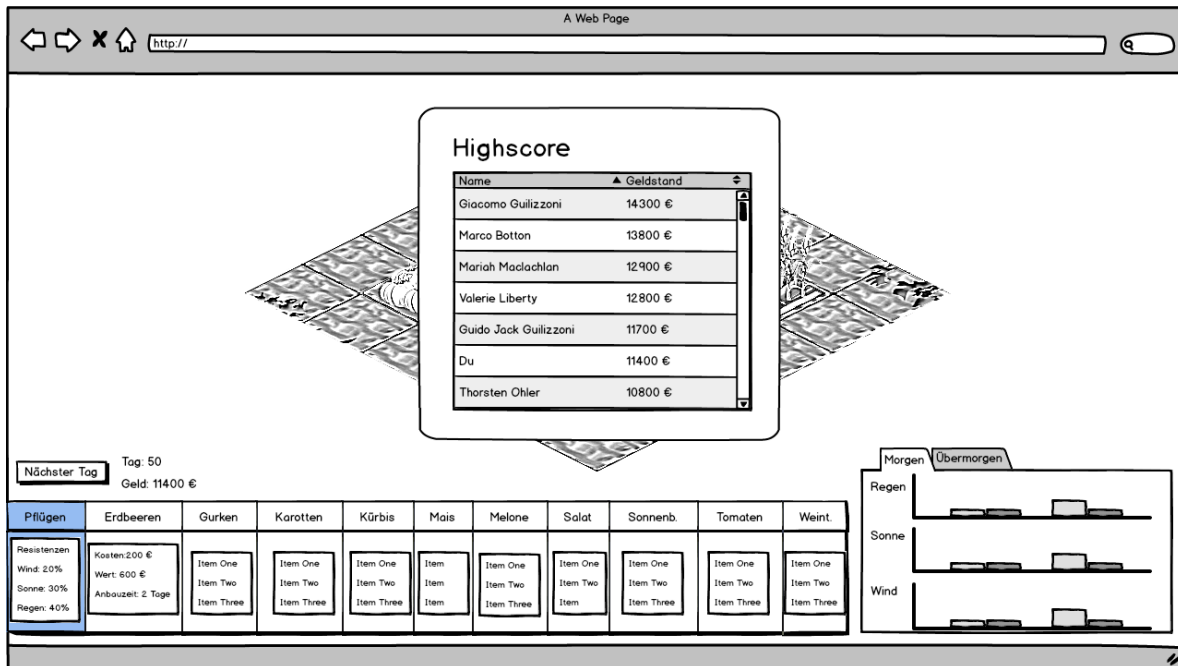


Abbildung A.12.: Der Highscore der bisher gespielten Partien werden dem Spieler gezeigt.



# Literaturverzeichnis

- [AA03] S. P. Apanovitch A., McCarthy D. Using message framing to motivate HIV testing among low-income, ethnic minority women. *Health Psychology*, 22(1):60–67, 2003. (Zitiert auf Seite 18)
- [AT92] D. K. A. Tversky. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323, 1992. (Zitiert auf Seite 54)
- [BD03] K. P. B. E. Berry D., Raynor D. Patients understanding of risk associated with medication use: Impact of European Commission guidelines and other risk scales. *Drug Safety*, 26(1):1–11, 2003. (Zitiert auf Seite 17)
- [BS95] G. S. Banks SM., Salovey P. The effects of message framing on mammography utilization. *Health Psychology*, 14(2):178–184, 1995. (Zitiert auf Seite 18)
- [CO02] J. D. M. Chris Olston. Visualizing Data with Bounded Uncertainty. *IEEE Symposium on Information Visualization*, 2002. (Zitiert auf Seite 22)
- [Cou06] N. R. Council. *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts. Report from the Committee on Estimating and Communicating Uncertainty in Weather and Climate Forecasts*. The National Academies Press, 2006. (Zitiert auf Seite 13)
- [DA84] K. D., T. A. Choices, values, and frames. *American Psychologist*, 39:341–350, 1984. (Zitiert auf Seite 18)
- [DK79] A. T. D. Kahneman. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979. (Zitiert auf Seite 54)
- [FL92] A. R. Forrow L., Taylor WC. Absolutely relative: how research results are summarized can affect treatment decisions. *American Journal of Medicine*, 92:121–124, 1992. (Zitiert auf Seite 18)
- [FP13] J. T. P. S. D. D. Florian Pappenberger, Elisabeth Stephens. Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication. *Hydrological Process.*, 27:132–146, 2013. (Zitiert auf den Seiten 23 und 26)
- [GG03] E. A. Gigerenzer G. Simple tools for understanding risks: from innumeracy to insight. *BMJ: British Medical Journal*, 327:741–744, 2003. (Zitiert auf den Seiten 18 und 19)
- [GG05] E. v. d. B. B. F. K. V. K. Gerd Gigerenzer, Ralph Hertwig. A 30Probabilistic Weather Forecasts. *Risk Analysis*, 25(3):623–630, 2005. (Zitiert auf den Seiten 13 und 20)

- [GM10] G.-R. R. Galesic M. Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Arch Intern Med*, 2010. (Zitiert auf den Seiten 11 und 20)
- [HI87] M. G. M. Harald Ibrekk. Graphical Communication of Uncertain Quantities to Nontechnical People. *Risk Analysis*, 7(4):519–529, 1987. (Zitiert auf den Seiten 21, 26 und 27)
- [JAW08] P. H. J. P. K. A. C. P. J. Arjan Wardekker, Jeron P. van der Sluijs. Uncertainty communication in environmental assessments views from the Dutch science-policy interface. *Environmental Science&Policy II*, S. 627–641, 2008. (Zitiert auf den Seiten 15 und 22)
- [JL12] S. L. Joslyn, J. E. LeClerc. Uncertainty Forecasts Improve Weather-Related Decisions and Attenuate the Effects of Forecast Error. *Journal of Experimental Psychology: Applied*, 18(1):126–140, 2012. (Zitiert auf den Seiten 15, 63 und 74)
- [KI93] J. L. K. A. Kirsch IS., Jungeblut A. Adult Literacy in America. A First Look at the Results of the National Adult Literacy Survey. *U. S. Department of Education*, 1993. (Zitiert auf Seite 16)
- [LI01] R. B. S. G. R. B. Lipkus IM., Samsa G. General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21(1):37–44, 2001. (Zitiert auf Seite 16)
- [MB82] S. H. J. T. A. McNeil BJ., Pauker SG. On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, 306(21):1259–1262, 1982. (Zitiert auf Seite 18)
- [MD93] J. S. W. J. R. J. Malenka DJ., Baron JA. The framing effect of relative and absolute risk. *Journal of General Internal Medicine*, 8(10):543–548, 1993. (Zitiert auf Seite 18)
- [MSR06] A. N. K. A. L. S.-C. Mark S. Roulston, Gary E. Bolton. A Laboratory Study of the Benefits of Including Uncertainty Information in Weather Forecasts. *American Meteorological Society*, 21:116–122, 2006. (Zitiert auf den Seiten 15, 52, 63 und 74)
- [Myl02] K. R. Mylne. Decision-making from probability forecasts based on forecast value. *Meteorol. Appl.*, 9:307–315, 2002. (Zitiert auf Seite 14)
- [Neu44] O. M. John von Neumann. *Theory of Games and Economic Behavior*. Princeton University Press, 1944. (Zitiert auf Seite 54)
- [RA93] A. C. K. K. M. C. Rothman AJ., Salovey P. The influence of message framing on intentions to perform health behaviors. *Journal of Experimental Social Psychology*, 29:408–433, 1993. (Zitiert auf Seite 18)
- [RA97] S. P. Rothman AJ. Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin*, 121(1):3–19, 1997. (Zitiert auf Seite 18)
- [RA99] B. B. D. J. S. P. Rothman AJ., Martino SC. The systematic influence of gain- and loss-framed messages on interest in and use of different types of health behavior. *Personality and Social Psychology Bulletin*, 25(11):1355–1369, 1999. (Zitiert auf Seite 18)
- [REM08] J. K. L. Rebecca E. Morss, Julie L. Demuth. Communicating Uncertainty in Weather Forecasts A Survey of the U.S. Public. *American Meteorological Society, Weather and Forecasting*, 23:974–991, 2008. (Zitiert auf den Seiten 13, 16, 20, 22, 54 und 63)



- [REM10] J. L. D. Rebecca E. Morss, Jeffrey K. Lazo. Examining the use of weather forecasts in decision scenarios: results from a US survey with implications for uncertainty communication. *Meteorological Appl.*, 17:149–162, 2010. (Zitiert auf Seite 19)
- [SDM82] R. L. D. Sandra D. McNabb. Development and Test of a Motivational Approach and Materials for Increasing Use of Restraints. *National Technical Information Service*, 1982. (Zitiert auf Seite 19)
- [SL09] W. H. Schwartz LM., Woloshin S. Using a drug facts box to communicate drug benefits and harms two randomized trials. *Annals of Internal Medicine*, 150(8):516–527, 2009. (Zitiert auf Seite 17)
- [SP00] M. J. M. D. Slovic P. Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law & Human Behavior*, 24(3):271–296, 2000. (Zitiert auf Seite 18)
- [Tul09] N. W. S. Tulsa. General Forecast Terminology and Tables, 2009. URL [http://www.srh.noaa.gov/tsa/?n=forecast\\_terms](http://www.srh.noaa.gov/tsa/?n=forecast_terms). (Zitiert auf Seite 20)
- [UH00] R. H. G. G. Ulrich Hoffrage, Samuel Lindsey. Medicine. Communicating statistical information. *Science*, 290:2261–2262, 2000. (Zitiert auf den Seiten 18 und 19)
- [WT86] R. A. Z. R. F. B. Wallsten TS., Budescu DV. Measuring the vague meaning of probability terms. *Journal of Experimental Psychology: General*, 115(4):348–365, 1986. (Zitiert auf Seite 17)
- [ZFB08] U. P. Zikmund-Fisher B., Fagerlin A. Improving understanding of adjuvant therapy options via simpler risk graphics. *Cancer*, 113(12):3382–3390, 2008. (Zitiert auf Seite 17)

Alle URLs wurden zuletzt am 20. 10. 2014 geprüft.



## **Erklärung**

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

---

Ort, Datum, Unterschrift